



OPEN

DATA DESCRIPTOR

Harmonized aerosol size distribution, cloud condensation nuclei, chemistry and optical properties at 10 sites

Elisabeth Andrews^{1,2,6} , Inés Zabala^{3,4,6} , Gerardo Carrillo-Cardenas⁵ , Gloria Titos^{3,4} , Juan Andrés Casquero-Vera^{3,4} & A. Gannet Hallar⁵

A harmonized data set of aerosol particle number size distributions and cloud condensation nuclei (CCN) concentrations at 10 US Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) sites is presented. Additional complementary data including total particle number concentration, aerosol chemical composition and aerosol optical properties are included where available. Calculated parameters related to particle activation, (e.g., activated fraction, hygroscopicity parameter and critical diameter) are provided. DOE/ARM data are acquired and processed following DOE/ARM protocols which are based on international recommendations. Here, the data have undergone additional stages of quality assessment including various closure tests. This data set may be useful for studying aerosol processes and, with other data, assessing aerosol-cloud interactions. High-resolution (5-min) data is provided along with 1-h averaged data, making the data set suitable for exploring the importance of new particle formation (NPF) events on CCN concentrations. Further applications include model and satellite evaluation of aerosol size distributions and CCN concentrations, and exploring relationships among different aerosol properties which may be useful for improving model parametrizations.

Background & Summary

Atmospheric aerosols impact climate, air quality and visibility, and human health (e.g.^{1,2}). Aerosols affect the Earth's radiation balance directly by scattering and absorbing sunlight and indirectly through their role as cloud condensation nuclei (CCN)^{3–6}. To date, radiative forcing through aerosol–cloud interactions (ACI) constitutes the least understood anthropogenic influence on climate⁷. It remains a significant challenge to reduce these uncertainties and to thereby increase our confidence in predictions of global and regional climate change^{7,8}.

One leading cause of the large uncertainty in ACI is detailed knowledge about how particles evolve to become effective CCN under different conditions/environments. CCN are emitted to the atmosphere directly from both anthropogenic⁹ and biogenic sources¹⁰, and over 50% of CCN are thought to be formed by condensational growth of smaller particles¹¹. These smaller particles may originate from atmospheric new particle formation (NPF), anthropogenic combustion or other emissions. NPF is the formation of thermodynamically stable clusters from multi-component systems followed by condensational growth of these clusters to a detectable size¹². Ultimately, the newly formed particles can grow to sizes in the CCN range by coagulation and condensation of more vapors¹³. The occurrence of NPF events has been observed to take place almost everywhere in the atmosphere¹⁴. The total particle number concentration (N_{tot}) in regional background conditions, as well as in the global troposphere, is very likely to be dominated by NPF events¹⁵. The contribution of aerosols formed via NPF to overall CCN concentration is still poorly understood (e.g.¹⁶). Many modeling studies suggest that NPF events can have an impact on the abundance of the global CCN (e.g.^{11,17–23}) but the magnitude of the NPF contribution to CCN differs substantially in different models and for different environments^{11,18,21}.

¹University of Colorado, CIRES, Boulder, 80309, USA. ²NOAA, Global Monitoring Laboratory, Boulder, 80305, USA. ³Andalusian Institute for Earth System Research, IISTA-CEAMA, University of Granada, Junta de Andalucía, Granada, 18006, Spain. ⁴Department of Applied Physics, University of Granada, Granada, 18071, Spain. ⁵Department of Atmospheric Sciences, University of Utah, Salt Lake City, UT, 84112, USA. ⁶These authors contributed equally: Elisabeth Andrews, Inés Zabala. ✉e-mail: betsy.andrews@colorado.edu; gtitos@ugr.es

Combining information about aerosol microphysics (number concentration and size distribution) and aerosol composition with information about the CCN activity of those aerosols can help to improve our understanding of the particles that can form clouds. Undertaking such a harmonization across a variety of locations and aerosol types has the potential to reduce uncertainty in ACI for different environmental conditions. While satellite retrievals are important for understanding ACI processes across different scales, the retrievals are currently only capable of assessing highly uncertain proxies for CCN activity for limited conditions²⁴. Those CCN proxies require evaluation with *in-situ* observations of CCN across different environments in order to assess when and whether they provide accurate enough estimates of CCN to be used in ACI studies^{24,25}. In an enormous effort, Schmale *et al.*^{26,27} provided a first overview of co-located Particle number size distribution (PNSD) and CCN measurements at 11 observatories (8 in Europe, 2 in Asia and 1 in USA), based on 1-h time resolution data. Today, additional collocated measurements of CCN and PNSD exist which can enhance global coverage. In particular, the Atmospheric Radiation Measurement program of the US Department of Energy (<https://www.arm.gov/>) has extensively performed these measurements both as a long-term monitoring effort as well as on a short-term/campaign basis²⁸.

Here we present a harmonized data set of US Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) measurements that includes surface CCN concentrations (N_{CCN}) as a function of supersaturation (SS) and PNSD at 10 sites along with collocated, complementary particle chemical composition and aerosol optical properties where available. We provide both high resolution (5-min averages) and hourly averaged data for these measurements. The data averaging was done as part of our processing (see data handling section below). The high resolution dataset is useful for NPF related investigations while the hourly averaged data is comparable to the Schmale *et al.*²⁶ dataset and is useful for applications such as global model evaluation²⁹. We also provide CCN-related data products including the hygroscopicity parameter (κ_{CCN} ³⁰), critical diameter (D_{crit}) and activated fraction (AF), derived by combining measurements from different instruments at each site.

Methods

In the methods section, we first describe the overall study design and provide a general overview of the individual sites along with the associated campaign(s) if applicable. There is then a section that briefly describes the DOE/ARM data archive along with individual sections for each measurement type. Each measurement type section includes the original DOE/ARM citation(s) for the measurement and a description of the operating principles of the instrument(s) used to make the measurement. The data handling section presents information on how we processed the original ARM files to generate the final harmonized data files presented here. Following that, we describe the methodology we used to calculate some CCN parameters using the finalized quality-assured data sets. Finally, in the Technical Validation section, we discuss the closure tests and other data checks we used to assess the quality of the resulting harmonized data sets.

Study design. The benchmark data set presented here consists of multiple collocated measurements from various DOE/ARM campaigns or long-term sites which are relevant for improving our current knowledge on CCN spatio-temporal variability and CCN sources and for exploring (a) the relationship between NPF and CCN and the factors impacting that relationship and (b) aerosol-cloud interaction studies and measurement-model comparisons, among others. Specifically, for each site we include CCN and PNSD data as well as an independent measure of total particle number concentration. These data, in their primary form, are already available from DOE/ARM's data archive (<https://arm.gov/data/>) but here we have put them under further review, harmonized the data (as described below), calculated related data products and provided them together so the next researcher may find the data more easily and in a consistent format. Many of the sites measured other relevant parameters (i.e., aerosol composition and aerosol optical properties) so those data, where available, are included in this harmonized data set.

Sites. All sites were DOE/ARM deployments operating either on a campaign basis (e.g., via the ARM Mobile Facility, AMF) or as part of ARM's long-term monitoring efforts. Table 1 provides general site information, while Figs. 1, 2 show a map of site locations and the instrument operating periods at each site, respectively. As a prerequisite for being considered in this study, only data from sites where collocated and contemporaneous sub-micron PNSD and CCN measurements were available in the ARM archive are included in the harmonized data set. We also required the CCN measurements to scan multiple SS values between 0.1 and 1. One site initially considered (Cape Cod, MA during the Two-Column Aerosol Project (TCAP)³¹) met both these criteria but was eliminated due to irregularities with CCN supersaturation scanning. We also considered data from two other DOE/ARM deployments in Houston, Texas³² and San Diego, California³³; unfortunately these sites needed to be removed from the final dataset after DOE/ARM identified a leak in the CCN sample line. Figure 1 shows that the measurements occurred primarily in the western hemisphere (Europe and the Americas) and, combined with the coverage of similar measurements described in Schmale *et al.*²⁷, highlights the need for these types of measurements across Asia and the global south.

The PNSD and CCN data meeting these criteria were available from 10 stations. All of these stations provided additional aerosol chemical composition and/or optical measurements. Measurements at most sites were performed following standard operational procedures established by DOE/ARM, making the measurements among sites comparable and the data formats consistent. However, at some sites some of the instruments were operated by external DOE-funded principal investigators (PIs) to support the science needs during specific campaigns. The format of the PI data sets are different than the DOE/ARM operational data sets and are found in DOE/ARM's intensive operating procedure (IOP) archive. Individual characteristics of each site are briefly described

Site ID	Location (Campaign/site name)	Coordinates	Time period ¹	Site type	Site reference
ANX	Andenes, Norway (COMBLE)	69.14 N 15.68 E, 2 masl ²	Dec 2019-Apr 2020	Polar/Marine	34
ASI	Ascension Island (LASIC)	7.97 S 14.35 W, 341 masl	Jun 2016-Oct 2017	Marine	35
COR	Cordoba, Argentina (CACTI)	32.13 S 64.73 W, 1141 masl	Oct 2018-Apr 2019	Continental	36,37
ENA	Azores, Portugal (Eastern North Atlantic)	39.09 N 28.02 W, 30 masl	Nov 2022-Dec 2023	Marine	38,106
GUC	Colorado, USA (SAIL)	38.90 N 106.94 W, 3137 masl	Nov 2021-Jun 2023	Continental/Mountain	39
MAO	Manacapuro, Brazil (GoAMAZON)	3.21 S 60.60 W, 50 masl	Jan 2014-May 2015	Continental/Urban	107
MOS	Arctic Ocean ³ (MOSAiC)	85 N 15 E, 12 masl	Oct 2019-Oct 2020	Polar/Marine	48
SBS-CP ⁴	Colorado, USA (STORMVEX)	40.45 N 106.79 W, 2438 masl	Oct 2010-Apr 2011	Continental/Mountain	49
SBS-SPL ⁵	Colorado, USA (STORMVEX)	40.45 N 106.73 W, 3220 masl	Dec 2010-Mar 2011	Continental/Mountain	49
SGP	Oklahoma, USA (Southern Great Plains)	36.61 N 97.49 W, 318 masl	Apr 2017-Aug 2023	Continental	108

Table 1. General site information. ¹Time range indicates period of CCN and PNSD overlap, but all measurements may not span that time range. ²masl = meters above sea level. ³MOSAiC moved over course of year as ship moved with ice - latitude and longitude provided represent midpoint of shiptrack. ⁴CP = Christy Peak. ⁵SPL = Storm Peak Laboratory.

below and in more detail in the provided references if available. The sites are alphabetically ordered by their three letter identification.

ANX, Andenes, Norway – COMBLE. The Cold-Air Outbreaks in the Marine Boundary Layer Experiment (COMBLE) was an AMF deployment designed to study boundary layer convection and air mass transformation during Arctic cold air outbreaks over a 6 month period (Dec. 2019 - May 2020)³⁴. The ANX sampling site was located on the coast of northern Norway.

ASI, Ascension Island, British Overseas Territories – LASIC. The overarching goal of the Layered Atlantic Smoke Interactions with Clouds (LASIC) campaign was to investigate absorbing aerosol and their interactions with clouds. The AMF was installed on Ascension Island and made measurements between June 2016 - October 2017³⁵. The AMF was deployed approximately 1.5 km inland on the windward side of the island, in the path of the biomass burning plume that typically flows off the African continent in the March-May timeframe.

COR, Córdoba, Argentina – CACTI. The Cloud Aerosol and Complex Terrain Interactions (CACTI) campaign was a 7 month deployment of the AMF in Argentina from September 2018 - April 2019. CACTI was designed to investigate cloud life cycle in complex terrain in order to improve parameterizations of cloud and aerosol processes in models³⁶. Measurements were made both at the surface and in an instrumented airplane, but we focus on the surface measurements. The surface sampling site was located on the eastern side of the Sierras de Córdoba mountain range and likely sampled biogenic, biomass burning and/or anthropogenic aerosol depending on the synoptic patterns³⁷.

ENA, Azores, Portugal - Eastern North Atlantic. Measurements at the ENA site on Graciosa Island began as one of the AMF deployments in 2009–2010, however our focus is on more recent data after ENA became one of DOE/ARM's long-term atmospheric observatories. The data used covers the time period between November 2022 - December 2023. The aerosol instrumentation is located on the north side of the island approximately 1 km inland³⁸.

GUC, Colorado, USA – SAIL. The Surface Atmosphere Integrated field Laboratory (SAIL) was an AMF deployment near Crested Butte, Colorado. The campaign went from September 2021 - June 2023 and was designed to investigate the water cycle and energy budget in complex terrain³⁹. The aerosol measurements were originally at the campaign's main facility (M1) but were moved about 6 weeks into the campaign to the S2 location. In this study, we focus on data from the S2 location which covers the time period October 2021 - June 2023. The S2 site was located at the Crested Butte ski resort in the Rocky Mountains.

MAO, Manacapuro, Brazil – GOAMAZON. The Observations and Modeling of the Green Ocean Amazon (GOAMAZON) was a 23 months deployment (2014–2015) in Brazil to study aerosol and cloud life cycles in the Amazon basin⁴⁰. The AMF and a new DOE/ARM mobile aerosol observing system (MAOS) were both deployed to this site with facility designators M1 and S1, respectively. These two facilities made some duplicate measurements (e.g., optical properties, N_{tot}). We utilized the data from the S1 facility for all instruments except the CCN counter (CCNC) for which there was only data available from the M1 facility. Both facilities were located in a pasture 70 km downwind of the city of Manaus and sampled both the urban plume and regional aerosol from the

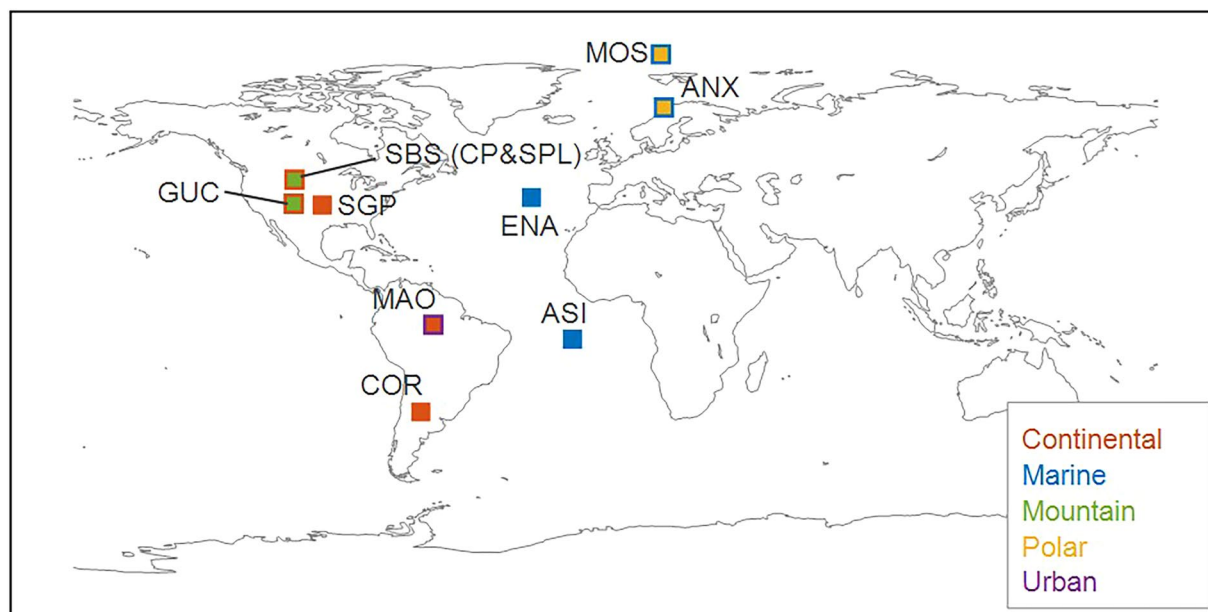


Fig. 1 Map of DOE/ARM sites with collocated simultaneous CCN and SMPS data. Site type is indicated with different colors; if the outline is different than the fill color the site could be described by more than one type (e.g., polar and marine). MOS is a mobile deployment so the location shown on the map is approximate. See Tables 1, 2 for more details.

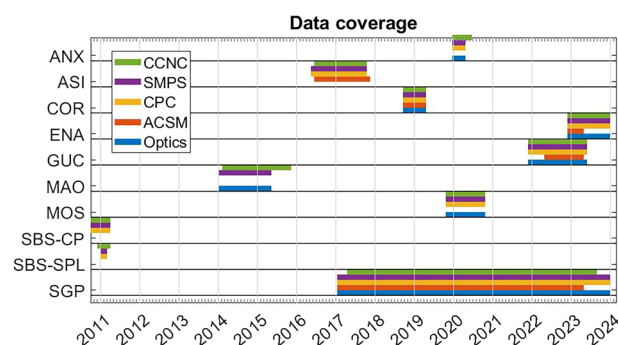


Fig. 2 Data coverage of the instruments used in this study for all the sites considered. The main objective is the analysis of CCNC and SMPS measurements so, although there may be longer CPC, ACSM and optics measurements, these have not been taken into account.

surrounding rainforest. A large amount of NPF and CCN analysis from this campaign and IOP measurements has already been published^{41–47}.

MOS, Arctic Ocean – MOSAiC. The Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC) campaign was a deployment of the AMF on an icebreaker frozen into and moving with the ice. The goal was to collect data to better understand climate change in the Arctic and to characterize the cryosphere-atmosphere relationship in the region⁴⁸. The 1 year campaign began in October 2019.

SBS, Colorado, USA – STORMVEX. Storm Peak Lab Cloud Property Validation Experiment measured cloud and aerosol properties in complex terrain at multiple sites over 6 months (Nov. 2010–Apr. 2011)⁴⁹. One focus of the campaign was to investigate the role of different types of aerosols on cloud and precipitation processes. There were two sites with CCN and PNSD measurements on the mountain at the Steamboat Springs ski resort, one at Christy Peak (CP) and one at the University of Utah's Storm Peak Laboratory (SPL) (https://atmos.utah.edu/storm_peak_lab). We refer to these two sites as SBS-CP and SBS-SPL, respectively. PI datasets for raw CCNC and condensation particle counter (CPC) from SPL and raw SMPS data from SPL and CP are in the DOE/ARM IOP archive.

Site ID	SMPS	CCN	CPC	ACSM	Neph	PSAP
ANX	TSI 3938	DMT2C	TSI 3772	n/u ¹	TSI 3563	RR 3W
ASI	TSI 3936	DMT2C	TSI 3772	Q-ACSM	n/u ²	n/u ²
COR	TSI 3936	DMT2C	TSI 3772	Q-ACSM	TSI 3563	RR 3W
ENA	TSI 3938	DMT2C	TSI 3772	Q-ACSM	TSI 3563	RR 3W
GUC	TSI 3938	DMT2C	TSI 3772	Q-ACSM	TSI 3563	RR 3W
MAO	TSI 3936	DMT1C	n/u ³	n/u ⁴	TSI 3563	RR 3W
MOS	TSI 3938	DMT2C	TSI 3772	n/a	TSI 3563	RR 3W
SBS-CP	TSI 3936	DMT1C	TSI 3772	n/a	n/u ⁵	n/u ⁶
SBS-SPL	TSI 3936	DMT1C	TSI 3010	n/a	n/a	n/a
SGP	TSI 3938	DMT2C	TSI 3772	Q-ACSM	TSI 3563	RR 3W

Table 2. Summary of site specific instrument information for measurement types included in harmonized data set. n/a indicates data not available, n/u indicates data not used due to quality issues: ¹No valid data; ²Sample RH >40%; ³DQR indicates nozzle clog; ⁴DQR marks data as suspect - recommend using AMS from 2 shorter IOP deployments; ⁵Time series analysis indicated unrealistic spectral and angular dependencies of scattering; ⁶PSAP requires nephelometer for standard corrections.

SGP, Oklahoma, USA - Southern Great Plains. This is the longest set of measurements and the most complete as it is where DOE/ARM tests instruments before deploying further afield. For comparability reasons, at SGP we limited the analysis to the operational measurements from 2017 to 2023 at central facility 'E13'. Note however that there are CCN and PNSD data from the SGP C1 facility prior to this, e.g.^{50,51}, albeit in different format data files and with a more limited auxiliary data set. The central facility site is located in a rural agricultural region more than 100 km from any large cities (e.g., Wichita, Oklahoma City, Tulsa).

DOE/ARM Datastreams and measurements. In this section we first briefly provide some details about the DOE/ARM data archive files. These files were the starting point for our evaluation and the further processing that we did to generate the final harmonized dataset. We then describe the individual instruments used for each measurement included in the harmonized data set; these instrument descriptions include information on the data processing performed by DOE/ARM. For some instruments we note where we had to replicate the DOE/ARM processing in order to have a consistent starting part for further data treatment. The additional data processing and evaluation that we performed (beyond that performed by DOE/ARM) to generate the harmonized data set is described in the data handling and data validation sections below. Table 2 summarizes the information about specific instruments used for each parameter on a site by site basis.

DOE/ARM Datastreams. The starting page for finding and downloading data from the DOE/ARM data archive is: <https://arm.gov/data/>. The DOE/ARM data archive stores both 'operational' and principal investigator (PI) data files from ARM field campaigns and permanent measurement locations. The operational data files are consistent across the sites in terms of data acquisition, QA and formatting. These operational files became available in the 2014–2015 time frame. Prior to that the DOE/ARM data sets were collected using data acquisition, processing and formatting procedures, often by entities external from DOE/ARM.

DOE/ARM uses a consistent three component naming convention for their operational data files. Each file name starts with a three letter identifier representing the site (*stn*). The site identifier is followed by the name of the datastream and then a second location identifier indicating the specific facility/building (*fac*) at the site where the measurement was made. The second part of the file name indicates data processing level and the third part of the file name is a date string which is the start time of the data contained in the file. For example, the file name *sgpaosnephdry1mE13.b1.20221213.000030.nc* represents data from the SGP site at the E13 facility. The datastream is for the aerosol observing system (AOS) dry nephelometer measurements averaged to 1 min. The data level is 'b1' which indicates the data have been quality controlled, reviewed and standard corrections have been applied. The data are for the day of 13 December 2022. To find the original DOE/ARM data files used in this study we recommend searching by datastream name (e.g., in this example, *aosnephdry1m* on the DOE/ARM data archive (<https://arm.gov/data/>)). Table 3 provides the names of the operational DOE/ARM datastreams and data citations for the files used to generate the harmonized dataset. More information on DOE/ARM's file naming and data levels are available here: <https://www.arm.gov/guidance/datause/formatting-and-file-naming-protocols>.

In addition to the actual data files for each operational datastream, a text file containing data quality reports (DQRs) accompanies downloaded data. These DQRs are generated by each DOE/ARM instrument mentor (the person responsible for reviewing the data) when the mentor identifies a problem with the data (e.g., that it's missing, that it's invalid due to an instrument issue or that it's potentially suspect). Users of DOE/ARM data need to incorporate review of the DQRs into their analysis to ensure they are using the best possible data. Typically DOE/ARM automatically sends additional DQR information to people who have downloaded the data. Based on our experience with the Houston and San Diego campaign data, where that did not happen, we recommend doing a final check of data availability/quality before publishing results using DOE/ARM data.

Additionally, as noted above, the ARM data archive includes links to measurement data generated by DOE-funded PIs to support the science needs at individual sites. Such data sets utilize PI-specific file naming conventions, processing, and formats and are found on the IOP section of the ARM data archive. Where

Measurement type	Operational filenames
Size distribution	[stn]aossmps[fac].b1.[timestamp].nc ¹⁰⁹
N_{CCN} (1 column)	[stn]aosccn1col[fac]**.[timestamp].nc ¹¹⁰ [stn]aosccn100[fac].a1.[timestamp].nc ¹
N_{CCN} (2 column)	[stn]aosccn2cola[fac].b1.[timestamp].nc ¹¹¹ [stn]aosccn2colb[fac].b1.[timestamp].nc ¹¹² [stn]aos2colaspectra.b1.[timestamp].nc ¹¹³
$N_{tot,CPC}$	[stn]aoscpf[fac]**.[timestamp].nc ⁵⁶
Chemistry	[stn]acsmcde[fac].c2.[timestamp].nc ⁵⁸
Scattering	[stn]aosnephdry1m[fac].b1.[timestamp].nc ¹¹⁴
Absorption	[stn]aoppsap1flynn1m[fac].c1.[timestamp].nc ¹¹⁵

Table 3. List of DOE/ARM filenames for operational data sets, with links to data citation. [stn] indicates a three letter identifier representing the site and [fac] indicates the facility at the site. **Indicates the data stream data level varied. For the CPC and CCNC measurements we used the level b1 data stream. ¹For SBS-CP CCNC only we used this a1 data file as that was all that was available in the DOE/ARM archive.

Measurement type	Site	IOP archive directories for data in PI format (on https://iop.archive.arm.gov/)
PNSD	SBS	/arm-iop/2010/sbs/stormvex/Christy/hallar-smmps/ /arm-iop/2010/sbs/stormvex/StormPeak/hallar-smmps/
N_{CCN}	SBS	/arm-iop/2010/sbs/stormvex/StormPeak/hallar-ccn/
$N_{tot,CPC}$	SBS	/arm-iop/2010/sbs/stormvex/StormPeak/hallar-cpc_3010/

Table 4. List of IOP archive locations for PI formatted files.

applicable, Table 4 lists the DOE/ARM IOP data archive locations for PI datastreams. PI data does not come with DQRs, although there is often a README file.

Particle number size distribution. The submicron PNSD is measured using a Scanning Mobility Particle Sizer (SMPS) spectrometer. Table 2 lists the various PNSD instruments that operated at each site. The working principle of the instrument is based on the counting of particles of different sizes that are selected based on their electrical mobility. The instrument consists of a bipolar diffusion charger (neutralizer), a differential mobility analyzer (DMA), and a CPC in series. The bipolar diffusion charger brings the particles into a near bipolar charge equilibrium. By scanning or stepping the voltage in the DMA, particles of certain electrical mobility are selected, and counted with the CPC (the CPC is often referred to as the detector). Depending on configuration the instrument has the capability of scanning from 10 - 1000 nm, although that full size range was not employed in the data sets here, and it scans 64 bins per diameter decade. Each full scan takes the SMPS 5 minutes to complete. Humidity in the sample line is maintained below 40% using a Nafion™ dryer. Data acquisition, instrument control and data inversion are accomplished through the manufacturer's software (Aerosol Instrument Manager, TSI).

The SMPS system undergoes calibration prior to deployment, including DMA sheath flow, voltage control, and particle sizing verification. DMA particle sizing is verified with polystyrene latex particle size standards (150 nm). The CPC characterization includes verifying the inlet flow rate and determining the size-dependent particle counting efficiency. The DOE/ARM SMPS handbook describes the SMPS and operations in more detail⁵².

The two SMPS data sets from the STORMVEX campaign at SBS were not from the DOE/ARM operational measurements but rather were from an individual PI (see Table 4) so more effort was required to identify, process and QA the data because the formats and processing differed from those of the 'operational' instruments. As described in the data handling section, similar evaluation and validation analyses were applied to the PI PNSD data as were used for the DOE/ARM operational PNSD measurements.

Cloud condensation nuclei concentration. The N_{CCN} as a function of SS is measured with a CCNC. There are two modes of operation for the CCNC: monodisperse operation mode in which particles are size selected by means of a DMA prior to entering the CCNC (providing size-resolved CCN concentrations) and polydisperse mode in which activation is measured for the whole aerosol population. In this study, all CCN measurements were performed in the polydisperse operation mode. Over the years DOE/ARM has operated two different versions of the Droplet Measurement Technology (DMT) CCNC. Table 2 lists the version of the CCNC that operated at each site. The DMT1C (i.e., model DMT100), a single column CCNC, typically scanned across 4–6 SS values over the course of an hour providing N_{CCN} at SS = 0.1% on the low end to approximately SS = 1% on the high end. More recently, the DMT2C (i.e., model DMT200), a dual column CCNC, has been deployed. For DOE/ARM deployments of the DMT2C, one column 'column A' typically scans across 4–6 SS values over the course of an hour and the second column 'column B' measures at constant SS (typically 0.4%). In our data handling the single column of the DMT1C is considered column A. For both the DMT1C and DMT2C over the course of an SS scan, the CCNC typically spends about 10 min at a given SS (6.5 minutes of measurement and 3.5 minutes for transition between SS). The DOE/ARM CCNC handbook describes the CCNC and operations in more detail⁵³. Below we describe generally how the CCNC works which is same for both column A and column B so we do

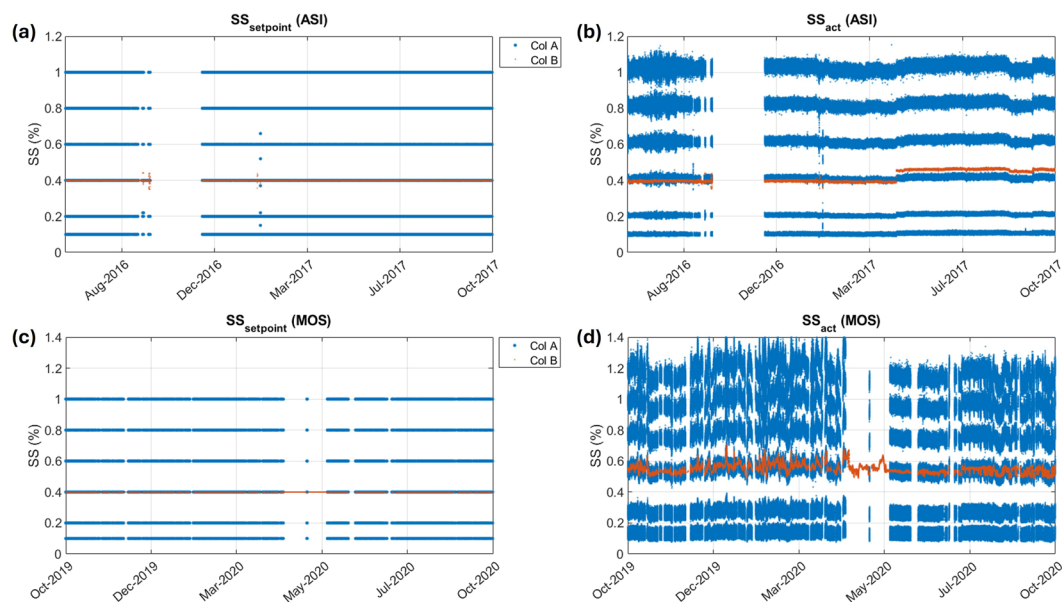


Fig. 3 CCNC supersaturation time series for ASI (a,b) and MOS (c,d). Plots a) and c) show the SS_{setpoint} time series for columns A (blue dots) and B (orange dots) of the CCNC at both sites. Plots b) and d) display the SS_{act} time series for columns A and B.

not specify column. We do however specify the column in the data handling section where there might be some confusion and in the harmonized data files.

Briefly, the operating principle of the CCNC is based on a cylindrical continuous-flow thermal-gradient diffusion chamber where constant temperature gradients are applied, generating different SS conditions⁵⁴. Inside the instrument, the aerosol sample flow is guided through the center of the cylinder by a particle-free laminar sheath flow. The centerline SS is generated by applying a controlled (and constant) streamwise temperature gradient at the cylinder wall; if the inner wall is kept wet, heat and water vapor continuously diffuse towards the center of the tube. Because water vapor has a lower molecular weight than moist air, diffusion of water vapor is faster than heat and the centerline becomes supersaturated⁵⁴. Particles that activate at a critical SS lower than the measurement SS form droplets. At all sites, the CCNC instruments were operated at similar conditions, with an aerosol to sheath flow ratio of 10 as recommended by the manufacturer (volumetric aerosol flow $45 \text{ cm}^3 \text{ min}^{-1}$, volumetric sheath flow $450 \text{ cm}^3 \text{ min}^{-1}$).

Operationally, the user chooses a specific SS value or values for the CCNC to measure at and also sets the timing of the measurements if the instrument is scanning across multiple SS. These specified SS values are referred to as the setpoint SS (SS_{setpoint}). Ideally, SS_{setpoint} would correspond to the actual supersaturation at the centerline of the CCNC column but in reality this is not the case. Actual SS (SS_{act}) in the CCNC depends on the instrument calibration and operating conditions (e.g., flows and temperature gradient) and can differ significantly from SS_{setpoint} ⁵⁵. Figures 3, 4a,b show the instrument SS_{setpoint} and SS_{act} for the measurements at ASI, MOS and GUC, as examples. While SS_{act} can be similar to the SS_{setpoint} there is a fair amount of variability in SS_{act} for each scan and it's also clear from these figures that the relationship between SS_{setpoint} and SS_{act} can change with time. At most sites the percent difference between SS_{setpoint} and the SS_{act} was less than 20%. The largest difference was observed for column B at MOS (Fig. 3c,d), where differences of up to 40% between the instrument $SS_{\text{setpoint},B}$ and the $SS_{\text{act},B}$ were observed, likely due to variations in column temperature gradient related to instrument operating conditions or calibration issues. The difference between SS_{setpoint} and SS_{act} varied among sites and also across time for some sites. Several supersaturation values are included in the harmonized data files (see Data Handling and Data Records sections for definition of the SS provided and the description of the harmonized data files).

Ideally CCNC calibration occurs before and after the deployment and/or approximately annually. Calibration of the CCNC includes flow and pressure calibrations and procedures to determine the thermal characteristics of the instrument. The flow, pressure and thermal characteristics are then used to derive the SS_{act} ⁵⁵. All DOE/ARM level b1 operational CCN data sets (i.e., *aosccn2cola.b1*, *aosccn2colb.b1*, and *aos2colaspectra.b1*) used in this study provide the SS_{act} , which takes into account the thermal characteristics of the column⁵⁵ and the instrument calibrations. In contrast, the two SBS sites, CP and SPL, report N_{CCN} referenced to the instrument SS_{setpoint} rather than the SS_{act} . SBS-CP only had the data level a1 datastream available which does not include SS_{act} and the CCN dataset from SBS-SPL was sourced from a PI rather than DOE/ARM operational measurements (see Table 4). These data were processed, quality-assured and put into the same harmonized format as we used for the DOE/ARM operational CCNC data, albeit with some assumptions related to SS values (see data handling section for specific details).

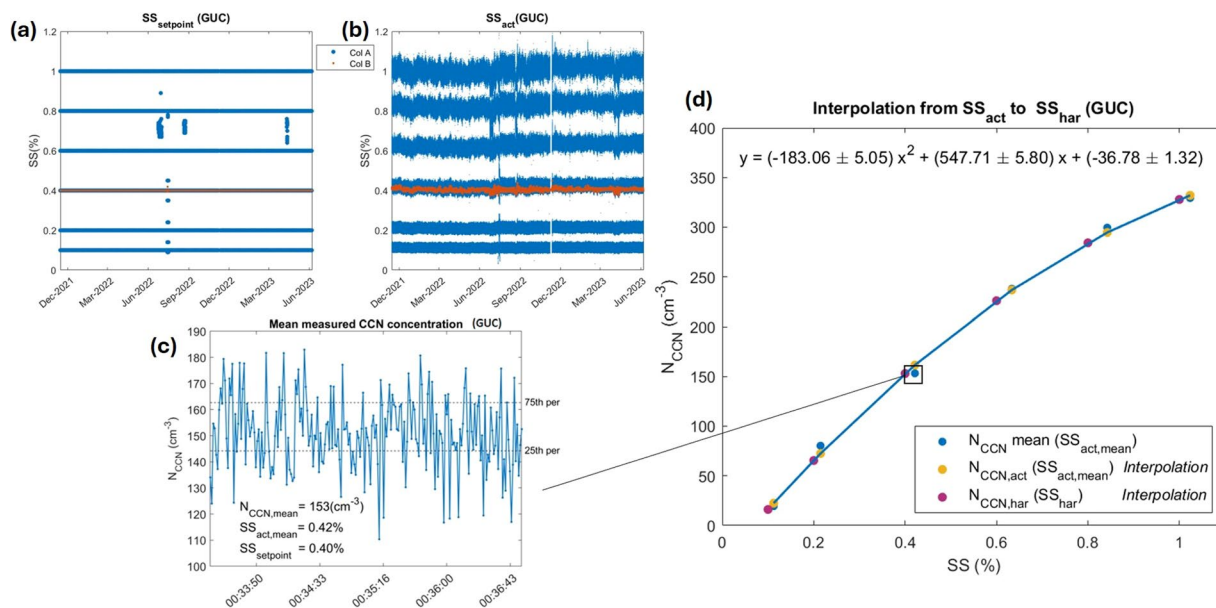


Fig. 4 (a) Time series for $SS_{\text{setpoint},A}$ (blue dots) and $SS_{\text{setpoint},B}$ (orange dots); (b) Time series for $SS_{\text{act},A}$ and $SS_{\text{act},B}$; (c) example time series of $N_{\text{CCN},A}$ for one $SS_{\text{setpoint},A}$ with horizontal dashed lines indicating the 25th and the 75th percentiles of $N_{\text{CCN},A}$. (d) Example of CCN interpolation for one column A SS scan. The blue dots are the mean measured $N_{\text{CCN},A}$ between the 25th and 75th percentile at the mean $SS_{\text{act},A}$ (Fig. (c)), the blue line is the two degree polynomial fit line based on the blue dots, the yellow dots are the interpolated (i.e., on the fit line) $N_{\text{CCN},A}$ at the mean $SS_{\text{act},A}$ and the pink dots are the $N_{\text{CCN},\text{har}}$ values corresponding to each SS_{har} value. All figures are for CCNC measurements at GUC.

Particle number concentration. A CPC measures the total number concentration ($N_{\text{tot},\text{CPC}}$) of particles above a certain minimum particle size. Table 2 lists the model of CPC (TSI 3010 or TSI 3772) that operated at each site. The manufacturer's specifications for these two CPC models indicate that they have the same minimum size detection limit of 10 nm.

The general principle of a CPC instrument is that it enables the operating fluid (typically butanol) to condense onto particles in the sample flow, creating aerosol droplets large enough to be detected efficiently using a light-scattering technique. Upon entering the CPC instrument, the sample air stream passes through a saturation block where the operating fluid evaporates into the sample stream, saturating the flow with vapor. The sample air stream then passes into a condenser tube cooled by thermoelectric coolers. Here, the operating fluid vapor supersaturates and condenses onto the particles. The minimum diameter of the original particles that will grow into droplets depends on the temperature difference between the saturator and the condenser. Droplets leaving the condenser tube pass one at a time through a single particle counting optical detector. Although the CPC may suffer counting issues at extremely high concentration ($>10^4$ for the TSI 3010 and 3772 models), these data have not been removed or flagged in the database. The CPC datastream is dried to low RH prior to entering the instrument. The DOE/ARM CPC handbook describes the CPC and operations in more detail⁵⁶.

The CPC dataset from SBS-SPL was sourced from a PI rather than DOE/ARM operational measurements archive (see Table 5). The same processing and quality assurance criteria as were applied to the operational CPC instruments were applied to the SBS-SPL CPC.

Sub-micron aerosol chemical composition. Aerosol chemical composition was measured by a Quadrupole Aerosol Chemical Speciation Monitor (Q-ACSM, hereafter referred to as ACSM) at many of the sites contributing to the harmonized data set. The ACSM measures sub-micron particle mass and chemical composition in real time. Specifically, it measures mass concentrations of particulate organics, sulfate, nitrate, ammonium, and chloride (M_{organics} , M_{sulfate} , M_{nitrate} , M_{ammonium} , and M_{chloride}) from 40 nm to 1 μm in aerosol aerodynamic diameter and a mass range (m/z) of 10 to 200 amu. DOE/ARM operations and protocols for the instrument are found in the ACSM instrument handbooks^{57,58}.

The main ACSM inlet system has a flow rate of 3 lpm from which a sample flow rate of approximately 0.1 lpm is picked-off to create near isokinetic sampling conditions. There is a Nafion dryer upstream of the ACSM at each site to keep the sample RH lower than 30%. The ACSM sample flow is focused by an aerodynamic lens, creating a 1 mm diameter aerosol stream that impacts a hot (600 °C) vaporizer. The vaporized aerosol is then ionized to produce the ions which are analyzed by the quadrupole mass spectrometer. The ACSM uses a naphthalene source for reference during calibrations and to determine an effective ion transmission efficiency through the quadrupole mass spectrometer. This ion source operates continuously resulting in background mass that must be subtracted from the particle mass spectra to accurately determine the particle mass composition.

The ACSM does not detect all particles in the air stream but rather a fraction of aerosol particles. This fraction is defined by a parameter known as the collection efficiency (CE) that should be applied to account for this

Instrument type	Uncertainty	comments	source
SMPS	10–20%	depends on flow rate and charging efficiency	80,109,116
CCNC	5–40% SS, 5% N_{cen}	depends on flow, SS calibration and Kohler model assumed	26,117
CPC	10–14%	depends on flow, T and number concentration	80
ACSM	9–10% mass, 10–40% species concentration	depends on species	26,80,118
Nephelometer	10%	depends on particle size, loading and averaging period	63,81
PSAP	20–100%	depends on loading, averaging and aerosol single scattering albedo	64,119

Table 5. Uncertainties for different measurements based on literature and DOE/ARM instrument and uncertainty handbooks.

incomplete detection. A primary cause of a CE less than 1 is the particles bouncing off the vaporizer impaction surface prior to vaporization and detection⁵⁹.

One approach to deal with this undersampling is to assume a constant CE, e.g., CE = 0.5 to indicate only half of the particles are detected, e.g.⁶⁰.

More recently composition-dependent collection efficiencies (CDCE) have been used to express CE as function of chemical composition. CDCE helps account for the phase properties of ambient aerosol particles. Particles that exhibit more liquid-like properties tend to stick onto the vaporizer resulting in a CE closer to 1.0, whereas solid particles bounce off the vaporizer resulting in lower CE values⁶¹. For the DOE/ARM ACSM operational data, DOE/ARM data level c files (datastream *acsmcdce*) apply a CDCE parameterization^{58,61}, while the level b files (datastream *aosacsm*) assume a constant CE of 1.0. For most sites (ASI, COR, ENA, GUC, and SGP) the level c files were available. For MAO, only ACSM level b data files were available. We derived CDCE values for this level b data using the CDCE methodology^{58,61}, and checked our processing against another site which had both level b and level c data.

Aerosol optical properties. Measurements of aerosol scattering and absorption coefficients were made at most sites included in this harmonization project. The aerosol scattering coefficients were measured by an integrating nephelometer, specifically a TSI 3563 integrating nephelometer which measures aerosol scattering (σ_{sp}) and backscattering coefficients (σ_{bsp}) at three wavelengths (450, 550, and 700 nm). The nephelometer measures aerosol scattering by detecting light scattered by the aerosol and subtracting light scattered by the gas, the walls of the instrument and the background noise in the detector. Due to instrument angular limitations the TSI nephelometer only measures scattering between 7° and 170°, requiring the so-called truncation correction to account for the angles not measured. Anderson *et al.*⁶² and Anderson and Ogren⁶³ provide an overview of the nephelometer and the truncation correction, respectively. In the DOE/ARM operational files (*aosnephdry1m*) the nephelometer data are reported at standard temperature and pressure (STP, $T_{std} = 0^\circ\text{C}$ and $P_{std} = 1013\text{ hPa}$) and have already been corrected for angular truncation and lamp non-idealities⁶³.

The aerosol absorption coefficients (σ_{ap}) were obtained from the three wavelength (464, 529, and 648 nm) Radiance Research (RR) Particle Soot Absorption Photometer (PSAP⁶⁴). This is a filter-based instrument which relates the decrease in light transmission through a filter due to particle loading to the aerosol absorption coefficient. Bond *et al.*⁶⁴ note that several corrections are required to transform the change in transmittance to an σ_{ap} value, including accounting for the aerosol scattering artifact, properly measuring the size of the filter spot, and a calibrated flow rate. Virkkula *et al.*⁶⁵ additionally account for the single scattering albedo of the sampled aerosol. We utilized the DOE/ARM developed, value added product datastream *[stn/aoppsap1flynn1m[fac].c1.[timestamp].nc]*⁶⁶ which is produced operationally. These files contain three different versions of corrected PSAP data (i) average of Virkkula^{65,67} and Bond/Ogren^{64,68} correction; (ii) Virkkula correction alone, (iii) Bond/Ogren correction alone. Note that DOE/ARM recommends the average of the two types of corrections “due to beneficial cancellation of offsetting correction artifacts” so that is the value included in the harmonized data set⁶⁶. The DOE/ARM nephelometer and PSAP handbooks describe these two instruments and their operations in more detail^{69,70}.

At all sites studied here, both the nephelometer and PSAP were downstream of a switched impactor system. Over the course of an hour the size cut for the σ_{sp} and σ_{ap} measurements switched between 10 μm and a 1 μm in diameter, referred to here as the PM₁₀ and PM₁ size cut, respectively. This enables quantification of the contribution of coarse mode aerosol to σ_{sp} and σ_{ap} , complementing information derived from the spectral dependence of σ_{sp} and σ_{ap} . The timing of the switching between PM₁₀ and PM₁ was designed to sample both size cuts over the course of an hour, though not necessarily for the same amount of time. For example, a typical repeating three hour pattern was the first hour measured 7 min on PM₁₀ and 52 min on PM₁, the second hour measured 7 min on PM₁₀, 45 min on PM₁ and then 7 more min on PM₁₀ and the third hour measured 7 min on PM₁ and 52 min on PM₁₀. Despite the datastream name (*aosnephdry1m*), the nephelometer sample stream was not actively dried for many of these deployments. This resulted in a sample RH for the scattering measurements that varied significantly depending on the both ambient temperature and RH conditions and the temperature conditions inside the measurement trailer and instrument. RH conditions were part of our further assessment as described in the data handling section.

Data handling. To simplify sharing of original and harmonized data, all data sets from DOE/ARM were downloaded to a server at the University of Utah to which team members had access. Individual members of team were responsible for evaluating data from specific instruments. Once the data were further quality controlled and

averaged the processed files were uploaded to a different directory on the Utah server and were then available for technical validation (see next section). In what follows, we first outline the additional data processing (beyond that performed by DOE/ARM) and quality control procedures we followed in order to develop the final harmonized data. We then discuss data averaging, and finally we describe our calculation of some key CCN parameters from the harmonized CCN and PNSD data sets.

Data review. Reliability and comparability of the operational data sets presented here were assured by the application of the DOE/ARM technical standards and protocols^{52,53,56,69,70} and by further data checks we performed as described in the data handling and validation sections. The PI data sets used were unprocessed. For the SMPS and CPC PI data we performed the same processing and evaluation as was applied by DOE/ARM to the DOE/ARM operational data sets. For the CCNC PI data we had to make some assumptions as not all necessary information was available to replicate the DOE/ARM processing (see description of CCNC data handling below for more details).

The DOE/ARM operational data files come with instrument-dependent flags tied to data quality. Here we've taken the approach that if any of the DOE/ARM data flags were set for a given timestamp then the data for that timestamp were removed prior to our own QA processing and averaging. Additionally, data identified as invalid or suspect in the separate DQRs provided when the data files were downloaded were removed. There are a few exceptions to this as noted below. In the supplemental materials we have included sections for each instrument that list the standard parameters which might cause an automatic instrument flag to be set in the data file.

In addition to using the DOE/ARM provided flags we applied some additional constraints and processing. For example, all instruments are adjusted to and reported at STP in the harmonized data files, if not already reported at STP in the DOE/ARM files. Additionally, in order to ensure consistency across instruments, if the measurements were made at RH higher than 40% they were considered invalid and not included in the harmonized dataset (this only impacted some of the optical measurements). The SMPS and CCNC data required the most processing effort, so we discuss those data sets first and then describe the processing and constraints applied to the other instrument data sets.

SMPS data handling. The SMPS software allows the user to change the size range over which the instrument measures. While processing different data sets it became clear that different sites used slightly different size ranges and, sometimes, the size range setting changed over the course of time for individual sites. To harmonize the datasets we generated files with columns for the entire range of possible diameter bins for each site (113 bins, ranging from a minimum diameter of 9.14 nm to a maximum diameter of 514 nm). If the SMPS measurements were shifted to larger (or smaller) diameters and did not include particles in the smallest (or largest) size bins, those bins were filled with missing value codes. Thus, the harmonized SMPS data sets will always have the same number of columns, with each size bin column always representing a consistent diameter.

CCNC data handling. One of our end goals for the CCNC data is to generate files containing mean N_{CCN} values interpolated to the same harmonized SS (SS_{har}) for all the sites, regardless of the SS_{act} . We chose SS_{har} values of 0.1, 0.2, 0.4, 0.6, 0.8 and 1%. This harmonization required a three-step process: (1) calculating the mean N_{CCN} and mean SS_{act} for each $SS_{setpoint}$ in a scan; (2) deriving a fit to those values for each scan so that mean N_{CCN} can be interpolated to different mean SS_{act} ; (3) calculating $N_{CCN,har}$ by interpolating mean N_{CCN} to the harmonized SS (SS_{har}). Step 1 is done for both column A and column B, while steps 2 and 3 can only be done for column A. For most sites steps 1 and 2 are done by the DOE/ARM processing with the mean N_{CCN} , SS_{act} and fit parameters available in the operational data files, while step 3 is done by us to get the mean N_{CCN} values interpolated to our chosen SS_{har} . For a few sites (ANX, MAO, SBS-CP and SBS-SPL), we performed all three steps in the harmonization process as the needed files or other information was not available from DOE/ARM or the data provider. We describe these steps in more detail below.

At most sites we only used DOE/ARM's CCN spectra file (either *aos1colaspectra* or *aos2colaspectra*, depending whether it is a 1 column or 2 column instrument). The spectra file provides the mean value of $N_{CCN,A}$ between the 25th and 75th percentile (i.e., the interquartile range, IQR) for each SS and the mean value of various parameters ($SS_{setpoint,A}$, $SS_{act,A}$ and $N_{tot,CPC}$) coincident with the $N_{CCN,A}$ falling within interquartile range. Figure 4c is a graphical example of how step 1 in the harmonization process is performed. The figure shows a 6.2 minute time series of 1-sec $N_{CCN,A}$ data for one $SS_{setpoint,A}$ value during a SS scan, with the horizontal dashed lines indicating the 25th and 75th percentiles of the plotted data. The mean $N_{CCN,A}$ is the mean of the data points falling between the two dashed lines and the mean $SS_{act,A}$ is the mean of the $SS_{act,A}$ associated with $N_{CCN,A}$ values contributing to the mean $N_{CCN,A}$. The mean $SS_{act,AB}$ associated with $N_{CCN,B}$ are determined the same way, with the column B data being temporally matched with the mean $N_{CCN,A}$.

The spectra file also includes DOE/ARM's fit parameters that describe the relationship between mean $N_{CCN,A}$ values and the mean $SS_{act,A}$ values for each hourly scan across the range of SS. Figure 4 shows a fit to the mean $N_{CCN,A}$ and mean $SS_{act,A}$ for a single SS scan and provides a graphical example for steps 2 and 3 in the harmonization process. The step 2 fit is performed using a two degree polynomial (Eq. 1):

$$N_{CCN} = aSS_{act}^2 + bSS_{act} + c \quad (1)$$

where a, b, and c are the fit parameters.

The fit parameters are arrived at iteratively. For each scan, the fitted N_{CCN} values are compared against individual mean $N_{CCN,A}$ data for each $SS_{act,A}$ and, if the difference is >20% for a given $SS_{act,A}$, the corresponding mean $N_{CCN,A}$ and SS_{act} data is removed and a new fit is performed on the remaining mean $N_{CCN,A}$ and $SS_{act,A}$ data points.

This calculation is repeated until no more $N_{CCN,A}$ outliers are observed but there are still 3 or more mean $N_{CCN,A}$ and $SS_{act,A}$ values available to perform the fit. This procedure allows better data quality assessment than looking at $N_{CCN,A}$ for individual SS values because looking at the SS spectra permits identification (and exclusion) of spectra during periods when instrument conditions were unstable.

Figure 4d shows an example of the step 3 interpolation procedure for one SS scan at GUC. The blue dots correspond to the mean measured $N_{CCN,A}$ between the 25th and 75th percentile at the mean $SS_{act,A}$ (calculated in step 1). The blue line is the two degree polynomial fit line based on the blue dots (calculated in step 2). The yellow dots are the interpolated (i.e., on the fit line) $N_{CCN,A}$ at the mean $SS_{act,A}$ and are all within 20% of the blue dots. The pink dots are the $N_{CCN,A,har}$ values, i.e., the mean $N_{CCN,A}$ values interpolated to SS_{har} (interpolation done in step 3). If the interpolation led to unreasonable $N_{CCN,A,har}$ values (e.g., negative concentrations, or lower concentrations at higher SS), the whole scan is reported as invalid.

We developed code to reproduce the ARM processing (steps 1 and 2) so that we could generate harmonized CCN data for sites where the spectra file was unavailable (ANX, MAO, and both SBS sites). To check our code, we compared the ARM operational spectra files with the spectra files we generated to be sure the results were identical. For MAO and ANX we use the level b1 CCN operational file (*aosccn1col* for MAO and *aosccn2col* for ANX) which includes 1-sec resolution data of N_{CCN} concentration as well as housekeeping parameters such as SS, thermal efficiency information, temperatures, pressure and flows. For SBS-CP we use the level a *aosccn100* files. For SBS-SPL we use the 1-sec PI file. Note that the CCNC data sets for the two SBS sites only reported instrument $SS_{setpoint}$. Due to the limited information for the SBS sites (no thermal characterization) it was not possible to calculate the SS_{act} within the instrument so the harmonization procedure is performed assuming $SS_{setpoint}$ was equal to SS_{act} for those two sites. To reflect this in the harmonized files for the two SBS sites, the columns in the data file corresponding to the SS_{act} are reported as missing data (see Data Records section for harmonized datafile descriptions).

At some sites due to narrower $SS_{setpoint,A}$ ranges the SS_{har} also covered a narrower SS range to minimize interpolation errors. The site specific $SS_{setpoint,A}$ values and resulting range of SS_{har} for each site are given in Table S1. Note that the column B SS cannot be interpolated to a harmonized SS value as this column only measured at a single SS.

It should be noted that in the case of high N_{CCN} ($>5000\text{ cm}^{-3}$), the SS and droplet sizes can decrease due to higher water vapor depletion⁷¹. The water vapor depletion can result in lower N_{CCN} . In the harmonized data set, the water depletion effect has not been taken into account. At SBS-SPL, ANX and ASI there are no N_{CCN} values $>5000\text{ cm}^{-3}$. At COR, ENA, MOS, GUC, SBS-CP and EPC, the percentage of N_{CCN} measurements $>5000\text{ cm}^{-3}$ is smaller than 0.1%. For SGP, less than 0.8% of measurements exceed this threshold, while at MAO, this percentage is 2%.

It is important to highlight that, during our initial data review, several of the ARM CCN data sets were identified as suspicious and required reprocessing due to incorrect instrument calibrations being applied. The ARM CCNC instrument mentors revised and updated the datasets in the ARM data portal and those are the datasets used here.

CPC data handling. The original CPC measurements are reported at 1 second resolution. In all cases they have been converted to STP in order to be consistent with the other measurements which are also at STP. We used temperature and pressure information from the SMPS or the CCN to do the STP correction as described in the Data Records section.

ACSM data handling. The DOE/ARM ACSM data were provided at different temporal resolutions, representative of the average number of sets of sampling cycles. All operational ACSM datasets were converted to 1-h resolution for simplicity due to different averaging time periods within each dataset. The original ACSM datasets were not corrected to STP conditions and the ACSM files did not include sample and measurement temperature or pressure. To do the ACSM STP adjustment, the sample temperature and pressure measurements from the collocated SMPS were used to convert ACSM measurements to STP at each site.

Optical data handling. Optical property aerosol measurements where the sample RH was greater than 40% is reported as missing in the harmonized data set. We use the nephelometer sample RH as the RH constraint for both the nephelometer and PSAP because the PSAP does not have its own measurement of RH and the PSAP requires the nephelometer scattering measurement in its corrections. This RH constraint is based on the Global Atmospheric Watch (GAW) programme recommendation to make aerosol measurements at low relative humidity (RH < 40%). It was often challenging to meet this criteria at many of the DOE/ARM sites. At ASI there is no harmonized optical data available due to this RH constraint, and some sites have limited optical data, particularly in warmer months when it is more difficult to control RH.

Other constraints and processing. Below is an itemized list of some exceptions to or issues that arose with the general QC described above

- At ASI, COR, MAO, and SBS (CP and SPL), sample air RH was not available for the PNSD measurements because the SMPS instrument model (TSI 3936) used at those sites did not include a sample air RH sensor. As noted previously an upstream Nafion™ drying system was used to lower the SMPS sample RH, however, the resulting sample RH was not recorded (A. Singh, pers. comm., 2024). Because the dryer was used, we assume that these PNSD measurements were also made at low (<40%) RH.

- At ASI, the SMPS measurement range changed from 11–461 nm to 22–461 nm. For the time period when the minimum diameter started at 11 nm, most bins had zero particle counts up to 22 nm (possibly indicating a problem with the DMA). This results in small concentration values for the nucleation mode and many $AF > 1$, with no results for D_{crit} and κ_{CCN} . However, the closure study with the CPC looks acceptable (see discussion and figures in Technical Validation section).
- At SGP, most SS_{act} values for CCNC column B (fixed SS) are missing during 2019 and 2020. Because the measured column temperatures are available, the SS_{act} is calculated by considering the thermal efficiency table and equation 16 in Lance *et al.*⁵⁵.
- At ANX, we did not apply bit19 of the DOE/ARM quality control flag to the CCN files. This flag is related to suspect instrument thermal efficiency. If the QC flag were to be applied, all CCN measurements for the site would be marked as invalid. The instrument mentor noted that bit19 is just an indicator to look more closely at the data. Due to this bit being set in the operational files, all the DOE/ARM operational spectra files also report missing value codes for the fitted values of SS_{act} and N_{CCN} . At this site we started with the basic CCN operational file *aosccn2cola* to generate the final harmonized files. Closure tests utilizing the harmonized ANX CCNC data, described in the technical validation section below, look good.
- At ANX there was also a period where CCN column A data was described as suspect in a DQR due to a possible internal leak in the sheath air line. The closure comparison of both columns at 0.4% SS is acceptable (Fig. S4 in supplemental materials) so we have not flagged this data as invalid.
- For both MAO and SBS-SPL there is a period where the CCNC counter made both increasing SS and decreasing SS scans. Due to limitations of our interpolation processing scheme, the decreasing SS scans are not considered.
- At GUC, we did not apply bit17 ($ambient_pressure < ambient_pressure_min_warning$) of QC for CPC which occurred due to low ambient pressures (station was at 3137 masl, ambient pressure was approximately 670 mb). Takegawa and Sakurai⁷² suggest for 20 nm and higher particles, butanol-based CPCs should have same counting efficiency at this pressure as at sea level.
- Evaluation of the MAO ACSM data after the CDCE was applied using closure studies and comparison with other collocated datasets suggest the data were problematic. As the DQR for this data set also suggest it is suspect, MAO ACSM data have not been further considered in this study. Users interested in chemistry data from MAO are recommended to instead download PI aerosol mass spectrometer (AMS) data from 2 shorter IOP deployments available in the ARM IOP archive directory: /arm-iop/2014/mao/goamazon/T3/alexander-ams.

Data averaging and uncertainties. Following Schmale *et al.*²⁶, we have provided the data at hourly resolution for all instruments studied. These data can be used for Earth system model evaluation^{73,74} and climatological/phenomenological analyses^{27,75–78}. Additionally, we have provided 5-min averages of PNSD, CCN, N_{tot} and scattering and absorption coefficients which have the added advantage of being useful for investigating faster time scale processes such as NPF^{16,29,79}.

For both averaging times the timestamp for the data represents the start of the averaging period. The 1-h and 5-min files for a given instrument file have the same format. Where data from different instruments is merged the measurements represent the same time period. Note: we cannot provide 5-min averages of CCNC column A because the SS scanning does not nicely line up with 5-min intervals. Similarly, the sampling cycles for the ACSM did not lend themselves to 5-min averaging.

It is also important to provide information on measurement uncertainties. For the DOE/ARM instruments in this study, Sisterson⁸⁰ provides information on measurement uncertainty primarily related to instrument calibration. Calibration uncertainty relates to the measurement error determined from testing the instrument with reference standards. Schmale *et al.*²⁶ also provide general uncertainty information for PNSD, CCN and chemical composition measurements along with citations for more in-depth uncertainty information for the SMPS, CCNC and ACSM. Sherman *et al.*⁸¹ included detailed information about uncertainty calculations for the TSI nephelometer and PSAP in their supplemental materials. Providing individual uncertainty calculations for the 6 instruments and 10 sites are outside the scope of this manuscript as they depend not only on the individual instruments, but also environmental conditions (e.g., atmospheric aerosol loading) which varies from location to location. Based on the references listed above we have included some very general estimates of uncertainty for the individual instruments along with comments and further references in Table 5.

Calculated parameters. From the processed and harmonized CCN and PNSD data, we have calculated the activated fraction, critical diameter and hygroscopicity parameter for each site. These calculations are described below.

The AF_{inst} is the ratio of the CCN concentration to the total number concentration (Eq. 2):

$$AF_{inst}(SS) = CCN(SS)/N_{tot,inst} \quad (2)$$

where $inst$ indicates the instrument providing the value of N_{tot} (either SMPS or CPC). AF is calculated for each SS. In the data files we provide AF values based on N_{tot} from both the CPC (AF_{CPC}) and the SMPS (AF_{SMPS}).

Critical diameter for each SS was estimated by integrating the PNSD backwards from largest SMPS diameter (D_{max}) to smaller diameters. The D_{crit} for a given SS is the diameter when the integrated number from the PNSD is equal to the CCN concentration at that SS (Eq. 3). This approach has been used previously, e.g.^{82,83}. Calculating D_{crit} in this way provides a step-function activation cut-off diameter (all particles larger than this size act as CCN) assuming an internal mixture.

Instrument file name	measured parameters	housekeeping	flags	calculated parameters
<i>SiteID_smeps_avgt.csv</i>	$dN/d\log D, N_{tot,SMPS}$	SMPS sample T, P, and RH	total number closure (CPC and SMPS), scattering number closure (neph and SMPS), PM_{10} mass closure (ACSM + PSAP and SMPS)	$N_{tot,CPC}$
<i>SiteID_ccn_cola.csv</i>	$N_{CCN,A}, N_{CCN,har}, N_{CCN,AB}, N_{tot,CPC}, N_{tot,SMPS}$	$SS_{act,A}$ and $SS_{har}, SS_{setpoint,B}$ and $SS_{act,AB}$ CCNC sample T and P for column A, sample flow and sheath flow for column A	column closure, CCN number closure ($SS=0.4, 1$)	$D_{crit}, \kappa_{CCN}, AF_{CPC}, AF_{SMPS}$
<i>SiteID_ccn_colb_avgt.csv</i>	$N_{CCN,B}, N_{tot,CPC}, N_{tot,SMPS}$	$SS_{setpoint,B}$ and $SS_{act,B}$ CCNC sample T and P for column B, sample flow and sheath flow for column B	CCN number closure ($SS=0.4$), total number closure (CPC and SMPS)	$D_{crit}, \kappa_{CCN}, AF_{CPC}, AF_{SMPS}$
<i>SiteID_acsm.csv</i>	$M_{Ammonium}, M_{Chloride}, M_{Organics}, M_{Nitrate}, M_{Sulfate}$	sample T, P, and RH (from SMPS)	PM_{10} mass closure (ACSM + PSAP and SMPS)	$M_{ACSM}, M_{SMPS}, M_{BC,PSAP}, M_{ACSM+PSAP}$
<i>SiteID_neph_avgt.csv</i>	σ_{sp} and σ_{bsp} at three wavelengths and two size cuts	nephelometer sample T, P, and RH	size cut (PM_{10}, PM_{10})	n/a
<i>SiteID_psap_avgt.csv</i>	σ_{ap} at three wavelengths and two size cuts	sample RH (from nephelometer)	size cut (PM_{10}, PM_{10})	n/a

Table 6. Summary of harmonized data files. *SiteID* is the short site identifier (e.g., SGP for Southern Great Plains), and *avgt* is the averaging time of the file where appropriate (i.e., either 5 min or 1 h). All variables are reported in standard temperature and pressure conditions. The housekeeping T, P and RH values are reported measurement conditions for the specified instrument. See Tables 7–12 for more detailed descriptions.

$$N_{CCN}(SS) = \int_{D_{crit}(SS)}^{D_{max}} \frac{dN}{d\log D} d\log D \quad (3)$$

The κ_{CCN} value is calculated using κ -Köhler theory⁸⁴. This is done in an iterative approach: D_{crit} is first calculated as explained above, then D_{crit} is put into the Köhler Eq. (4), and then κ_{CCN} is varied until the corresponding critical SS equals the SS value at which N_{CCN} was measured.

$$S = \frac{D^3 - D_{crit}^3}{D^3 - D_{crit}^3(1 - \kappa_{CCN})} \exp\left(\frac{4\sigma_{s/a}M_w}{RT\rho_w D}\right) \quad (4)$$

where S is the saturation vapor pressure of water over an aqueous solution droplet of diameter D , $\sigma_{s/a}$ is the surface tension at the surface/air interface, M_w is the molecular weight of water, R is the universal gas constant, T is temperature and ρ_w is density of water. We used the surface tension of pure water in all of our calculations since at the point of activation we assume that the solution is diluted. Equation 4 can be expressed in terms of the SS, which is defined as $SS = (S-1)*100$. These CCN-derived κ_{CCN} values quantify the effective hygroscopicity of activated particles in the CCNC and show a dependence on SS.

We should note that DOE/ARM provides a value-added product (VAP) data stream (*aoscscnsmppskappa*) for many of these locations that includes calculated κ_{CCN} values. We do not use this data stream because in it the D_{crit} and κ_{CCN} are calculated for SS values different than our SS_{har} values.

In order to have a more complete representation of aerosol mass concentration in our validation analysis, we also calculated the black carbon mass concentration $M_{BC,PSAP}$ from the PSAP σ_{ap} measurements (Eq. 5):

$$M_{BC,PSAP} = \sigma_{apG,1}/MAC \quad (5)$$

where $\sigma_{apG,1}$ is the green wavelength (529 nm) PM_{10} absorption coefficient measured by the PSAP and MAC is the mass absorption cross section which was assumed to be $10 \text{ m}^2 \text{ g}^{-1}$ ⁸⁵.

Data Records

We developed 6 types of aerosol data files in the course of the harmonization process. Table 6 provides an overview of all the harmonized files (file names and contents), while Tables 7–12 provide a more detailed description of what is in each file. Each of these files is also briefly described below. The file names have the following format: *SiteID_Instrument_avgt.csv* where *SiteID* is the three letter identifier given in Table 1, *Instrument* is the instrument type and *avgt* is the time resolution where appropriate (i.e., 5 min or 1 h for 5-min averaged and 1-h averaged), respectively. For most sites the 1–2 years of data are concatenated into a single file and no year is included in the file name. For SGP we have included the year in the file name (*SGP_Instrument_avgt_Year.csv*) as there are seven years of data and we've provided the data for SGP split by year. For the MOS dataset there is a separate file containing latitude and longitude as those measurements were taken aboard a moving ship. The MOS position data file was generated from the campaign navigation files^{86–90} which we averaged over 5-min and 1-h to match our harmonized data files. For all time series files the primary time stamp is given in the form: YYYY-MM-DD hh:mm:ss. We also provide *README* files for each site and instrument that briefly describes the variables and contain additional metadata (e.g., instrument model, site location, etc). All harmonized data files are available in the figshare repository⁹¹.

Variable	Description
Start_date	Start date in UTC (yyyy-mm-dd hh:mm:ss format)
N_CN_SMPS_STP	Total particle number concentration from SMPS at STP (see P_SMPS, T_SMPS) (cm^{-3}). Referred to as $N_{\text{tot,SMPS}}$ in manuscript text.
P_SMPS	Sample pressure from SMPS (hPa).
T_SMPS	Sample temperature from SMPS ($^{\circ}\text{C}$).
RH_SMPS	Sample relative humidity from SMPS (%).
N_CN_CPC_STP	Total particle concentration from CPC at STP (cm^{-3}). Referred to as $N_{\text{tot,CPC}}$ in manuscript text.
Qc_CPC_SMPS	Quality check of number closure. Flag indicates comparison of N_CN_SMPS_STP and N_CN_SMPS_STP. Qc = 0 differences smaller than 50%; Qc = 1 differences bigger than 50%.
Qc_ACSM_SMPS	Quality check of PM_{10} mass closure. Flag indicates comparison of Mass_ACSM_PSAP_STP and Mass_SMPS_STP. Qc = 0 for differences smaller than 50%; Qc = 1 for differences bigger than 50%.
Qc_Neph_SMPS	Quality check of scattering closure. Flag indicates comparison of calculated aerosol scattering coefficient and BsG1_STP measured by the integrating nephelometer. Qc = 0 differences smaller than 50%; Qc = 1 differences bigger than 50%.
$dN/d\log D_i$ (bin1, bin2, ..., bin113)	Particle number size concentration at STP for each size bin ¹ (cm^{-3}). Referred to as $dN/d\log D_i$ in manuscript text.

Table 7. Description of *SiteID_smgs_avgt.csv* file. ¹Midpoint diameter for each size bin is given in header for each. $dN/d\log D_i$ column, e.g., if the midpoint diameter is 101.8 nm then the column header is D_101_8.

The harmonized SMPS files (see Table 7) contain $N_{\text{tot,SMPS}}$, the particle number concentrations for each individual size bin i ($dN/d\log D_i$), some housekeeping information (T, P, RH), $N_{\text{tot,CPC}}$, and several closure flags for number closure, scattering closure and mass closure (see Technical Validation section). The harmonized file name for the SMPS datasets is *SiteID_smgs_avgt.csv*. The $N_{\text{tot,CPC}}$ value is the mean between the 25th and 75th percentile of the instrument's measurement period in the file. Here we used the SMPS measure of sample T and P to correct $N_{\text{tot,CPC}}$ to STP.

There are two types of files for the harmonized CCN measurements - one for column A data (see Table 8) and one for column B data (see Table 9). The harmonized file name for the CCN datasets are *SiteID_ccn_col_a.csv* and *SiteID_ccn_col_b_avgt.csv*. If the CCNC was a one column instrument (i.e., at MAO and both SBS sites) then only the column A file was generated, with NaN values for fields related to column B values. The *SiteID_ccn_col_a.csv* file contains $N_{\text{CCN,A}}$, $N_{\text{CCN,AB}}$, $N_{\text{tot,CPC}}$ and $N_{\text{tot,SMPS}}$ in (cm^{-3}) and $\text{SS}_{\text{act,A}}$ and SS_{har} values as well as some housekeeping parameters and flags indicating results of the quality checks (see Technical Validation section) and calculated parameters. As with the SMPS files, the $N_{\text{tot,CPC}}$ value is the mean between the 25th and 75th percentile of the instrument's measurement period for each time stamp in the file (for example, in column A file, the CPC value given is the mean value at each SS time). Here we used the CCNC measure of sample T and P to correct $N_{\text{tot,CPC}}$ to STP.

The *SiteID_ccn_col_b_avgt.csv* file contains similar information, but only for the second CCNC column (column B) which is held at constant SS. It should be noted that the averaging time for the column B parameters is different in the two CCN files. The column B parameters included in the *SiteID_ccn_col_a.csv* file (e.g., $N_{\text{CCN,AB}}$, and $\text{SS}_{\text{act,AB}}$) are calculated from values temporally matched to the filtered column A parameters (i.e., they are matched to $N_{\text{CCN,A}}$ between the 25th and 75th percentile for each $\text{SS}_{\text{act,A}}$). In the *SiteID_ccn_col_b_avgt.csv* file, the mean $N_{\text{CCN,B}}$ represents the average of the high resolution $N_{\text{CCN,B}}$ between the 25th and 75th percentile for the averaging time of the file (5-min or 1-h). The other mean parameters (SS, flows, etc.) in the column B file are calculated from values temporally matched with the filtered high resolution $N_{\text{CCN,B}}$.

The harmonized ACSM files (see Table 10) contain the mass concentration values ($\mu\text{g m}^{-3}$) of the particle organics, sulfate, ammonium, nitrate, chloride, from the ACSM and black carbon derived from the PSAP data. Total mass concentrations calculated from the ACSM, SMPS, and combined ACSM and PSAP are included, as well as a quality assurance flag (qc_ACSM_SMPS) indicative of how well the combined mass concentration of the ACSM and PSAP compares with the mass concentration derived from the PNSD data (see Technical Validation section). The harmonized file name for the ACSM datasets is *SiteID_acsm.csv*.

The harmonized nephelometer data sets (see Table 11) include the truncation-corrected scattering and backscattering coefficient values (in Mm^{-1}) at the three measurement wavelengths (450, 550 and 700 nm) for the two measurement size cuts (PM_{10} and $\text{PM}_{2.5}$). Additionally, they include nephelometer housekeeping values (T, P, RH) and the number of nephelometer data points making up each average value. Nephelometer RH > 40% was the primary reason scattering data are reported as missing in the harmonized data file. The harmonized file name for the nephelometer datasets is *SiteID_neph_avgt.csv*.

The harmonized PSAP data sets (see Table 12) include the corrected absorption coefficient values (in Mm^{-1}) at the three PSAP measurement wavelengths (464, 529 and 648 nm) for the two measurement size cuts (PM_{10} and $\text{PM}_{2.5}$). The PSAP files also include the nephelometer sample RH value as nephelometer RH > 40% was primary reason absorption data were excluded. The PSAP correction requires the nephelometer data, so if the nephelometer data were invalid due to RH > 40%, then the PSAP data were also marked as invalid. The harmonized file name for the PSAP datasets is *SiteID_psap_avgt.csv*.

Variable	Description
Start_date	Start date in UTC (yyyy-mm-dd hh:mm:ss format)
End_date	End date in UTC (yyyy-mm-dd hh:mm:ss format) ¹
SS_harmonized_A	Column A supersaturation harmonized value (%). Referred to as SS_{har} in manuscript text.
SS_actual_A	Column A mean value of the supersaturation actual values corresponding to the mean filtered $N_{CCN,A}$ (see $N_{CCN_mean_STP_A}$) (%). Referred to as mean $SS_{act,A}$ in manuscript text.
$N_{CCN_mean_STP_A}$	Column A mean value of the high resolution $N_{CCN,A}$ for each mean $SS_{act,A}$ filtered to only include the values between the 25th and 75th percentile of the high resolution $N_{CCN,A}$ values. Converted to STP (see P_{CCNC_A} , T_{CCNC_A}) (cm^{-3}). Referred to as mean $N_{CCN,A}$ in manuscript text.
$N_{CCN_har_STP_A}$	Interpolation of mean $N_{CCN,A}$ to SS_{har} and converted to STP (see P_{CCNC_A} , T_{CCNC_A}) (cm^{-3}). Referred to as $N_{CCN,har}$ in manuscript text.
SS_setpoint_B	Column B supersaturation setpoint value (%). Referred to as $SS_{setpoint,B}$ in manuscript text.
SS_actual_AB	Column B mean value of the supersaturation actual values of the filtered $N_{CCN,B}$ (see $N_{CCN_mean_STP_AB}$) (%). Referred to as mean $SS_{act,AB}$ in manuscript text.
$N_{CCN_mean_STP_AB}$	Column B mean value of the high resolution $N_{CCN,B}$ calculated from data temporally matched with the filtered $N_{CCN,A}$. Converted to STP with P_{CCNC_A} and T_{CCNC_A} (cm^{-3}). Referred to as mean $N_{CCN,AB}$ in manuscript text.
P_{CCNC_A}	CCNC column A sample pressure. Mean value of the pressure for the mean filtered $N_{CCN,A}$ (see $N_{CCN_mean_STP_A}$) (hPa).
T_{CCNC_A}	CCNC column A sample temperature. Mean value of the temperature for the mean filtered $N_{CCN,A}$ (see $N_{CCN_mean_STP_A}$) ($^{\circ}C$).
Q_{sample_A}	CCNC column A sample flow. Mean value of the volumetric flow rate of sample air for the mean filtered $N_{CCN,A}$ (see $N_{CCN_mean_STP_A}$) ($cm^3 min^{-1}$).
Q_{sheath_A}	CCNC column A sheath flow. Mean value of the volumetric flow rate of sheath air for the mean filtered $N_{CCN,A}$ (see $N_{CCN_mean_STP_A}$) ($cm^3 min^{-1}$).
$N_{CN_CPC_STP_A}$	Total particle concentration from CPC at STP. Mean value of total particle concentration from CPC temporally matched with the filtered $N_{CCN,A}$. Converted to STP with P_{CCNC_A} and T_{CCNC_A} (cm^{-3}). Referred to as $N_{tot,CPC}$ in manuscript text.
$N_{CN_SMPS_STP_A}$	Total particle number concentration calculated from SMPS size distribution at STP. Mean value of 10-minute SMPS size distribution (two closest SMPS scans to Start_date) (cm^{-3}). Referred to as $N_{tot,SMPS}$ in manuscript text.
Critical_diameter_A	Critical diameter calculated with $N_{CCN_har_STP_A}$ and mean value of 10-minute SMPS size distribution (two closest SMPS scans to Start_date) (nm). Referred to as D_{crit} in manuscript text.
Kappa_A	Hygroscopicity parameter obtained from D_{crit} and Eq. 4. Referred to as κ_{CCN} in manuscript text.
Activation_fraction_CPC_A	Ratio of activated particles. $N_{CCN_har_STP_A} / N_{CN_CPC_STP_A}$. Referred to as AF_{CPC} in manuscript text.
Activation_fraction_SMPS_A	Ratio of activated particles. $N_{CCN_har_STP_A} / N_{CN_SMPS_STP_A}$. Referred to as AF_{SMPS} in manuscript text.
Qc_column_AB	Quality check of CCN column closure at 0.4% SS. Flag indicates comparison between $N_{CCN_har_STP_A}$ for column A and $N_{CCN_mean_STP_AB}$ for column B. Qc = 0 differences smaller than 50%; Qc = 1 differences bigger than 50%.
Qc_CCN04_N80_A	Quality check of CCN number closure at 0.4% SS. Flag indicates comparison of $N_{CCN_har_STP_A}$ at 0.4% SS for column A CCNC and mean value of number concentration of particles greater than 80 nm (N80) calculated from mean value of 10-minute SMPS data (2 closest SMPS scans to Start_date) at STP for SMPS. Qc = 0 differences smaller than 50%; Qc = 1 differences bigger than 50%.
Qc_CCN1_N30_A	Quality check of CCN number closure at 1% SS. Flag indicates comparison of $N_{CCN_har_STP_A}$ at 1% SS for column A CCNC and mean value of number concentration of particles greater than 30 nm (N30) calculated from mean value of 10-minute SMPS data (2 closest SMPS scans to Start_date) at STP for SMPS. Qc = 0 differences smaller than 50%; Qc = 1 differences bigger than 50%.

Table 8. Description of *SiteID_ccnCola.csv* file. ¹Start and end dates are provided in this file corresponding to the start and end of each individual SS measurement during the hourly scan across all SS. There is no 5-min averaged file for this file type.

Technical Validation

Following the initial data review and handling, a variety of checks and closure experiments were performed to assess the data quality and further screen the data as described below. Successful closure experiments indicate that the data sets are internally consistent and are commonly used as part of the QA/QC process^{26,92–96}. Only sites and periods of measurements passing our detailed QA/QC process and closure tests as described below are included in the final harmonized data sets presented in this paper. We have also included flags (described below) that the user may want to apply as further data constraints. We use the suite of measurements from ASI, MOS and GUC to provide an example for some of our time series analyses (e.g., Figs. 3, 4a,b) and for GUC for each closure test (Fig. 5). Similar plots for all the sites can be found in supplemental materials.

Time series analysis. Plots of the full data time series for each parameter, monthly and diurnal patterns and interannual variability (where possible) were used to identify potential irregularities in the data. For example, time series analysis of the wavelength dependence of the nephelometer scattering data at SBS-CP led to removal of this data set from consideration. Specifically, total scattering coefficient at the red wavelength was often higher than total scattering coefficient for the blue and green wavelengths which is unrealistic. Additionally, at that site the backscattering coefficient values were often similar to and sometimes greater than the total scattering

Variable	Description
Start_date	Start date in UTC (yyyy-mm-dd hh:mm:ss format)
End_date	End date in UTC (yyyy-mm-dd hh:mm:ss format)
SS_setpoint_B	Column B supersaturation setpoint value (%). Referred to as $SS_{setpoint,B}$ in manuscript text.
SS_actual_B	Column B mean value of the supersaturation actual values corresponding to the mean filtered $N_{CCN,B}$ (see $N_{CCN_mean_STP_B}$) (%). Referred to as mean $SS_{act,B}$ in manuscript text.
$N_{CCN_mean_STP_B}$	Column B mean value of the high resolution $N_{CCN,B}$ for each mean $SS_{act,B}$, filtered to only include the values between the 25th and 75th percentile of the high resolution $N_{CCN,B}$ values for the file averaging time (5-min or 1-h). Converted to STP (see P_{CCNC_B} , T_{CCNC_B}) (cm^{-3}). Referred to as mean $N_{CCN,B}$ in manuscript text.
P_{CCNC_B}	CCNC column B sample pressure. Mean value of the pressure of the mean filtered $N_{CCN,B}$ (see $N_{CCN_mean_STP_B}$) (hPa).
T_{CCNC_B}	CCNC column B sample temperature. Mean value of the temperature of the mean filtered $N_{CCN,B}$ (see $N_{CCN_mean_STP_B}$) ($^{\circ}C$).
Q_{sample_B}	CCNC column B sample flow. Mean value of the volumetric flow rate of sample air of the mean filtered $N_{CCN,B}$ (see $N_{CCN_mean_STP_B}$) ($cm^3 min^{-1}$).
Q_{sheath_B}	CCNC column B sheath flow. Mean value of the volumetric flow rate of sheath air of the mean filtered $N_{CCN,B}$ (see $N_{CCN_mean_STP_B}$) ($cm^3 min^{-1}$).
$N_{CN_CPC_STP_B}$	Total particle concentration from CPC at STP temporally matched with mean filtered $N_{CCN,B}$. Converted to STP with P_{CCNC_B} and T_{CCNC_B} (cm^{-3}). Referred to as $N_{tot,CPC}$ in manuscript text.
$N_{CN_SMPS_STP_B}$	Total particle number concentration calculated from SMPS size distribution at STP. Mean value of 5-min/1-h SMPS measurement (cm^{-3}). Referred to as $N_{tot,SMPS}$ in manuscript text.
Critical_diameter_B	Critical diameter calculated with $N_{CCN_mean_STP_B}$ and the corresponding time period of SMPS data (nm). Referred to as D_{crit} in manuscript text.
Kappa_B	Hygroscopicity parameter obtained from D_{crit} and Eq. 4. Referred to as κ_{CCN} in manuscript text.
Activation_fraction_CPC_B	Ratio of activated particles. $N_{CCN_mean_STP_B} / N_{CN_CPC_STP_B}$. Referred to as AF_{CPC} in manuscript text.
Activation_fraction_SMPS_B	Ratio of activated particles. $N_{CCN_mean_STP_B} / N_{CN_SMPS_STP_B}$. Referred to as AF_{SMPS} in manuscript text.
$Qc_CCN04_N80_B$	Quality check of CCN number closure at 0.4% SS. Flag indicates comparison of $N_{CCN_mean_STP_B}$ at 0.4% SS for column B CCNC and mean value of number concentration of particles greater than 80 nm (N80) of the corresponding time period data at STP for SMPS. $Qc = 0$ differences smaller than 50%; $Qc = 1$ differences bigger than 50%.
Qc_CPC_SMPS	Quality check of number closure. Flag indicates comparison of $N_{tot,SMPS}$ and $N_{tot,CPC}$. $Qc = 0$ differences smaller than 50%; $Qc = 1$ differences bigger than 50%.

Table 9. Description of *SiteID_ccn_colb_avgt.csv* file. The variable names in this file are similar to those in the *SiteID_ccn_colc.csv* file, but the averaging time is different for this file so we have provided start and end times to make that clearer - see text for more detail.

Variable	Description
Start_date	Start date in UTC (yyyy-mm-dd hh:mm:ss format)
Mass_organics_STP	Mass concentration of particle organics at STP ($\mu g m^{-3}$). Referred to as $M_{Organics}$ in manuscript text.
Mass_sulfate_STP	Mass concentration of particle sulfate at STP ($\mu g m^{-3}$). Referred to as $M_{Sulfate}$ in manuscript text.
Mass_ammonium_STP	Mass concentration of particle ammonium at STP ($\mu g m^{-3}$). Referred to as $M_{Ammonium}$ in manuscript text.
Mass_nitrate_STP	Mass concentration of particle nitrate at STP ($\mu g m^{-3}$). Referred to as $M_{Nitrate}$ in manuscript text.
Mass_chloride_STP	Mass concentration of particle chloride at STP ($\mu g m^{-3}$). Referred to as $M_{Chloride}$ in manuscript text.
Mass_ACSM_STP	Mass concentration from ACSM at STP ($\mu g m^{-3}$). Referred to as M_{ACSM} in manuscript text.
Mass_SMPS_STP	Mass concentration from SMPS at STP ($\mu g m^{-3}$). Referred to as M_{SMPS} in manuscript text.
Mass_BC_PSAP_STP	Mass concentration of black carbon calculated from PSAP absorption coefficient using green wavelength (529 nm) and PM_{10} size cut at STP ($\mu g m^{-3}$). Referred to as $M_{BC,PSAP}$ in manuscript text.
Mass_ACSM_PSAP_STP	Sum of mass concentration from ACSM and PSAP at STP ($\mu g m^{-3}$). Referred to as $M_{ACSM+PSAP}$ in manuscript text.
Qc_ACSM_SMPS	Quality check of PM_{10} mass closure. Flag indicates comparison of $M_{ACSM+PSAP}$ and M_{SMPS} . $Qc = 0$ for differences smaller than 50%; $Qc = 1$ for differences bigger than 50%.
P_{SMPS}	Sample pressure from SMPS (hPa).
T_{SMPS}	Sample temperature from SMPS ($^{\circ}C$).
RH_{SMPS}	Sample relative humidity from SMPS (%).

Table 10. Description of *SiteID_acsm.csv* file. Sample pressure and temperature used for STP conversion were taken from SMPS.

coefficient values; for atmospheric aerosols measured at low RH the backscattering coefficient is normally 10–20% of the total scattering coefficient⁸¹. Additional time series plots of instrument specific diagnosis variables such as supersaturations (e.g., Fig. 3 for ASI and MOS and Fig. 4a,b for GUC), flow ratio, temperatures, RH, and so on, were also inspected to identify possible malfunctioning of each instrument.

Variable	Description
Start_date	Start date in UTC (yyyy-mm-dd hh:mm:ss format)
BsB10_STP	Scattering coefficient for 450 nm at STP, PM ₁₀ size cut (Mm^{-1}).
BsG10_STP	Scattering coefficient for 550 nm at STP, PM ₁₀ size cut (Mm^{-1}).
BsR10_STP	Scattering coefficient for 700 nm at STP, PM ₁₀ size cut (Mm^{-1}).
BbsB10_STP	Backscattering coefficient for 450 nm at STP, PM ₁₀ size cut (Mm^{-1}).
BbsG10_STP	Backscattering coefficient for 550 nm at STP, PM ₁₀ size cut (Mm^{-1}).
BbsR10_STP	Backscattering coefficient for 700 nm at STP, PM ₁₀ size cut (Mm^{-1}).
T_neph_inlet10	Nephelometer inlet temperature for PM ₁₀ size cut ($^{\circ}C$).
RH_neph10	Nephelometer sample RH for PM ₁₀ size cut (%).
T_neph10	Nephelometer sample temperature for PM ₁₀ size cut ($^{\circ}C$).
P_neph10	Nephelometer sample pressure for PM ₁₀ size cut (hPa).
BsB1_STP	Scattering coefficient for 450 nm at STP, PM ₁ size cut (Mm^{-1}).
BsG1_STP	Scattering coefficient for 550 nm at STP, PM ₁ size cut (Mm^{-1}). Referred to as $\sigma_{spG,1}$ in manuscript text.
BsR1_STP	Scattering coefficient for 700 nm at STP, PM ₁ size cut (Mm^{-1}).
BbsB1_STP	Backscattering coefficient for 450 nm at STP, PM ₁ size cut (Mm^{-1}).
BbsG1_STP	Backscattering coefficient for 550 nm at STP, PM ₁ size cut (Mm^{-1}).
BbsR1_STP	Backscattering coefficient for 700 nm at STP, PM ₁ size cut (Mm^{-1}).
T_neph_inlet1	Nephelometer inlet temperature for PM ₁ size cut ($^{\circ}C$).
RH_neph1	Nephelometer sample RH for PM ₁ size cut (%).
T_neph1	Nephelometer sample temperature for PM ₁ size cut ($^{\circ}C$).
P_neph1	Nephelometer sample pressure for PM ₁ size cut (hPa).
num_pts10	Number of points in PM ₁₀ average.
num_pts1	Number of points in PM ₁ average.

Table 11. Description of *SiteID_neph_avgt.csv* file.

Variable	Description
Start_date	Start date in UTC (yyyy-mm-dd hh:mm:ss format)
BaB10_STP	Absorption coefficient for 464 nm at STP, PM ₁₀ size cut (Mm^{-1}).
BaG10_STP	Absorption coefficient for 529 nm at STP, PM ₁₀ size cut (Mm^{-1}).
BaR10_STP	Absorption coefficient for 648 nm at STP, PM ₁₀ size cut (Mm^{-1}).
RH_neph10	Nephelometer sample RH for PM ₁₀ size cut (%).
BaB1_STP	Absorption coefficient for 464 nm at STP, PM ₁ size cut (Mm^{-1}).
BaG1_STP	Absorption coefficient for 529 nm at STP, PM ₁ size cut (Mm^{-1}). Referred to as $\sigma_{apG,1}$ in manuscript text.
BaR1_STP	Absorption coefficient for 648 nm at STP, PM ₁ size cut (Mm^{-1}).
RH_neph1	Nephelometer sample RH for PM ₁ size cut (%).
num_pts10	Number of points in PM ₁₀ average.
num_pts1	Number of points in PM ₁ average.

Table 12. Description of *SiteID_psap_avgt.csv* file.

Number closure. For each site included in the analysis, the collocated total particle number concentration (N_{tot}) from a CPC ($N_{tot,CPC}$) was compared with that calculated by integrating PNSD ($N_{tot,SMPS}$). Ideally, the integrated PNSD should be highly correlated and within a few percent of the total number concentration. Here we flagged the PNSD data when the comparison between the integrated PNSD and the independent $N_{tot,CPC}$ was greater than 50%. The flag in the harmonized files (*SiteID_smeps_avgt.csv*) for this closure test is Qc_CPC_SMPS. Because such a comparison does not definitively indicate which measurement is problematic (SMPS or CPC), other closure experiments with independent measurements were also performed in order to isolate the problem measurement. Figure 5a for GUC shows the scatter plot of the total particle number concentration measured with the SMPS and the CPC for 1-h time resolution (this closure test was also done with the 5-min data for all the sites). The results show very good agreement between both instruments, with most of the data confined within the $\pm 50\%$ deviation from the 1:1 line. The slope is < 1 indicating that the CPC instrument typically reported higher N_{tot} than the SMPS. Figure S1 in the supplementary material presents number closure plots for all the sites analysed. All the sites show a slope < 1 except ENA (Fig. S1d) and MOS (Fig. S1g). This may be due to instrument issues that were not caught by the automatic QC or mentor DQR reports. Alternatively, it could also be due to problematic sampling conditions; for example, at MOS the system developed for excluding ship emissions did not work, necessitating development of a post-processing pollution detection algorithm⁹⁷. That contamination masking algorithm is not applied here, but should be considered when the data are used.

Scattering closure. Another test of the PNSD data quality was performed by calculating the aerosol scattering coefficient from the PNSD using Mie theory⁹⁸ and comparing to the sub-micron green wavelength scattering ($\sigma_{spG,1}$) measured by the integrating nephelometer. For all sites we assumed the particles were spherical and had a constant index of refraction of $1.55 + 0.01i$. If the number closure for a time period looked problematic but the scattering closure was good then the problem is likely with the CPC measurements rather than the PNSD measurements. Scattering closure was only performed for hourly averaged data when there was a PM_{10} size cut for scattering and the nephelometer sample RH < 40%. These two constraints limit where scattering closure can be done for some locations and time periods. The closure flag in the harmonized SMPS data files (*SiteID_smgs_avgt.csv*) is noted as *Qc_Neph_SMPS*. The scattering closure result for GUC is shown in Fig. 5b. This test shows very good results at GUC, with a slope near 1 and high correlation coefficient ($R = 0.9$), indicating good agreement between the SMPS and the nephelometer. The good scattering closure and number closure tests confirm the good performance of the SMPS at GUC. Figure S2 in the supplementary materials presents scattering closure plots for the other sites. For all sites the correlation coefficient is higher than 0.8 except at ENA (Fig. S2d), with a value of $R = 0.64$. The data dispersion observed for this closure test, together with the dispersion for the number closure test, leads us to believe that the SMPS measurements at ENA may have a problem.

PM_{10} mass closure. A final PNSD check was performed by performing PM_{10} mass closure tests. The use of the level c CDCE-corrected ACSM data is expected to result in a better mass closure result than if a constant CE is used⁶¹. For this closure test, the sub-micron mass calculated from the PNSD (M_{SMPS}) was compared with the PM_{10} mass from the ACSM plus $M_{BC,PSAP}$ estimated from the PSAP absorption coefficient measurement ($M_{ACSM+PSAP}$). If the number and scattering closure tests appeared to be good but the mass closure was questionable then this suggests a problem with the ACSM measurements. A mass concentration was calculated from the PNSD for aerosols larger than 40 nm to remain consistent with the aerosol range of the ACSM (40 nm to 1 μm)⁵⁷. The density used to calculate (M_{SMPS}) was a weighted density based on the densities listed in the ACSM instrument handbook⁵⁷ and the mass concentrations of the species measured by the ACSM. The densities of aerosol organics, sulfate, nitrate, ammonium, and chloride are 1.2, 1.77, 1.72, 1.77, and 1.52 g cm⁻³, respectively⁵⁸. The density of BC used in this study is 1.7 g cm⁻³⁹⁹. In the harmonized ACSM files (*SiteID_acsm.csv*), we flagged ACSM values (flag = *Qc_ACSM_SMPS*) when $M_{ACSM+PSAP}$ differed from M_{SMPS} by more than 50%. This flag serves as an additional quality control criterion to ensure the accuracy and agreement of ACSM and SMPS measurements. It should be noted that the SMPS measurements are referenced to aerosol mobility diameter while the ACSM measurements are reported for aerosol aerodynamic diameter. We have not accounted for the difference in diameter type in this closure test.

Correlation coefficients between the two mass values are also included as an additional criterion for quality assurance. The result of the PM_{10} mass closure for GUC is shown in Fig. 5c. The closure results for GUC show good agreement (slope = 0.91, $R = 0.95$) between $M_{ACSM+PSAP}$ and M_{SMPS} , suggesting a good performance of both instruments during the measurement period. Figure S3 in the supplementary materials presents the PM_{10} mass closure plots for the other sites. Again, all the sites show a good agreement between measurements except ENA (Fig. S3d) (slope = 0.34 and $R = 0.48$). The three closure studies applied to the SMPS data for all the sites suggest a good agreement for all the sites except at ENA, for which a further filtering of the SMPS data based on our closure test flags in the harmonized data sets is recommended.

CCN column closure. As noted above, some sites operated a two column CCNC. When this was the case, we compared $N_{CCN,B}$ which operated at constant $SS_{setpoint,B}$ to $N_{CCN,A}$ at the same SS_{har} from column A. This comparison is expected to show consistency between the two columns but will not be perfect because while the column $SS_{setpoint}$ values are identical the SS_{act} in each column differ. For column A we used $N_{CCN,har}$ while for column B we used $N_{CCN,act,B}$ (recall $N_{CCN,B}$ cannot be harmonized to a specific SS because column B was operated at a single SS). If the N_{CCN} for the two columns are not consistent this suggests an issue with the calibration of one or both columns. The closure flag in the *SiteID_ccn_col.csv* files is *Qc_column_AB*. Figure 5d for GUC shows the results of this comparison. In the case of GUC, the agreement is exceptional, with a slope of 0.99 and $R = 1$. Figure S4 in the supplemental materials presents the column closure results for all the sites analysed. The excellent agreement observed at GUC is not always found for the other sites. Note: at MOS (Fig. S4g), the $SS_{setpoint,B}$ value is 0.4% SS but, as the corresponding $SS_{act,AB}$ mean values are approximately 0.6%, this column is compared with the column A measurements SS_{har} at 0.6%.

CCN number closure at high SS. CCN number closure was performed by comparing $N_{tot,SMPS}$ with the $N_{CCN,A}$ measured at the highest supersaturation value. The CCN concentration should always be equal to or less than N_{tot} as there cannot be more CCN than there are particles. At 1% SS most particles are likely to be activated, although this is highly dependent on the aerosol properties, especially on the aerosol hygroscopicity and size. The presence of very small particles can negatively impact the comparison as they can be too small to activate in the CCN counter, so this comparison was performed only for particles with diameters larger than 30 nm (N_{30}). The closure flag in the *SiteID_ccn_col.csv* file is *Qc_CCN1_N30*. Figure 5e for GUC and Fig. S5 of the supplemental material for all the sites analysed show the results of this closure. For most sites, the results are consistent and the slopes of the comparison are > 1 (except ASI, where all the particles activate). The largest slope is obtained for MOS (Fig. S5g, respectively) indicating the significant contribution of non-CCN active particles at these three sites.

CCN number closure at 0.4% SS. We found some cases in which the two columns of the CCNC did not agree (as in SGP where some points were outside the 50% bounds (Fig. S4j)). In order to elucidate which column

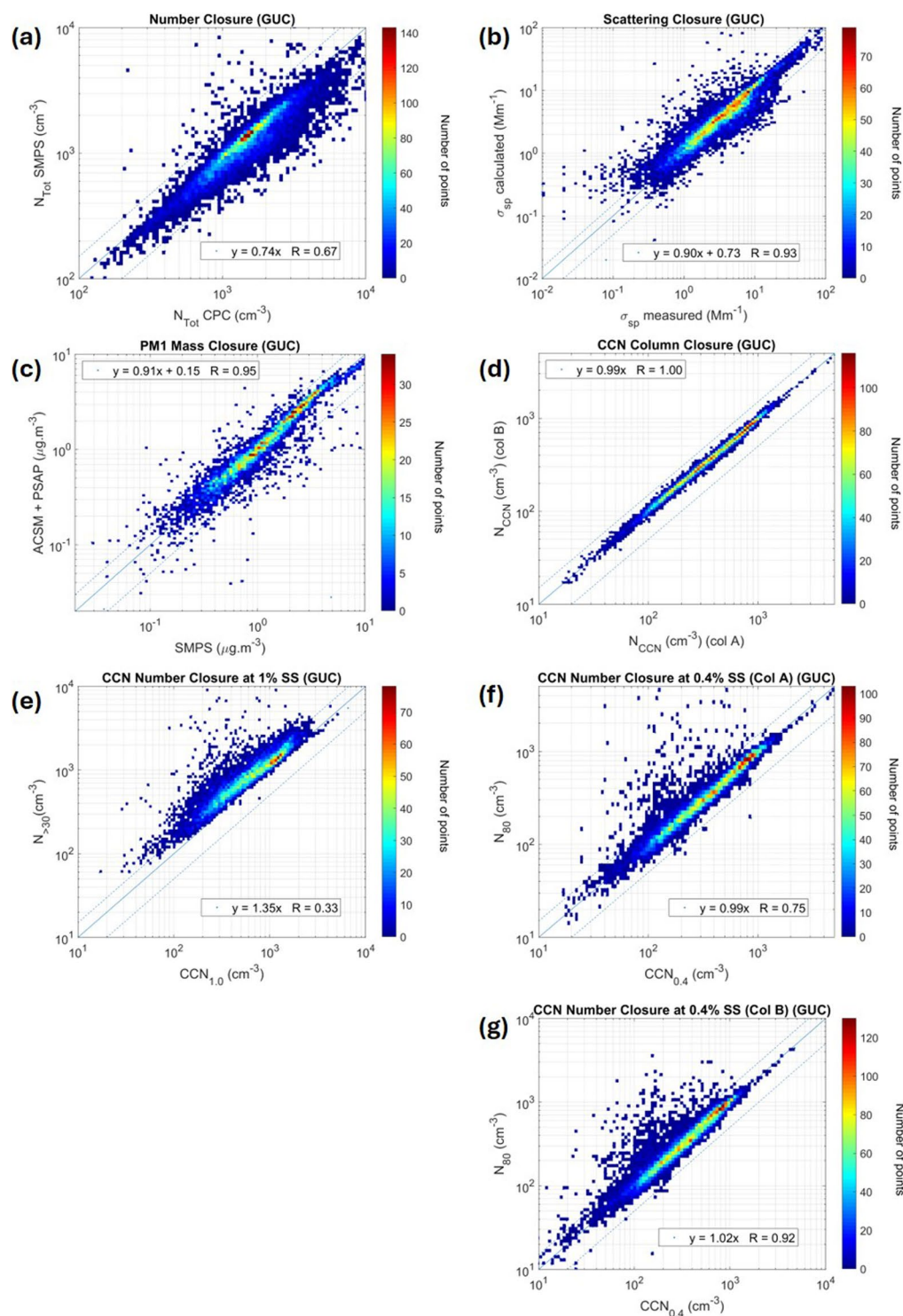


Fig. 5 Example of data closure studies for GUC. **(a)** Number closure; **(b)** Scattering closure; **(c)** PM₁ mass closure; **(d)** CCN column closure; **(e)** CCN number closure at 1% SS; **(f)** CCN number closure at 0.4% SS for column A; **(g)** Same as f but for column B. The solid line is the 1:1 line and the dashed lines indicate $\pm 50\%$ deviation from the 1:1 line. Plots a, b, c and g use 1-h averaged data, while plots d, e, and f used 1-sec data when both CCN columns overlapped (**d**) or when SMPS and CCN column A overlapped (**e** and **f**).

is performing properly, we performed an additional CCN number closure test with the PNSD. In this case, we compared the CCN concentration at 0.4% (which is the common SS in both columns) against the N_{80} (particle number concentration of particles with diameter above 80 nm from the SMPS). This comparison is not definitive since the 80 nm threshold might be more suitable for some sites than for others. However, it offers additional insight into the behaviour of both columns and the activation properties at the different sites. Data were flagged if

the difference was greater than 50%. The closure flag for this test in the harmonized files (*SiteID_ccn_cola.csv*) is Qc_CCN04_N80. Results of this quality check are shown for GUC in Fig. 5f,g for columns A and B, respectively, and in Figs. S6, S7 of the supplementary material for all the sites analysed. For GUC, the agreement of both columns with N_{80} is very good, and the slopes are near 1 indicating that at 0.4% the critical diameter is around 80 nm. Comparing slopes for the different sites, the clean marine sites (ANX, ASI) have slopes < 1 , likely associated with critical diameters less than 80 nm.

Usage Notes

The harmonized csv format and consistent date stamp (YYYY-MM-DD hh:mm:ss) across instrument files should make the files readily usable in most software packages (R, Python, Matlab, IDL, Igor...) as well as more basic spreadsheet tools like Microsoft Excel. As noted above, each file type for each station comes with a README file describing the data and includes other metadata about the site and the instruments. The harmonized file names and contents are summarized in Table 6.

The time series files of the variables are designed so that they can be used without further processing. We have also provided calculated values used in the closure analyses (e.g., mass_SMPS_STP and mass_ACSM_PSAP_STP) as well as parameters related to CCN (AF , D_{crit} and κ_{CCN}) to simplify analyses related to CCN processes. The multi-year SGP data set enables some assessment of inter-annual variability and climatology. For the rest of the locations the shorter time period of measurements means the user should assess the climatological representativity of the data by determining whether conditions were 'normal' or 'exceptional' during the time period of the available data using meteorological and other data not included in this harmonized dataset. Tables S3, S4 provide information on how much data would be removed from the SMPS and CCN datasets depending on which closure flags are applied. We recommend the user consider their application when deciding what closure flags to apply. For example, SMPS and CPC data are often flagged during NPF events when CPC concentrations are 50% higher than the SMPS total particle number concentration. This is likely due to the different cut-off sizes of the SMPS and CPC, as well as the effect of the CPC's counting efficiency curve, rather than a malfunction of the instruments.

We encourage the user to use the data for the following types of analyses (this list should be considered a starting point rather than being comprehensive):

- *Phenomenological studies of CCN across a wide variety of site types*: It could be especially interesting to combine the data set described here with a similar dataset developed by Schmale *et al.*²⁶ and extend the analyses reported on in Schmale *et al.*²⁷. The two datasets together provide significantly more global coverage.
- *Development of a machine learning (ML) algorithm for NPF classification using the data sets with 5 min temporal resolution*: Most NPF classification is done visually and can be quite time consuming (e.g.¹⁰⁰), so an ML approach could be useful. This data set, which covers a variety of aerosol and environment types would provide a useful training and evaluation data set.
- *Quantification of the impact of NPF on CCN and identification of the method or methods that work best to describe the impact of NPF on CCN*: There are various methods for doing so^{29,101–103}, but, to our knowledge, no previous comparison of these methods exists.
- *Improvement of processes relating to PNSD, NPF, and CCN in models*: This could be done for global climate models such as was done for CCN⁷³, PNSD¹⁰⁴ and NPF^{17,105} or for more regionally focussed models such as WRF-Chem.

Code availability

Code for processing the data for different instruments was developed by different parts of research team and different programming languages were used. The CCN, PNSD and CPC data were processed in MATLAB. Python was used to process and harmonize the ACSM data. Code for the aerosol optical data harmonized files is written in IDL and also made use of bash shell scripts. All processing code used to create the harmonized datasets is provided alongside the data.

Received: 4 December 2024; Accepted: 31 March 2025;

Published online: 04 June 2025

References

1. Akimoto, H. Global air quality and pollution. *Science* **302**, 1716–1719, <https://doi.org/10.1126/science.1092666> (2003).
2. Masson-Delmotte, V. *et al.* (eds.) *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, UK and New York, NY, USA, 2021).
3. Twomey, S. Pollution and the planetary albedo. *Atmospheric Environ.* (1967) **8**, 1251–1256, [https://doi.org/10.1016/0004-6981\(74\)90004-3](https://doi.org/10.1016/0004-6981(74)90004-3) (1974).
4. Twomey, S. A., Piepgrass, M. & Wolfe, T. L. An assessment of the impact of pollution on global cloud albedo. *Tellus B* **36B**, 356–366, <https://doi.org/10.1111/j.1600-0889.1984.tb00254.x> (1984).
5. Albrecht, B. A. Aerosols, cloud microphysics, and fractional cloudiness. *Science* **245**, 1227–1230, <https://doi.org/10.1126/science.245.4923.1227> (1989).
6. Charlson, R. J. *et al.* Climate forcing by anthropogenic aerosols. *Science* **255**, 423–430, <https://doi.org/10.1126/science.255.5043.423> (1992).
7. IPCC. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013).

8. Seinfeld, J. H. *et al.* Improving our fundamental understanding of the role of aerosol-cloud interactions in the climate system. *Proc. Natl. Acad. Sci.* **113**, 5781–5790, <https://doi.org/10.1073/pnas.1514043113> (2016).
9. Paasonen, P. *et al.* Continental anthropogenic primary particle number emissions. *Atmospheric Chem. Phys.* **16**, 6823–6840, <https://doi.org/10.5194/acp-16-6823-2016> (2016).
10. Després, V. R. *et al.* Primary biological aerosol particles in the atmosphere: a review. *Tellus B: Chem. Phys. Meteorol.* **64**, 15598, <https://doi.org/10.3402/tellusb.v64i0.15598> (2012).
11. Gordon, H. *et al.* Causes and importance of new particle formation in the present-day and preindustrial atmospheres. *J. Geophys. Res. Atmospheres* **122**, 8739–8760, <https://doi.org/10.1002/2017JD026844> (2017).
12. Stolzenburg, D. *et al.* Atmospheric nanoparticle growth. *Rev. Mod. Phys.* **95**, 045002, <https://doi.org/10.1103/RevModPhys.95.045002> (2023).
13. Riipinen, I. *et al.* Organic condensation: A vital link connecting aerosol formation to cloud condensation nuclei (CCN) concentrations. *Atmospheric Chem. Phys.* **11**, 3865–3878, <https://doi.org/10.5194/acp-11-3865-2011> (2011).
14. Nieminen, T. *et al.* Global analysis of continental boundary layer new particle formation based on long-term measurements. *Atmospheric Chem. Phys.* **18**, 14737–14756, <https://doi.org/10.5194/acp-18-14737-2018> (2018).
15. O'Dowd, C. D. *et al.* A dedicated study of new particle formation and fate in the coastal environment (PARFORCE): Overview of objectives and achievements. *J. Geophys. Res. Atmospheres* **107**, PAR 1–1–PAR 1–16, <https://doi.org/10.1029/2001JD000555> (2002).
16. Kerminen, V.-M. *et al.* Atmospheric new particle formation and growth: Review of field observations. *Environ. Res. Lett.* **13**, 103003, <https://doi.org/10.1088/1748-9326/aadf3c> (2018).
17. Spracklen, D. V. *et al.* The contribution of boundary layer nucleation events to total particle concentrations on regional and global scales. *Atmospheric Chem. Phys.* **6**, 5631–5648, <https://doi.org/10.5194/acp-6-5631-2006> (2006).
18. Merikanto, J., Spracklen, D. V., Mann, G. W., Pickering, S. J. & Carslaw, K. S. Impact of nucleation on global CCN. *Atmospheric Chem. Phys.* **9**, 8601–8616, <https://doi.org/10.5194/acp-9-8601-2009> (2009).
19. Pierce, J. R. & Adams, P. J. Uncertainty in global CCN concentrations from uncertain aerosol nucleation and primary emission rates. *Atmospheric Chem. Phys.* **9**, 1339–1356, <https://doi.org/10.5194/acp-9-1339-2009> (2009).
20. Yu, F. & Luo, G. Simulation of particle size distribution with a global aerosol model: Contribution of nucleation to aerosol and CCN number concentrations. *Atmospheric Chem. Phys.* **9**, 7691–7710, <https://doi.org/10.5194/acp-9-7691-2009> (2009).
21. Luo, G. & Yu, F. Sensitivity of global cloud condensation nuclei concentrations to primary sulfate emission parameterizations. *Atmospheric Chem. Phys.* **11**, 1949–1959, <https://doi.org/10.5194/acp-11-1949-2011> (2011).
22. Makkonen, R. *et al.* BVOC-aerosol-climate interactions in the global aerosol-climate model ECHAM5.5-HAM2. *Atmospheric Chem. Phys.* **12**, 10077–10096, <https://doi.org/10.5194/acp-12-10077-2012> (2012).
23. Dunne, E. M. *et al.* Global atmospheric particle formation from CERN CLOUD measurements. *Science* **354**, 1119–1124, <https://doi.org/10.1126/science.aaf2649> (2016).
24. Rosenfeld, D. *et al.* Satellite retrieval of cloud condensation nuclei concentrations by using clouds as CCN chambers. *Proc. Natl. Acad. Sci.* **113**, 5828–5834, <https://doi.org/10.1073/pnas.1514044113> (2016).
25. Rosenfeld, D. *et al.* Global observations of aerosol-cloud-precipitation-climate interactions. *Rev. Geophys.* **52**, 750–808, <https://doi.org/10.1002/2013RG000441> (2014).
26. Schmale, J. *et al.* Collocated observations of cloud condensation nuclei, particle size distributions, and chemical composition. *Sci. Data* **4**, <https://doi.org/10.1038/sdata.2017.3> (2017).
27. Schmale, J. *et al.* Long-term cloud condensation nuclei number concentration, particle number size distribution and chemical composition measurements at regionally representative observatories. *Atmospheric Chem. Phys.* **18**, 2853–2881, <https://doi.org/10.5194/acp-18-2853-2018> (2018).
28. Uin, J. *et al.* Atmospheric Radiation Measurement (ARM) Aerosol Observing Systems (AOS) for surface-based *in situ* atmospheric aerosol and trace gas measurements. *J. Atmospheric Ocean. Technol.* **36**, 2429–2447, <https://doi.org/10.1175/JTECH-D-19-0077.1> (2019).
29. Kalkavouras, P. *et al.* Regional new particle formation as modulators of cloud condensation nuclei and cloud droplet number in the eastern mediterranean. *Atmospheric Chem. Phys.* **19**, 6185–6203, <https://doi.org/10.5194/acp-19-6185-2019> (2019).
30. Petters, M. D. & Kreidenweis, S. M. A single parameter representation of hygroscopic growth and cloud condensation nucleus activity. *Atmospheric Chem. Phys.* **7**, 1961–1971, <https://doi.org/10.5194/acp-7-1961-2007> (2007).
31. Berg, L. K. *et al.* The two-column aerosol project: Phase I—overview and impact of elevated aerosol layers on aerosol optical depth. *J. Geophys. Res. Atmospheres* **121**, 336–361, <https://doi.org/10.1002/2015JD023848> (2016).
32. Jensen, M. P. *et al.* A succession of cloud, precipitation, aerosol, and air quality field experiments in the coastal urban environment. *Bull. Am. Meteorol. Soc.* **103**, 103–105, <https://doi.org/10.1175/BAMS-D-21-0104.1> (2022).
33. Russell, L. *et al.* Eastern Pacific Cloud Aerosol Precipitation Experiment (EPCAPE) Science Plan. Tech. Rep. DOE/SC- ARM-21-009, ARM user facility. <https://doi.org/10.2172/1804710> (2021).
34. Geerts, B. *et al.* The COMBLE Campaign: A Study of Marine Boundary Layer Clouds in Arctic Cold-Air Outbreaks. *Bull. Am. Meteorol. Soc.* **103**, E1371–E1389, <https://doi.org/10.1175/BAMS-D-21-0044.1> (2022).
35. Zuidema, P. *et al.* Smoke and clouds above the Southeast Atlantic: Upcoming field campaigns probe absorbing aerosol's impact on climate. *Bull. Am. Meteorol. Soc.* **97**, 1131–1135, <https://doi.org/10.1175/BAMS-D-15-00082.1> (2016).
36. Varble, A. C. *et al.* Utilizing a Storm-Generating Hotspot to Study Convective Cloud Transitions: The CACTI Experiment. *Bull. Am. Meteorol. Soc.* **102**, E1597–E1620, <https://doi.org/10.1175/BAMS-D-20-0030.1> (2021).
37. Fast, J. D. *et al.* Large spatiotemporal variability in aerosol properties over central Argentina during the CACTI field campaign. *Atmospheric Chem. Phys.* **24**, 13477–13502, <https://doi.org/10.5194/acp-24-13477-2024> (2024).
38. Wood, R. *et al.* Clouds, Aerosols, and Precipitation in the Marine Boundary Layer: An ARM Mobile Facility Deployment. *Bull. Am. Meteorol. Soc.* **96**, 419–440, <https://doi.org/10.1175/BAMS-D-13-00180.1> (2015).
39. Feldman, D. R. *et al.* The Surface Atmosphere Integrated Field Laboratory (SAIL) Campaign. *Bull. Am. Meteorol. Soc.* **104**, E2192–E2222, <https://doi.org/10.1175/BAMS-D-22-0049.1> (2023).
40. Martin, S. T. *et al.* Introduction: Observations and Modeling of the Green Ocean Amazon (GoAmazon2014/5). *Atmospheric Chem. Phys.* **16**, 4785–4797, <https://doi.org/10.5194/acp-16-4785-2016> (2016).
41. Pöhlker, M. L. *et al.* Long-term observations of cloud condensation nuclei in the Amazon rain forest – part 1: Aerosol size distribution, hygroscopicity, and new model parametrizations for CCN prediction. *Atmospheric Chem. Phys.* **16**, 15709–15740, <https://doi.org/10.5194/acp-16-15709-2016> (2016).
42. Thalman, R. *et al.* CCN activity and organic hygroscopicity of aerosols downwind of an urban region in central Amazonia: Seasonal and diel variations and impact of anthropogenic emissions. *Atmospheric Chem. Phys.* **17**, 11779–11801, <https://doi.org/10.5194/acp-17-11779-2017> (2017).
43. Pöhlker, M. L. *et al.* Long-term observations of cloud condensation nuclei over the Amazon rain forest – part 2: Variability and characteristics of biomass burning, long-range transport, and pristine rain forest aerosols. *Atmospheric Chem. Phys.* **18**, 10289–10331, <https://doi.org/10.5194/acp-18-10289-2018> (2018).
44. Wimmer, D. *et al.* Ground-based observation of clusters and nucleation-mode particles in the Amazon. *Atmospheric Chem. Phys.* **18**, 13245–13264, <https://doi.org/10.5194/acp-18-13245-2018> (2018).
45. Zhao, B. *et al.* Impact of urban pollution on organic-mediated new-particle formation and particle number concentration in the Amazon rainforest. *Environ. Sci. & Technol.* **55**, 4357–4367, <https://doi.org/10.1021/acs.est.0c07465> (2021).

46. Almeida, G. P. The role played by the bulk hygroscopicity on the prediction of the cloud condensation nuclei concentration inside the urban aerosol plume in Manaus, Brazil: From measurements to modeled results. *Atmospheric Environ.* **295**, 119517, <https://doi.org/10.1016/j.atmosenv.2022.119517> (2023).
47. Zhao, B. *et al.* Formation process of particles and cloud condensation nuclei over the Amazon rainforest: The role of local and remote new-particle formation. *Geophys. Res. Lett.* **49**, e2022GL100940, <https://doi.org/10.1029/2022GL100940> (2022).
48. Shupe, M. D. *et al.* Overview of the MOSAiC expedition: Atmosphere. *Elem. Sci. Anthropocene* **10**, <https://doi.org/10.1525/elementa.2021.00060>. 00060 (2022).
49. Mace, J. *et al.* STORMVEX: The Storm Peak Lab Cloud Property Validation Experiment Science and Operations Plan. Tech. Rep., DOE Office of Science Atmospheric Radiation Measurement (ARM) Program. <https://doi.org/10.2172/989263> (2010).
50. Marinescu, P. J., Levin, E. J. T., Collins, D., Kreidenweis, S. M. & van den Heever, S. C. Quantifying aerosol size distributions and their temporal variability in the Southern Great Plains, USA. *Atmospheric Chem. Phys.* **19**, 11985–12006, <https://doi.org/10.5194/acp-19-11985-2019> (2019).
51. Jefferson, A. Empirical estimates of CCN from aerosol optical properties at four remote sites. *Atmospheric Chem. Phys.* **10**, 6855–6861, <https://doi.org/10.5194/acp-10-6855-2010> (2010).
52. Kuang, C. Scanning mobility particle spectrometer instrument handbook. DOE/SC-ARM-TR-147, ARM Climate Research Facility. <https://doi.org/10.2172/1245993> (2016).
53. Uin, J. Cloud condensation nuclei particle counter instrument handbook, DOE/SC-ARM-TR-168, U.S. Department of Energy, Atmospheric Radiation Measurement user facility (2022).
54. Roberts, G. & Nenes, A. A continuous-flow streamwise thermal-gradient CCN chamber for atmospheric measurements. *Aerosol Sci. Technol.* **39**, 206–221, <https://doi.org/10.1080/027868290913988> (2005).
55. Lance, S., Medina, J., Smith, J. & Nenes, A. Mapping the operation of the DMT continuous flow CCN counter. *Aerosol Sci. Technol.* **40**, 242–254, <https://doi.org/10.1080/02786820500543290> (2006).
56. Kuang, C. Condensation particle counter (CPC) instrument handbook. DOE/SC-ARM-TR-145, U.S. Department of Energy, Atmospheric Radiation Measurement user facility (2016).
57. Watson, T. B. Aerosol chemical speciation monitor (ACSM) instrument handbook. Tech. Rep. DOE/SC-ARM-TR-196, U.S. Department of Energy, Brookhaven National Laboratory (2017).
58. Shilling, J. E. & Levin, M. S. Aerosol chemical speciation monitor (ACSM) composition-dependent collection efficiency (CDCE) value-added product report. Tech. Rep. DOE/SC-ARM-TR-271, U.S. Department of Energy, Pacific Northwest National Laboratory (2021).
59. Alfarra, M. R. *et al.* Characterization of urban and rural organic particulate in the Lower Fraser Valley using two Aerodyne aerosol mass spectrometers. *Atmospheric Environ.* **38**, 5745–5758, <https://doi.org/10.1016/j.atmosenv.2004.01.054>. The Pacific 2001 Air Quality Study (2004).
60. Allan, J. D. *et al.* Submicron aerosol composition at Trinidad Head, California, during ITCT 2K2: Its relationship with gas phase volatile organic carbon and assessment of instrument performance. *J. Geophys. Res. Atmospheres* **109**, <https://doi.org/10.1029/2003JD004208> (2004).
61. Ann M. Middlebrook, J. L. J., Roya Bahreini & Canagaratna, M. R. Evaluation of composition-dependent collection efficiencies for the Aerodyne aerosol mass spectrometer using field data. *Aerosol Sci. Technol.* **46**, 258–271, <https://doi.org/10.1080/02786826.2011.620041> (2012).
62. Anderson, T. *et al.* Performance characteristics of a high-sensitivity, three-wavelength, total scatter/backscatter nephelometer. *J. Atmospheric Ocean. Technol.* **13**, 967–986, 10.1175/1520-0426(1996)013<0967:PCOAH>2.0.CO;2 (1996).
63. Anderson, T. & Ogren, J. Determining aerosol radiative properties using the TSI 3563 integrating nephelometer. *Aerosol Sci. Technol.* **29**, 57–69 (1998).
64. Bond, T. C., Anderson, T. L. & Campbell, D. Calibration and intercomparison of filter-based measurements of visible light absorption by aerosols. *Aerosol Sci. Technol.* **30**, 582–600, <https://doi.org/10.1080/027868299304435> (1999).
65. Virkkula, A. *et al.* Modification, calibration and a field test of an instrument for measuring light absorption by particles. *Aerosol Sci. Technol.* **39**, 68–83, <https://doi.org/10.1080/027868290901963> (2005).
66. Flynn, C., Chand, D., Ermold, B. & Koontz, A. The ARM aerosol optical properties (AOP) value-added product. Tech. Rep. DOE/SC-ARM-TR-211, U.S. Department of Energy, Atmospheric Radiation Measurement user facility (2020).
67. Virkkula, A. Correction of the calibration of the 3-wavelength particle soot absorption photometer (3 λ PSAP). *Aerosol Sci. Technol.* **44**, 706–712, <https://doi.org/10.1080/02786826.2010.482110> (2010).
68. Ogren, J. A. Comment on “Calibration and intercomparison of filter-based measurements of visible light absorption by aerosols”. *Aerosol Sci. Technol.* **44**, 589–591, <https://doi.org/10.1080/02786826.2010.482111> (2010).
69. Uin, J. Nephelometer instrument handbook. Tech. Rep. DOE/SC-ARM-TR-165, U.S. Department of Energy, Atmospheric Radiation Measurement user facility (2016).
70. Springston, S. R. Particle soot absorption photometer (PSAP) instrument handbook. Tech. Rep. DOE/SC-ARM-TR-176, U.S. Department of Energy, Atmospheric Radiation Measurement user facility (2018).
71. Lathem, T. L. & Nenes, A. Water vapor depletion in the DMT continuous-flow CCN chamber: Effects on supersaturation and droplet growth. *Aerosol Sci. Technol.* **45**, 604–615, <https://doi.org/10.1080/02786826.2010.551146> (2011).
72. Takegawa, N. & Sakurai, H. Laboratory evaluation of a TSI condensation particle counter (model 3771) under airborne measurement conditions. *Aerosol Sci. Technol.* **45**, 272–283, <https://doi.org/10.1080/02786826.2010.532839> (2011).
73. Fanourgakis, G. S. *et al.* Evaluation of global simulations of aerosol particle and cloud condensation nuclei number, with implications for cloud droplet formation. *Atmospheric Chem. Phys.* **19**, 8591–8617, <https://doi.org/10.5194/acp-19-8591-2019> (2019).
74. Gliß, J. *et al.* AeroCom phase III multi-model evaluation of the aerosol life cycle and optical properties using ground- and space-based remote sensing as well as surface *in situ* observations. *Atmospheric Chem. Phys.* **21**, 87–128, <https://doi.org/10.5194/acp-21-87-2021> (2021).
75. Rose, C. *et al.* Seasonality of the particle number concentration and size distribution: A global analysis retrieved from the network of global atmosphere watch (GAW) near-surface observatories. *Atmospheric Chem. Phys.* **21**, 17185–17223, <https://doi.org/10.5194/acp-21-17185-2021> (2021).
76. Shen, Y. *et al.* Estimating cloud condensation nuclei number concentrations using aerosol optical properties: role of particle number size distribution and parameterization. *Atmospheric Chem. Phys.* **19**, 15483–15502, <https://doi.org/10.5194/acp-19-15483-2019> (2019).
77. Laj, P. *et al.* A global analysis of climate-relevant aerosol properties retrieved from the network of global atmosphere watch (GAW) near-surface observatories. *Atmospheric Meas. Tech.* **13**, 4353–4392, <https://doi.org/10.5194/amt-13-4353-2020> (2020).
78. Ervens, B. *et al.* CCN predictions using simplified assumptions of organic aerosol composition and mixing state: A synthesis from six different locations. *Atmospheric Chem. Phys.* **10**, 4795–4807, <https://doi.org/10.5194/acp-10-4795-2010> (2010).
79. Ren, J. *et al.* The NPF effect on CCN number concentrations: A review and re-evaluation of observations from 35 sites worldwide. *Geophys. Res. Lett.* **48**, e2021GL095190, <https://doi.org/10.1029/2021GL095190> (2021).
80. Sisterson, D. A unified approach for reporting ARM measurement uncertainties technical report: updated in 2016. Tech. Rep. DOE/SC-ARM-17-010, U.S. Department of Energy, Argonne National Laboratory (2017).

81. Sherman, J. P. *et al.* A multi-year study of lower tropospheric aerosol variability and systematic relationships from four North American regions. *Atmospheric Chem. Phys.* **15**, 12487–12517, <https://doi.org/10.5194/acp-15-12487-2015> (2015).
82. Vogelmann, A. M. *et al.* RACORO extended-term aircraft observations of boundary layer clouds. *Bull. Am. Meteorol. Soc.* **93**, 861–878, <https://doi.org/10.1175/BAMS-D-11-00189.1> (2012).
83. Jurányi, Z. *et al.* A 17 month climatology of the cloud condensation nuclei number concentration at the high alpine site Jungfraujoch. *J. Geophys. Res. Atmospheres* **116**, <https://doi.org/10.1029/2010JD015199> (2011).
84. Köhler, H. The nucleus in and the growth of hygroscopic droplets. *Transactions Faraday Soc.* **32**, 1152–1161, <https://doi.org/10.1039/TF9363201152> (1936).
85. Zanutta, M. *et al.* A European aerosol phenomenology-5: Climatology of black carbon optical properties at 9 regional background sites across Europe. *Atmospheric Environ.* **145**, 346–364, <https://doi.org/10.1016/j.atmosenv.2016.09.035> (2016).
86. Rex, M. Links to master tracks in different resolutions of POLARSTERN cruise PS122/1, Tromsø - Arctic Ocean, 2019-09-20 - 2019-12-13 (Version 2), <https://doi.org/10.1594/PANGAEA.924668> (2020).
87. Haas, C. Links to master tracks in different resolutions of POLARSTERN cruise PS122/2, Arctic Ocean - Arctic Ocean, 2019-12-13 - 2020-02-24 (Version 2), <https://doi.org/10.1594/PANGAEA.924674> (2020).
88. Kanzow, T. Links to master tracks in different resolutions of POLARSTERN cruise PS122/3, Arctic Ocean - Longyearbyen, 2020-02-24 - 2020-06-04 (Version 2), <https://doi.org/10.1594/PANGAEA.924681> (2020).
89. Rex, M. Master tracks in different resolutions of POLARSTERN cruise PS122/4, Longyearbyen - Arctic Ocean, 2020-06-04 - 2020-08-12, <https://doi.org/10.1594/PANGAEA.926829> (2021).
90. Rex, M. Master tracks in different resolutions of POLARSTERN cruise PS122/5, Arctic Ocean - Bremerhaven, 2020-08-12 - 2020-10-12, <https://doi.org/10.1594/PANGAEA.926910> (2021).
91. Andrews, E. *et al.* Harmonized aerosol size distribution, cloud condensation nuclei, chemistry and optical properties at 12 sites, <https://doi.org/10.6084/m9.figshare.27913806>.
92. Minguillón, M. C. *et al.* Chemical characterization of submicron regional background aerosols in the western Mediterranean using an aerosol chemical speciation monitor. *Atmospheric Chem. Phys.* **15**, 6379–6391, <https://doi.org/10.5194/acp-15-6379-2015> (2015).
93. Kassianov, E. *et al.* Airborne aerosol *in situ* measurements during TCAP: A closure study of total scattering. *Atmosphere* **6**, 1069–1101, <https://doi.org/10.3390/atmos6081069> (2015).
94. Lance, S. *et al.* Cloud condensation nuclei activity, closure, and droplet growth kinetics of Houston aerosol during the Gulf of Mexico Atmospheric Composition and Climate Study (GoMACCS). *J. Geophys. Res. Atmospheres* **114**, <https://doi.org/10.1029/2008JD011699> (2009).
95. Rissman, T. A. *et al.* Characterization of ambient aerosol from measurements of cloud condensation nuclei during the 2003 Atmospheric Radiation Measurement Aerosol Intensive Observational Period at the Southern Great Plains site in Oklahoma. *J. Geophys. Res. Atmospheres* **111**, <https://doi.org/10.1029/2004JD005695> (2006).
96. Andrews, E. *et al.* Comparison of methods for deriving aerosol asymmetry parameter. *J. Geophys. Res. Atmospheres* **111**, <https://doi.org/10.1029/2004JD005734> (2006).
97. Boyer, M. *et al.* A full year of aerosol size distribution data from the central Arctic under an extreme positive Arctic Oscillation: insights from the Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAIC) expedition. *Atmospheric Chem. Phys.* **23**, 389–415, <https://doi.org/10.5194/acp-23-389-2023> (2023).
98. Bohren, C. & Huffman, D. R. *Absorption and Scattering of Light by Small Particles* (Wiley Science Paperback Series, 1998).
99. Fierz-Schmidhauser, R. *et al.* Measured and predicted aerosol light scattering enhancement factors at the high alpine site Jungfraujoch. *Atmospheric Chem. Phys.* **10**, 2319–2333, <https://doi.org/10.5194/acp-10-2319-2010> (2010).
100. Dal Maso, M., Kulmala, M., Riipinen, I. & Wagner, R. Formation and growth of fresh atmospheric aerosols: Eight years of aerosol size distribution data from SMEAR II, Hyytiälä, Finland. *Boreal Environ. Res.* **10**, 323–336 (2005).
101. Hirshorn, N. S. *et al.* Seasonal significance of new particle formation impacts on cloud condensation nuclei at a mountaintop location. *Atmospheric Chem. Phys.* **22**, 15909–15924, <https://doi.org/10.5194/acp-22-15909-2022> (2022).
102. Damato de España, C. *et al.* Long-term quantitative field study of new particle formation (NPF) events as a source of cloud condensation nuclei (CCN) in the urban background of Vienna. *Atmospheric Environ.* **164**, 289–298, <https://doi.org/10.1016/j.atmosenv.2017.06.001> (2017).
103. Rose, C. *et al.* CCN production by new particle formation in the free troposphere. *Atmospheric Chem. Phys.* **17**, 1529–1541, <https://doi.org/10.5194/acp-17-1529-2017> (2017).
104. Mann, G. W. *et al.* Intercomparison and evaluation of global aerosol microphysical properties among AeroCom models of a range of complexity. *Atmospheric Chem. Phys.* **14**, 4679–4713, <https://doi.org/10.5194/acp-14-4679-2014> (2014).
105. Spracklen, D. V. *et al.* Explaining global surface aerosol number concentrations in terms of primary emissions and particle formation. *Atmospheric Chem. Phys.* **10**, 4775–4793, <https://doi.org/10.5194/acp-10-4775-2010> (2010).
106. Wang, J. *et al.* Aerosol and Cloud Experiments in the Eastern North Atlantic (ACE-ENA). *Bull. Am. Meteorol. Soc.* **103**, E619–E641, <https://doi.org/10.1175/BAMS-D-19-0220.1> (2022).
107. Almeida, G. P., Bittencourt, A. T., Evangelista, M. S., Vieira-Filho, M. S. & Fornaro, A. Characterization of aerosol chemical composition from urban pollution in Brazil and its possible impacts on the aerosol hygroscopicity and size distribution. *Atmospheric Environ.* **202**, 149–159, <https://doi.org/10.1016/j.atmosenv.2019.01.024> (2019).
108. Sheridan, P., Delene, D. & Ogren, J. Four years of continuous surface aerosol measurements from the Department of Energy's Atmospheric Radiation Measurement Program Southern Great Plains Cloud and Radiation Testbed site. *J. Geophys. Res.* **106**, 20735–20747, <https://doi.org/10.1029/2001JD000785> (2001).
109. Kuang, C., Singh, A. & Howie, J. Scanning mobility particle sizer (aossmps) dataset, <https://doi.org/10.5439/1476898>.
110. Uin, J., Salwen, C. & Enekwizu, O. Cloud condensation nuclei particle counter (aoscnc1col) dataset, <https://doi.org/10.5439/1984587>.
111. Koontz, A. & Uin, J. Cloud condensation nuclei particle counter (aoscnc2cola) dataset, <https://doi.org/10.5439/1323892>.
112. Koontz, A. & Uin, J. Cloud condensation nuclei particle counter (aoscnc2colb) dataset, <https://doi.org/10.5439/1323893>.
113. Koontz, A. *et al.* Cloud condensation nuclei particle counter (aoscnc2colasptra), <https://doi.org/10.5439/1323896>.
114. Uin, J., Salwen, C. & Senum, G. Nephelometer (aonephdry1m) dataset, <https://doi.org/10.5439/1984586>.
115. Koontz, A. & Flynn, C. Aerosol optical properties (aoppsap1flynn1m) dataset, <https://doi.org/10.5439/1369240>.
116. Wiedensohler, A. *et al.* Mobility particle size spectrometers: Calibration procedures and measurement uncertainties. *Aerosol Sci. Technol.* **52**, 146–164, <https://doi.org/10.1080/02786826.2017.1387229> (2018).
117. Uin, J., Andrews, E., Enekwizu, O., Hayes, C. & Salwen, C. Cloud condensation nuclei particle counter (aoscnc1colsptra), <https://doi.org/10.5439/1342134>.
118. Crenn, V. *et al.* ACTRIS ACSM intercomparison – part 1: Reproducibility of concentration and fragment results from 13 individual quadrupole aerosol chemical speciation monitors (Q-ACSM) and consistency with co-located instruments. *Atmospheric Meas. Tech.* **8**, 5063–5087, <https://doi.org/10.5194/amt-8-5063-2015> (2015).
119. Ogren, J. A., Wendell, J., Andrews, E. & Sheridan, P. J. Continuous light absorption photometer for long-term studies. *Atmospheric Meas. Tech.* **10**, 4805–4818, <https://doi.org/10.5194/amt-10-4805-2017> (2017).

Acknowledgements

First and foremost we would like to acknowledge the DOE/ARM mentors for providing help with data issues. These include: Janek Uin for help with getting CCN data file issues resolved; Connor Flynn and Gourihar Kulkarni for thoughts on CCN files; Maria Zawadowicz for help with the ACSM; and Ashish Singh for clarity on SMPS RH. At University of Utah, Megan Ostlie and Maria Garcia were invaluable for helping with data logistics. Fangqun Yu from SUNY Albany was also a helpful sounding board. This effort was supported through DOE/ASR funding via grant number DE-SC0022886.

Author contributions

E.A. coordinated the writing of the manuscript and dataset development and processed the aerosol optical data, I.Z. processed the CCN, PNSD and CPC data with much advice from G.T. and J.A.C.V. and did much of the writing related to those datasets. G.C.-C. processed the ACSM with support from A.G.H. and wrote the section describing the ACSM data. I.Z. and G.T. performed the closure analyses on the final harmonized datasets. G.T. also did much of the initial writing. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-04931-y>.

Correspondence and requests for materials should be addressed to E.A. or G.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025