https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

A 21-year global dataset of particle number concentrations for aerosol-cloud interaction studies

Aino Ovaska[1], Daniel Holmberg[2], Elio Rauth[2,3], Mansour Alghamdi[4], Paulo Artaxo[5], Eija Asmi[6], John Backman[6], Benjamin Bergmans[7], Matthew Boyer[1,8], Liezl Bredenkamp[9], Maurizio Busetto[10], Delano Campos De Oliveira[11], Juan Andrés Casquero-Vera[12], Darius Ceburnis[13], Tak Chan[14], Tommy Chan[1], Sebastien Conil[15], Daniele Contini[16], Suzanne Crumeyrolle[17], Valentin Duflot[18], Kostas Eleftheriadis[19], Johan Esveld[20], Ekaterina Ezhova[1], Markus Fiebig[18], Shahzad Gani[1,21], Olga Garmash[1,22], Francisco Gómez-Moreno[23], Roy M. Harrison[24], Andras Hoffer[25], Rakesh K. Hooda[6], Antti Hyvärinen[6], Tareq Hussein[1,26], Jorma Joutsensaari[27], Nikos Kalivitis[28,29], Heinz Kaminski[30], Jutta Kesti[6], Radovan Krejci[31,32], Adam Kristensson[33], Chongai Kuang[11], Markku Kulmala[1,34,35,36], Lauri Laakso[6,9], Ari Leskinen[37], Heikki Lihavainen[6,8], Andreas Massling[38], Maik Merkel[39], Steffen Noe[40], Jakub Ondracek[41], Noemí Perez[42], Jean-Eudes Petit[43], Tuukka Petäjä[1], Michael Pikridas[44], Christopher Pöhlker[45], Mira Pöhlker[39,46], Jean-Philippe Putaud[47], Ximeng Qi[35,36], Cristina Reche[42], Sergio Rodríguez[48,49], Petr Roztocil[41], Jean Sciare[44], Karine Sellegri[50], Dongjie Shang[51], Ashish Singh[11], Mikko Sipilä[1], Henrik Skov[38], Mar Sorribas[52], Tamanna Subba[11], Junying Sun[53], Peter Tunved[32], Ville Vakkari[6,9], Pieter G. Van Zyl[9], Aki Virkkula[6], Jens Voigtländer[39], Kay Weinhold[39], Alfred Wiedensohler[39], Hee-Jung Yoo[54], Putian Zhou[1, 6], Kai Puolamäki[1,2], Tuomo Nieminen[1,55], Veli-Matti Kerminen[1], Victoria A. Sinclair[1,✞], and Pauli Paasonen[1]

✞deceased[1]Institute for Atmospheric and Earth System Research (INAR)/Physics, University of Helsinki, Finland
[2]Department of Computer Science, University of Helsinki, Finland
[3]Department of Remote Sensing, Institute of Geography and Geology, University of Würzburg, Germany
[4]Department of Environment, Faculty of Environmental Sciences, King Abdulaziz University, Jeddah, Saudi Arabia
[5]Institute of Physics, University of São Paulo, Brazil
[6]Finnish Meteorological Institute (FMI), Erik Palménin aukio 1, 00560, Helsinki, Finland
[7]Institut Scientifique de Service Public (ISSeP), Liege, Belgium
[8]SIOS-Knowledge Centre, Longyearbyen, Norway
[9]Atmospheric Chemistry Research Group, Chemical Resource Beneficiation, North-West University, Potchefstroom, South Africa
[10]Institute of Atmospheric Sciences and Climate, National Research Council of Italy (CNR-ISAC), Bologna, Italy
[11]Environmental Science and Technologies Department, Brookhaven National Laboratory, U.S.A.
[12]Andalusian Institute for Earth System Research (IISTA-CEAMA), University of Granada, Autonomous Government of Andalusia, Granada, Spain
[13]School of Natural Sciences, Physics Unit, Ryan Institute's Centre for Climate and Air Pollution Studies, University of Galway, Ireland
[14]Atmospheric Composition Monitoring : Rural and Remote Regions, State of the Environment Directorate, Science and Technology Branch, Environment and Climate Change Canada
[15]Agence nationale pour la gestion des déchets radioactifs, France
[16]Institute of Atmospheric Sciences and Climate, National Research Council of Italy (CNR-ISAC), U.O.S. di Lecce, Lecce, Italy
[17]Univ. Lille, CNRS, UMR 8518 – LOA – Laboratoire d'Optique Atmosphérique, F-59000 Lille, France
[18]Centre for Atmospheric Data, NILU, Kjeller, Norway

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

[19]Environmental Radioactivity Laboratory, Institute of Nuclear and Radiological Science & Technology, Energy & Safety, N.C.S.R. "Demokritos", Attiki, Greece
[20]TNO, the Netherlands Organisation for Applied Scientific Research, Utrecht, the Netherlands
[21]Centre for Atmospheric Sciences, Indian Institute of Technology Delhi, New Delhi, India
[22]Department of Chemistry, University of Copenhagen, Copenhagen, Denmark
[23]Centre for Energy, Environmental and Technological Research (CIEMAT), Madrid, Spain
[24]School of Geography, Earth & Environmental Sciences, University of Birmingham, United Kingdom
[25]HUN-REN–PE Air Chemistry Research Group, University of Pannonia, Hungary
[26]University of Jordan, School of Science, Department of Physics, Environmental and Atmospheric Research Laboratory (EARL), Amman, 11942 Jordan
[27]Department of Technical Physics, University of Eastern Finland, Kuopio, Finland
[28]Department of Chemistry, University of Crete, Greece
[29]Institute for Environmental Research and Sustainable Development, National Observatory of Athens, Athens, Greece
[30]Institute of Energy and Environmental Technology (IUTA), Germany
[31]Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden
[32]Department of Environmental Science, Stockholm University, Stockholm, Sweden
[33]Department of Physics, Lund University, Lund, Sweden
[34]Aerosol and Haze Laboratory, Beijing Advanced Innovation Center for Soft Matter Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, China
[35]Nanjing-Helsinki Institute in Atmospheric and Earth System Sciences, Nanjing University, Nanjing, China
[36]School of Atmospheric Sciences, Nanjing University, Nanjing, China
[37]Finnish Meteorological Institute, Yliopistonranta 8, 70210, Kuopio, Finland
[38]Department of Environmental Science, Faculty of Technical Sciences, Aarhus University, Roskilde, Denmark
[39]Leibniz Institute for Tropospheric Research (TROPOS), Leipzig, Germany
[40]Institute of Forestry and Engineering, Estonian University of Life Sciences, Tartu, Estonia
[41]Research Group of Aerosol Chemistry and Physics, Institute of Chemical Process Fundamentals of the CAS, Prague, Czech Republic
[42]Institute of Environmental Assessment and Water Research (IDAEA), CSIC, Barcelona, Spain
[43]Laboratory for Climate and Environmental Sciences, Pierre-Simon Laplace Institute, France
[44]Climate and Atmosphere Research Center (CARE-C), The Cyprus Institute, Nicosia, Cyprus
[45]Multiphase Chemistry Dept., Max Planck Institute for Chemistry, Mainz, Germany
[46]Faculty of Physics and Earth Sciences, Leipzig Institute for Meteorology, Leipzig University, 04103 Leipzig, Germany
[47]Joint Research Centre, European Commission, Ispra, Italy
[48]Izaña Atmospheric Research Centre, AEMET, Tenerife, Spain
[49]Group of Atmosphere, Aerosols and Climate, IPNA CSIC, Tenerife, Spain
[50]Laboratoire de Météorologie Physique LaMP/CNRS, Université Blaise Pascal, Aubiere, France
[51]State Key Laboratory of Regional Environment and Sustainability, College of Environmental Sciences and Engineering, Peking University, China
[52]Centro Experim "El Arenosillo", National Institute of Aerospace Technology, Huelva, Spain
[53]Chinese Academy of Meteorological Sciences, Beijing, China
[54]Global Atmospheric Watch and Research Division, National Institute of Meteorological Sciences, Jeju-do, Korea
[55]Department of Physics, University of Helsinki, Finland

Correspondence: Aino Ovaska (aino.ovaska@helsinki.fi)

Abstract. Aerosol particles larger than roughly 50-100 nm in diameter are climatically important because they can act as cloud condensation nuclei (CCN), making their global number concentrations essential for understanding aerosol–cloud interactions. However, observationally constrained, long-term global datasets of particle number concentrations in this size range remain

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

scarce. In this investigation, we present a global dataset of ground-level particle number concentrations for the period 2003–2024, produced by combining in situ observations with a machine-learning approach. The dataset includes two variables: the number concentrations for particles larger than 100 nm (N_{100}) and larger than 50 nm (N_{50}), provided at $0.75° \times 0.75°$ spatial resolution and daily temporal resolution. To generate this dataset, we trained an eXtreme Gradient Boosting (XGB) model using measurements from 62 in situ stations as targets and reanalysis variables as predictors, enabling a data-driven representation of particle number concentrations at the global scale. We evaluated the dataset against independent observations from 12 additional stations. At 2/3 of these stations, the dataset shows good performance, capturing the median concentrations within a factor of 1.5 from the observations. Furthermore, we describe the main characteristics of the dataset in terms of global spatial patterns, temporal variability, and seasonal cycles, and demonstrate its ability to capture long-term trends in particle number concentrations, including both increasing and decreasing tendencies reported in the literature. This work provides the first observation-constrained, machine-learning-based global dataset of N_{50} and N_{100} at daily resolution over two decades, bridging the gap between sparse measurements and computationally expensive process-based models. The dataset, publicly available at https://doi.org/10.5281/zenodo.20202080, offers a valuable resource for evaluating model simulations, improving CCN-related parameterizations, and supporting weather and climate studies without the need for explicit knowledge of the aerosol particle microphysics.

1 Introduction

Atmospheric aerosol particles larger than 50 nm play a key role in the climate system because they influence the Earth's radiative budget both directly and indirectly. These particles are particularly important for aerosol–cloud interactions, which remain one of the largest sources of uncertainty in estimates of radiative forcing (Forster et al., 2021). Particles exceeding roughly 50-100 nm in diameter can activate as cloud condensation nuclei (CCN), with the exact activation size depending on the ambient water-vapour supersaturation (Dusek et al., 2006; Kerminen et al., 2012; Pöhlker et al., 2021). The activated CCN then serve as the seeds around which the cloud droplets form. During cloud formation, the number of CCN present at cloud base determines the size of the cloud droplets in the cloud. This, in turn, governs key cloud properties, including albedo, liquid water path, and lifetime of clouds, ultimately determining how the cloud affects radiative transfer in the atmosphere (e.g., Twomey, 1977; Albrecht, 1989; Forster et al., 2021; Rosenfeld et al., 2023; Stier et al., 2024).

Understanding aerosol-cloud interactions requires reliable quantification of CCN concentrations globally (e.g. Rosenfeld et al., 2014). However, comprehensive CCN measurements are difficult to obtain. Existing in situ datasets for CCN and particle number size distributions provide invaluable information, but they capture only a small portion of the spatial and temporal variability needed for global analyses (Rosenfeld et al., 2014; Schmale et al., 2018).

The global estimates of CCN rely primarily on satellite remote sensing, which is the only method that can provide global observational coverage. However, inferring aerosol number concentrations from satellite retrievals is challenging (e.g. Rosenfeld et al., 2014; Bellouin et al., 2020; Quaas et al., 2020; Rosenfeld et al., 2023). The main limitation of satellite retrievals is that they cannot directly measure aerosol particle number concentrations. Instead, they typically provide optical quantities such as

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

aerosol optical depth (AOD) or aerosol index (AI). Inferring number concentrations from these variables is inherently difficult, because they depend not only on aerosol number concentration but also, for example, on relative humidity and particle size (Quaas et al., 2020; Rosenfeld et al., 2023). Additionally, optical detection of particles with diameters smaller than or even slightly larger than 100 nm is challenging and in many cases impossible (Rosenfeld et al., 2023). They also cannot observe low aerosol concentrations (Rosenfeld et al., 2023). Conventional passive remote sensing methods produce only column-integrated values, but newer lidar-based active remote sensing techniques can retrieve vertically resolved aerosol profiles needed to estimate aerosol loading at cloud-relevant altitudes (Rosenfeld et al., 2023). Furthermore, satellite sensors cannot, or have limited ability to, observe aerosol beneath cloud layers (Quaas et al., 2020; Rosenfeld et al., 2023), limiting their usefulness precisely where cloud-relevant aerosol information would be most valuable.

While measuring CCN-size particles is difficult due to challenges in remote sensing observations and lack of comprehensive in situ measurements, modelling them with Earth System Models is equally challenging. Particles in this size range originate from a variety of natural and anthropogenic sources. Some form through atmospheric new particle formation and subsequently grow until they reach 50-100 nm, while others are emitted directly as particles, either already within this size range or as smaller particles that subsequently grow to this size range (e.g. Morawska et al., 1999; Kerminen et al., 2012; Paasonen et al., 2013). Capturing this growth in models is particularly difficult, as it requires detailed numerical representations of complex aerosol dynamics within Earth system models (Blichner et al., 2021). Fanourgakis et al. (2019) investigated 14 global chemistry transport models and general circulation models, and identified biogenic secondary organic aerosol formation as a major source of uncertainty in the models, along with hygroscopic properties of the organic material in summer, and dry deposition and cloud processing during winter. Compared to observations, the 14 models tended to underestimate CCN concentrations.

Given the challenges outlined above, different approaches are needed to complement existing methods and provide a more complete picture of global CCN concentrations. Several recent studies have advanced this effort, including both satellite-based and alternative techniques.

One example of a new satellite-based method is presented by Choudhury and Tesche (2022), who utilise aerosol retrievals from the CALIOP lidar aboard the CALIPSO satellite. Because CALIOP is an active remote-sensing instrument, it provides vertically resolved aerosol optical properties. Additionally, the retrievals are classified into six different aerosol types. By combining these observations with the aerosol microphysics in the CALIPSO aerosol model, Choudhury and Tesche (2022) derived vertically resolved number concentrations of particles larger than 50 nm (N_{50}) and 100 nm (N_{100}), which can be converted to CCN concentrations using additional assumptions. Building on this framework, Choudhury and Tesche (2023) produced global fields of vertically resolved CCN derived entirely from satellite observations. However, the method still inherits several key limitations of satellite-based measurements (Rosenfeld et al., 2023), and its temporal resolution is restricted to monthly averages.

Choudhury et al. (2025) compared these satellite-based CCN estimates with an entirely independent global dataset developed by Block et al. (2024). The Block et al. approach infers CCN concentrations from mass mixing ratios of different aerosol components, retrieved from reanalysis products, and calculated aerosol hygroscopicity, providing global fields of vertically resolved

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

CCN estimates with daily temporal resolution. The intercomparison in Choudhury et al. (2025) showed strong agreement in the large-scale spatial patterns across the two datasets, demonstrating consistency between these independent methods.

Machine learning approaches have also been used to generate CCN and particle number concentration estimates. Nair and Yu (2020) developed one such model, training a machine-learning model to predict CCN concentrations using 19 predictor variables from chemical transport model simulations. In a follow-up study, Yu et al. (2022) extended this framework to estimate particle number concentrations across the 1.2–120 nm size range inside an atmospheric composition model. More recently, Georgiades et al. (2025) introduced another particle number concentration dataset based on machine learning. Their model was trained directly on particle number concentration measurements and used predictors that combined air quality and meteorological variables from reanalysis data with land-use-related information. The resulting analysis produced global fields of annual mean particle number concentrations with 95% coverage intervals for the period 2010–2019.

These earlier studies provide valuable approaches and datasets for estimating CCN and particle number concentrations beyond the sparse in-situ measurement network. However, there remains a need for global, high–time-resolution estimates that are more directly tied to observations.

In this study, we address this gap by generating a global dataset for N_{100} and N_{50} using a machine learning model. Our work builds on Ovaska et al. (2025), who developed and validated a method for combining in-situ particle number concentration measurements with global reanalysis data to produce global fields of daily N_{100} estimates. While Ovaska et al. (2025) focused on method development, here we apply the approach to produce and release a complete global 21-year dataset. We incorporate a larger set of measurement sites, 74 stations in total, of which we use 62 stations to train an eXtreme Gradient Boosting (XGB) model. With the trained model, we generate global, daily-averaged N_{100} and N_{50} concentrations at ground level on a 0.75°×0.75° grid for the period 2003–2024. We validate the resulting dataset by comparing the modelled values with observations from the remaining 12 measurement sites. Finally, we provide an overview of the dataset, including the global spatial distributions of N_{100} and N_{50}, the temporal variability in the concentrations, and their seasonal characteristics. We also highlight several example regions to demonstrate trends in the modelled particle number concentrations.

2 Data and Methods

2.1 Particle number concentration measurements at the training and holdout stations

For this study, we gathered in situ particle number size distribution measurements from 74 stations (Fig. 1, Table 1). The temporal coverage of the measurements at each station is shown in Figure S1, and it ranges from 174 days to 19.5 years, with median duration of 4.6 years. The measurements were performed either at ground-level or near ground-level using a differential mobility particle sizer (DMPS) (Aalto et al., 2001), a scanning mobility particle sizer (SMPS) (Wiedensohler et al., 2012), or an electrical aerosol spectrometer (EAS) (Tammet et al., 2002) (Table 1). Of these 74 stations, 34 were also used in Ovaska et al. (2025), but we extended their measurement time-series if more data had become available. We also excluded one previously included station (Preila, Lithuania) because, shortly before finalising the updated training dataset, we discovered data-quality uncertainties in the version of the station's data available to us at that time. In addition to the 34 stations included in Ovaska

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Table 1. List of measurement stations included in this study and their information, including station name and country, abbreviation used in the text, coordinates (latitude, longitude) and altitude above sea level in meters, station environment type, instrumentation, and references. The measurements were conducted either with differential mobility particle sizer (DMPS), scanning mobility particle sizer (SMPS), or electrical aerosol spectrometer (EAS). Superscripts: a - coastal site, b - mountainous, c - dataset reference for part of the dataset, d - This reference pertains to measurements at the original Troll station (prior to 2014). The present study uses data from the relocated Trollhaugen station.

Station	Country	Abbre-viation	Lat-itude	Long-itude	Altitude (m a.s.l.)	Environ-ment Type	Instru-ment	Site description	Dataset Reference
Agia Marina	Cyprus	CAO	35.04	33.06	532	rural	smps	Pikridas et al. (2018)	Baalbaki et al. (2021)
Alert	Canada	ALE	82.49	-62.51	75	polar	smps	Leaitch et al. (2013)	Boyer et al. (2023)
Amazonas	Brazil	AMA	-2.15	-59.01	130	remote	smps	Martin et al. (2017)	Martin et al. (2017)
Amman	Jordan	AMM	32.01	35.87	1000	urban	smps	Hussein et al. (2019)	Hussein et al. (2020)
Anmyeon-do	South Korea	AMY	36.54	126.33	58	rural[a]	smps	Bu et al. (2025)	
Annaberg-Buchholz	Germany	ABZ	50.57	12.99	545	urban	smps	Schladitz et al. (2015)	Birmili et al. (2016)
Ascension Island	Ascension Island	ASI	-7.97	-14.35	76	marine	smps	Zuidema et al. (2018)	Singh et al. (2016a)
Aspvreten	Sweden	ASP	58.80	17.38	25	rural	dmps	Tunved and Ström (2019)	Nieminen et al. (2018)
Barcelona (Palau Reial)	Spain	BCN	41.39	2.12	130	urban	smps	in 't Veld et al. (2021)	Carnerero et al. (2021)
Beijing	China	BEI1	39.90	116.40	55	urban	dmps	Wu et al. (2007)	Wu et al. (2007)
Beijing	China	BEI2	39.94	116.30	20	urban	dmps	Liu et al. (2020)	
Birkenes	Norway	BIR	58.38	8.25	190	rural	dmps	Eckhardt et al. (2009)	Yttri et al. (2021)
Botsalano	South Africa	BOT	-25.50	25.80	1400	rural	dmps	Laakso et al. (2008)	Nieminen et al. (2018)
Bösel (Südoldenburg)	Germany	BSL	53.00	7.95	17	rural	smps	Asmi et al. (2011)	Birmili et al. (2016)
Cabauw Zijdeweg	Netherlands	CBW	51.97	4.93	1	rural	smps	Liu et al. (2024)	
Delhi	India	DEL	28.55	77.19	225	urban	smps	Gani et al. (2019)	Gani et al. (2020)
Demokritos (Athens)	Greece	DEM	38.00	23.82	270	urban	smps	Eleftheriadis et al. (2021)	Kecorius et al. (2024)
Dome-C	Antarctica	DMC	-75.10	123.40	3233	polar	dmps	Järvinen et al. (2013)	Nieminen et al. (2018)
Dresden-Wincke	Germany	DRE	51.04	13.73	112	urban	smps	Birmili et al. (2016)	
East Trout Lake	Canada	ETL	54.35	-104.99	493	rural	dmps	Isenor et al. (2025)	
Eastern North Atlantic	Azores	ENA	39.09	-28.03	30	marine	smps	Wang et al. (2022)	Singh et al. (2022)

Continued on next page

Open Access Earth System Science Data Discussions

Continued from previous page

Station	Country	Abbre-viation	Lat-itude	Long-itude	Altitude (m a.s.l.)	Environ-ment Type	Instru-ment	Site description	Dataset Reference
Egbert	Canada	EGB	44.20	-79.80	251	rural	smps	Slowik et al. (2010)	Pierce et al. (2014)
El Arenosillo	Spain	ARN	37.10	-6.73	41	rural[a]	smps	Sorribas et al. (2015)	Sorribas et al. (2015)
Finokalia	Greece	FKL	35.34	25.67	250	remote[a]	smps	Mihalopoulos et al. (1997)	Kalivitis et al. (2019)
Fonovaya	Russia	FON	56.42	84.07	80	rural	dmps	Antonovich et al. (2018)	Lampilahti et al. (2025)
Granada	Spain	UGR	37.16	-3.61	680	urban	smps	Casquero-Vera et al. (2021)	
Gual Pahari	India	GUA	28.43	77.15	243	suburban	dmps	Hyvärinen et al. (2010)	
Gunnison	United States	GUC	38.90	-106.94	2886	remote[b]	smps	Feldman et al. (2023)	Singh et al. (2021)
Hada al Sham	Saudi Arabia	HAD	21.80	39.73	500	rural	dmps	Lihavainen et al. (2016)	
Harwell	United Kingdom	HRW	51.57	-1.33	126	rural	smps	Charron et al. (2007)	Nieminen et al. (2018)
Helsinki	Finland	HEL	60.20	24.97	26	urban	dmps	Järvi et al. (2009)	SmartSMEAR
Hohenpeissenberg	Germany	HPB	47.80	11.00	988	rural	smps	Birmili et al. (2003)	Birmili et al. (2016)
Hyytiälä	Finland	HYY	61.85	24.29	181	rural	dmps	Hari and Kulmala (2005)	SmartSMEAR
Ispra	Italy	IPR	45.81	8.64	209	rural	dmps	Putaud et al. (2021)	
Izana	Spain	IZA	28.30	-16.48	2373	marine[b]	smps	García et al. (2014)	García et al. (2014)
Järvselja	Estonia	JRV	58.27	27.27	36	rural	eas	Noe et al. (2015)	Nieminen et al. (2018)
Košetice	Czech Republic	KCE	49.56	15.08	534	rural	smps	Zíková and Ždímal (2013)	
K-Puszta	Hungary	KPZ	46.97	19.55	125	rural	dmps	Yli-Juuti et al. (2009)	Nieminen et al. (2018)
La Jolla	United States	EPC	32.87	-117.26	7	urban[a]	smps	Russell et al. (2025)	Singh et al. (2023)
La Réunion	Réunion	REU	-21.08	55.38	2160	marine[b]	dmps	Baray et al. (2013)	Foucart et al. (2018)
Lecce	Italy	LEC	40.34	18.12	36	suburban	smps	Merico et al. (2025)	Dinoi et al. (2023)
Mace Head	Ireland	MHD	53.33	-9.90	10	remote[a]	smps	O'Connor et al. (2008)	Nieminen et al. (2018)
Madrid (CIEMAT)	Spain	MAD	40.46	-3.73	669	urban	smps	Barreiro et al. (2017)	Gómez-Moreno et al. (2024)
Maldives Climate Observatory of Hanimaadhoo	Maldives	MCOH	6.78	73.18	2	marine	dmps/ smps	Kesti et al. (2020)	
Manacapuru	Brazil	MAO	-3.21	-60.60	50	remote	smps	Martin et al. (2017)	Singh et al. (2014)
Marikana	South Africa	MAR	-25.70	27.50	1170	urban	dmps	Vakkari et al. (2013)	Nieminen et al. (2018)
Melpitz	Germany	MLP	51.53	12.90	87	rural	dmps	Engler et al. (2007)	Birmili et al. (2016)
Montseny	Spain	MSY	41.77	2.35	720	rural	smps	in 't Veld et al. (2021)	Carnerero et al. (2021)
Mukteshwar	India	MUK	29.40	79.60	2180	remote	dmps	Hooda et al. (2018)	Nieminen et al. (2018)

Continued on next page

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

Continued from previous page

Station	Country	Abbreviation	Latitude	Longitude	Altitude (m a.s.l.)	Environment Type	Instrument	Site description	Dataset Reference
Mülheim-Styrum	Germany	MUL	51.45	6.87	40	suburban	smps	Asbach et al. (2024)	Asbach et al. (2024)
Nanjing	China	NAN	32.10	118.90	40	urban	dmps	Qi et al. (2015)	Chen et al. (2021)
Neuglobsow	Germany	NEU	53.14	13.03	70	rural	smps	UBA (2013)	Birmili2016
Observatoire Perenne de l'Environnement	France	OPE	48.56	5.51	392	rural	smps	Conil et al. (2019)	Farah et al. (2020)
Paris (SIRTA)	France	PAR	48.71	2.16	156	suburban	smps	Petit et al. (2015)	
Prague-Suchdol	Czech Republic	PRA	50.13	14.38	277	suburban	smps	Kubelová et al. (2015)	Suchánková et al. (2025)
Puijo (Kuopio)	Finland	PUI	62.91	27.66	306	suburban	dmps	Leskinen et al. (2009)	SmartSMEAR
San Pietro Capofiume	Italy	SPC	44.65	11.62	11	rural	dmps	Hamed et al. (2007)	Nieminen et al. (2018)
Schauinsland	Germany	SCH	47.91	7.91	1205	rural[b]	smps	Asmi et al. (2011)	Birmili et al. (2016)
Southern Great Plains	United States	SGP	36.61	-97.49	300	rural	smps	Marinescu et al. (2019)	Singh et al. (2016b)
São Paulo	Brazil	SAO	-23.60	-46.60	750	urban	dmps	Backman et al. (2012)	Nieminen et al. (2018)
Tiksi	Russia	TKS	71.60	128.88	10	remote	dmps	Asmi et al. (2016)	
Trollhaugen	Antarctica	TRO	-72.01	2.54	1553	polar	dmps	Hansen et al. (2009)[d]	Fiebig et al. (2014)[d]
United Arab Emirates	United Arab Emirates	UAE	25.23	55.98	165	rural	dmps	Kesti et al. (2022)	Kesti et al. (2024)
Utö	Finland	UTO	59.78	21.38	7	rural[a]	dmps	Engler et al. (2007)	
Värriö	Finland	VAR	67.76	29.61	390	remote	dmps	Hari et al. (1994)	SmartSMEAR
Vavihill	Sweden	VHL	56.01	13.15	172	rural	dmps	Kristensson et al. (2008)	Nieminen et al. (2018)
Villeneuve d'Ascq	France	VDA	50.61	3.14	70	suburban	smps	Crumeyrolle et al. (2023)	Suchánková et al. (2025)
Vielsalm	Belgium	VIE	50.30	6.00	496	rural	smps	ACTRIS (2024)	
Villum Research Station	Greenland	VIL	81.60	-16.67	31	polar	smps	Pernov et al. (2022)	
Waliguan	China	WLI	36.28	100.90	3816	remote[b]	dmps	Kivekäs et al. (2009)	Nieminen et al. (2018)
Waldhof	Germany	WAL	52.80	10.76	75	rural	smps	Asmi et al. (2011)	Birmili et al. (2016)
Welgegund	South Africa	WGD	-26.57	26.94	1480	rural	dmps	Beukes et al. (2025)	Nieminen et al. (2018)
Zeppelin	Svalbard	ZPL	78.91	11.89	473	polar	dmps	Platt et al. (2022)	Karlsson et al. (2021)
Zotino	Russia	ZOT	60.80	89.35	114	remote	dmps	Heintzenberg et al. (2011)	

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

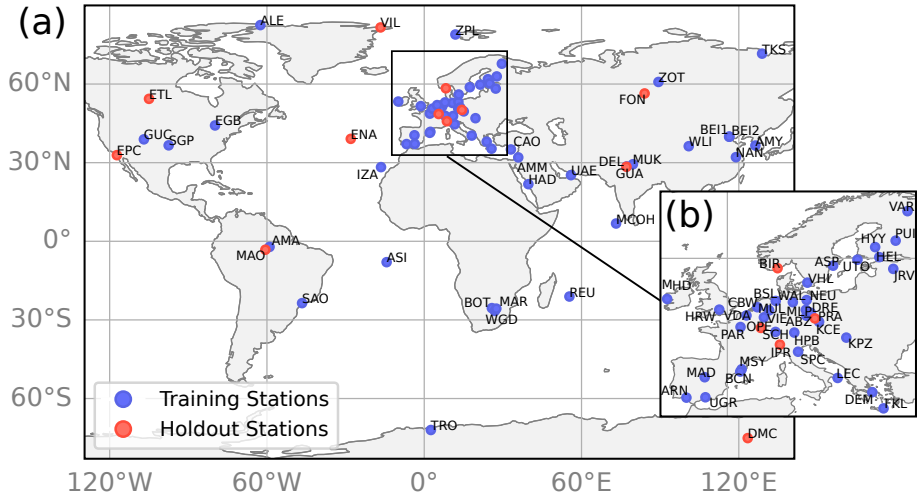

Figure 1. Map of measurement stations. Panel (a) shows the global map, and panel (b) shows a zoom-in over Europe. The station abbreviations are explained in Table 1

et al. (2025), we added 40 new stations to improve global coverage. We especially focused on adding stations in regions that had limited representation in our earlier work, such as polar and marine areas. With the new additional stations, the dataset generated in this study is more representative of the whole globe compared to the results from the previous study.

Of the added stations, we selected 12 to serve as holdout stations (Fig. 1), meaning that we excluded them from XGB model training and instead used them only to evaluate the final dataset (Sect. 3). We selected the holdout stations to represent a range of conditions globally, while also ensuring that their removal would not reduce the representativeness of the training set. In particular, we avoided selecting stations that were the only examples of their environment type. The remaining 62 stations were used as training stations for training the XGB model.

From the particle number size distribution measurements, we calculated the number concentrations of particles with diameters larger than 100 nm (N_{100}) and larger than 50 nm (N_{50}). The upper size limit was defined as the maximum particle diameter measured by the instrument, unless it exceeded 1000 nm, in which case the upper limit was set to 1000 nm. Only time periods for which the maximum measured diameter exceeded 300 nm were included in the analysis. Because the particle number size distribution measurements had a sub-daily time resolution, we calculated the daily-averaged N_{100} and N_{50} for days with at least 21 hours of data using UTC time. We performed a thorough quality control for the calculated N_{100} and N_{50} using quality flags provided in the particle number size distribution data and careful visual inspection. To address any remaining data-quality issues, we removed likely measurement-related outliers by discarding values that were outside three standard deviations from each station's mean. While some of these outliers may be real, we are confident that the positive influence of including such individual short term events on the model performance is negligible in comparison to the negative influence of including outliers caused by instrumental factors.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

2.2 Reanalysis data and input variable processing

For both training the XGB model and producing the global dataset, we used reanalysis variables as inputs to the XGB model. We used 19 variables that are linked to primary particle emissions or known to influence particle formation, growth, transport, dilution, or combinations of these processes. To align with the ground-level N_{100} and N_{50} measurements, we retrieved each variable from the lowest available level in the reanalysis datasets.

We used variables from two reanalysis datasets: the Copernicus Atmosphere Monitoring Service (CAMS) "CAMS global reanalysis (EAC4)" dataset (Inness et al., 2019a, b), and the "ERA5 hourly data on single levels from 1940 to present" dataset (Copernicus Climate Change Service, 2025). From ERA5 reanalysis data, we retrieved boundary layer height. From the CAMS reanalysis, we retrieved mixing ratios for the following aerosol species and trace gases: hydrophilic and hydrophobic organic matter aerosol, hydrophilic and hydrophobic black carbon aerosol, sulfate aerosol, dust aerosol (0.03–0.55 μm), sea salt aerosol (0.03–0.5 μm), carbon monoxide, sulphur dioxide, ammonia, nitrogen monoxide, nitrogen dioxide, isoprene, and terpenes. In addition, we retrieved or derived several meteorological variables from CAMS: specific rain water content, air temperature at 2 m height, relative humidity at 2 m height (derived), and 10-m wind speed (derived).

Reanalysis variables were processed as follows. For XGB model training, we first interpolated the global reanalysis fields to the coordinates of each measurement station (Table 1). We then sampled them to temporally align with the in situ observations. Finally, we computed the derived variables (relative humidity and wind speed), and calculated daily averages in UTC time.

For the global dataset, we applied the same procedure but without spatial interpolation or temporal sampling. Instead, because ERA5 has a finer resolution (0.25°) than CAMS (0.75°), we first regridded ERA5 to the coarser CAMS grid. We then used the full global fields, computed the derived variables, and produced daily averages.

We additionally log10-transformed reanalysis variables that have log-normal distributions. For variables that contained arbitrarily low outlier values much smaller than the rest of the distribution (e.g., 10^{-27}), we shifted the minimum to 10^{-17} because the outliers could be challenging for the machine learning model. Finally, all variables were centered and scaled using the mean and the standard deviation derived from the training data.

For further details about the reanalysis variables and their selection, retrieval, and processing, please refer to Ovaska et al. (2025).

2.3 Machine learning methods and XGB model

To generate the N_{100} and N_{50} global dataset, we trained an XGB regression machine learning model, which is known to perform well on complex, non-linear data. XGB is a decision tree based model that uses gradient boosting, where the model builds many small trees sequentially, and each new tree corrects the errors of the previous ones (Chen and Guestrin, 2016). We used a similar XGB model setup as in Ovaska et al. (2025): we had squared error as the loss function, and used the same XGB model hyperparameters as in Ovaska et al. (2025)), but with no early stopping.

We trained the XGB model using log10-converted daily N_{100} and N_{50} values from the 62 training stations as targets, and the corresponding spatially interpolated, daily averaged, and scaled reanalysis variables as predictors. To ensure that all

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

training stations contributed equally to model training, we weighted each datapoint in the training data with a weight inversely proportional to the number of days of data from the corresponding measurement station (details in Ovaska et al., 2025).

During model training, we evaluated the method using leave-one-station-out cross-validation, following the principles of spatial cross-validation in which the cross-validation folds are defined based on geographical information (e.g. Cho et al., 2020; Beigaitė et al., 2022). In this approach, each training station was iteratively selected as the target station, an intermediate model was trained on the remaining stations, and the resulting model was then tested on the held-out target station. This approach assesses the model's skill in a large variety of environments and offers an independent measure of model reliability. Note that the models that are trained during the leave-one-station-out cross validation are not used to produce the final global datasets.

After training the XGB model with data from the 62 training stations, we generated the final output using the global reanalysis dataset for 2003–2024 as input and validated the results against the 12 manually selected holdout stations (see Sect. 2.1). We generated the output for this time period because the CAMS reanalysis begins in 2003 and, at the time of our analysis, extended to 2024. In the future, the output can be updated to include 2025 and subsequent years as the CAMS reanalysis is extended annually by one additional year.

2.4 Performance metrics and statistical methods

To validate the method and the final model output, we compared the modelled number concentration values against the corresponding observations using the following performance metrics:

- *Root Mean Squared Error*: We evaluated model performance using the root mean squared error (RMSE) between the log10-transformed observed and modelled number concentration values at the measurement stations ($\mathrm{RMSE}_{\log 10}$). The log10-transformation was applied to prioritise obtaining the correct order of magnitude rather than exact values. It also reduces the scale dependence of RMSE, which would otherwise exaggerate errors at high number concentration levels.

- *Pearson correlation coefficient*: We computed the Pearson correlation coefficient (r) for the log10-transformed observed and modelled number concentration values.

- *Mean absolute percentage error*: The mean absolute percentage error (MAPE) is calculated as the absolute difference between the modelled and observed values scaled by the observed value and expressed as a percentage. We calculated this for the absolute number concentrations.

For the statistical analysis of our dataset, we used the following methods:

- *Mann-Kendall test*: To characterise the dataset, we examined long-term trends in particle number concentrations for selected regions. We estimated trends in the annual mean concentrations using the Mann–Kendall test, a widely used non-parametric, rank-based method that detects monotonic trends without assuming any specific data distribution (Kendall, 1938; Mann, 1945). This test is commonly applied in climatology.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

– *Coefficient of variation*: The coefficient of variation expresses the variability of a dataset relative to its mean, calculated as the standard deviation divided by the mean. It provides a dimensionless measure of relative dispersion.

3 Method and Model Output Validation

Although the method was carefully evaluated and the model results thoroughly validated in Ovaska et al. (2025), the substantially expanded station network used in this study required us to verify that the approach remains robust. To this end, we validated the method using leave-one-station-out cross-validation, and discuss these results in this section. Furthermore, we evaluated the final model output by comparing it with the observed particle number concentrations at the holdout stations. This comparison provides examples that we use to discuss the capabilities and limitations of the produced dataset.

3.1 Method validation

The leave-one-station-out cross-validation result confirms that the model reproduces the station-level N_{100} and N_{50} reasonably well, even when the target station is excluded from training (Fig. 2). The modelled log10-converted median concentrations were typically within a factor of 1.5 of the observed log10-converted medians, indicated by the dashed lines in Figure 2. Reproducing the median within a factor of 1.5 can be considered good performance, given that the observed medians span three orders of magnitude (approximately 10-10 000 cm^{-3} for N_{100} and 80-20 000 cm^{-3} for N_{50}). However, at a small number of stations, the XGB model underestimated the medians by more than a factor of 1.5 (10 out of 62 stations for N_{100} and 9 for N_{50}). Most of these stations had mid- to high concentration levels, suggesting that the XGB model may have a tendency to underestimate, especially at higher concentration levels. However, all of these stations are included in the training set of the final model, meaning that the produced global data set is expected to reproduce the concentrations better at these stations, and in other similar environments, than the leave-one-station-out cross-validation.

Figure 2 also shows the observed and modelled 10th–90th percentile ranges of daily concentrations. Collectively, these ranges illustrate the spread of observed versus modelled daily values during leave-one-out cross-validation. For individual stations, however, it is often more informative to examine actual scatterplots of observed-modelled daily values (examples shown in Figure S2), which typically reveal a stronger correlation between the model and observations than might be inferred from the percentile ranges alone. Nevertheless, Figure 2 enables direct comparison of how the modelled range of daily concentrations (vertical) corresponds to the observed range (horizontal) at each station. In a perfect model these ranges would be equal. In practice, the modelled ranges in Figure 2 are often narrower than the observed ones, indicating that the model tends to underrepresent the lowest and highest concentration values at these stations.

3.2 Global model output validation

We validated the final global model output against the 12 holdout stations, which were excluded from training and used only to assess the performance of the produced dataset. The holdout station locations are shown in Figure 1, and Table 2 summarises the global model output validation results at each site.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

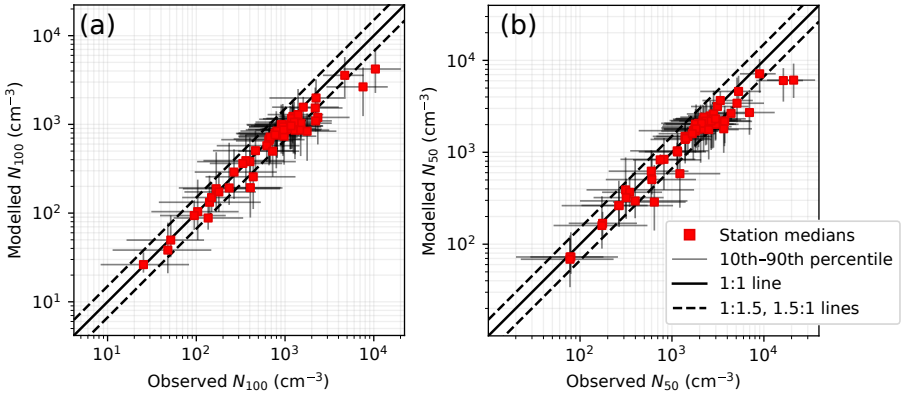

Figure 2. Figure contains the results for leave-one-station-out cross-validation, showing for each training station the median modelled number concentrations against median observed number concentrations, as well as the 10-90th percentile ranges of daily modelled values and observed values for (a) N_{100} and (b) N_{50}.

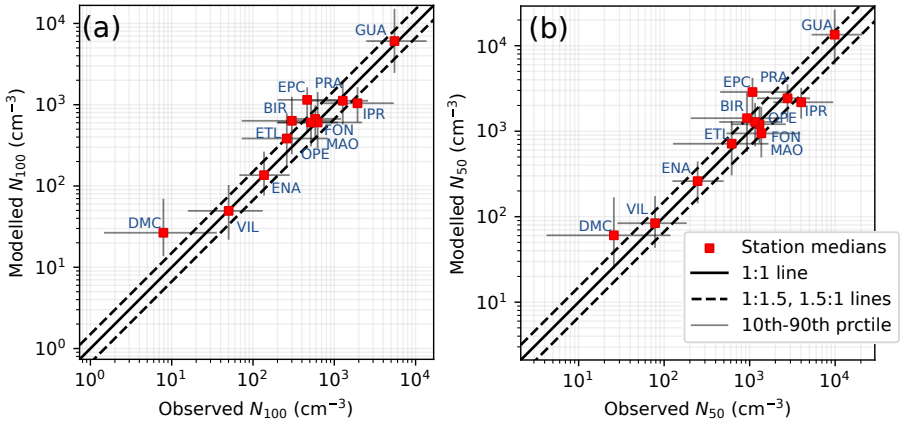

Figure 3. Figure shows the holdout station median estimated number concentrations against median observed number concentrations, as well as the 10-90th percentile ranges of daily model estimates and daily observations for (a) N_{100} and (b) N_{50}.

Figure 3 presents the model output validation results in the same format as the leave-one-out cross-validation results shown in Figure 2, including the observed and modelled medians as well as the 10th–90th percentile ranges for daily concentrations. Focusing first on the medians, most of the modelled medians fall within a factor of 1.5 of the observed medians. This result is promising, as it shows that the model output can capture different concentration levels in varying environments. For both N_{100} and N_{50}, the median modelled concentrations at eight of the 12 holdout stations are within a factor of 1.5 of the corresponding observed medians. Moreover, 11 out of 12 stations fall within a factor of 2.5 of the observed medians, the only exception being Dome-C (DMC) station in Antarctica, which has extremely low concentrations. However, it should be noted that some of the well-performing holdout stations are either geographically nearby training stations, or are otherwise well-represented in the

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

Table 2. Holdout station information and the global model output validation results at the holdout stations. The table lists each station's details (name, abbreviation, country), the error metrics for N_{50} and N_{100}, and a spatiotemporal comparison with the nearest training stations. The second-to-last column shows the geographically closest training station (for station abbreviations, see Table 1), the distance between them, and the percentage of days for which measurements at the holdout station temporally overlap with those at the training station. The last column lists the number of training stations within a 500 km radius of the holdout station and the percentage of days with temporally overlapping measurements with at least one of these nearby stations

Station (abbreviation), Country	RMSE$_{log10}$ (N_{100} / N_{50})	MAPE (%) (N_{100} / N_{50})	Pearson r (N_{100} / N_{50})	Nearest tr. station, distance, (temporal overlap)	No. tr. stations $<$ 500km (temporal overlap)
East Trout Lake (ETL), Canada	0.36 / 0.35	149.0 / 129.7	0.70 / 0.72	GUC, 1720 km (0 %)	0 (0 %)
La Jolla (EPC), USA	0.38 / 0.42	136.8 / 160.1	0.72 / 0.66	GUC, 1140 km (33 %)	0 (0%)
Villum Research Station (VIL), Greenland	0.28 / 0.27	63.3 / 66.2	0.61 / 0.55	ZPL, 610 km (58 %)	0 (0 %)
Fonovaya (FON), Russia	0.20 / 0.20	42.1 / 36.9	0.62 / 0.60	ZOT, 580 km (0 %)	0 (0 %)
Manacapuru (MAO), Brazil	0.22 / 0.26	36.2 / 36.1	0.88 / 0.84	AMA, 210 km (52 %)	1 (52 %)
Eastern North Atlantic (ENA), Azores	0.22 / 0.22	42.3 / 42.0	0.55 / 0.50	IZA, 1600 km (0 %)	0 (0 %)
Dome-C (DMC), Antarctica	0.68 / 0.54	419 / 266	0.78 / 0.84	TRO, 3170 km (0 %)	0 (0 %)
Gual Pahari (GUA), India	0.20 / 0.22	43.0 / 56.4	0.75 / 0.69	DEL, 10 km (0 %)	2 (56 %)
Birkenes (BIR), Norway	0.44 / 0.35	173 / 113	0.74 / 0.77	VHL, 400 km (77 %)	1 (77 %)
Prague–Suchdol (PRA), Czech Republic	0.17 / 0.16	27.4 / 26.2	0.81 / 0.80	KCE, 80 km (26 %)	7 (83 %)
Observatoire Perenne de l'Environnement (OPE), France	0.21 / 0.20	48.3 / 45.0	0.76 / 0.72	SCH, 190 km (23 %)	7 (84 %)
Ispra (IPR), Italy	0.41 / 0.37	48.8 / 46.0	0.43 / 0.37	SCH, 240 km (22 %)	3 (35 %)

training. Table 2 lists the geographically nearest training station with distance and temporal overlap information as well as the number of training stations within 500 km radius. Comparing these against the performance values, we can see that some of the best performing stations do indeed have training stations within 500 km radius with temporal overlap in measurements (e.g., Manacapuru (MAO) in Brazil, Prague-Suchdol in Czech Republic (PRA) and Gual Pahari (GUA) in India).

Focusing next on the percentile ranges, the results in Figure 3 are similar to the leave-one-out results (Figure 2). The modelled daily value 10th–90th percentile ranges in Figure 3 (vertical bars) are often narrower than the observed ranges (horizontal bars), indicating that the model can fail to capture the highest and lowest daily average concentrations. This is also evident in Figure S2, which compares the model output with observations at the holdout stations. However, Figure S2 also shows that the daily values generally fall within the factor-of-1.5 lines, and Table 2 shows that the correlations between the model output and the observations tend to be good.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

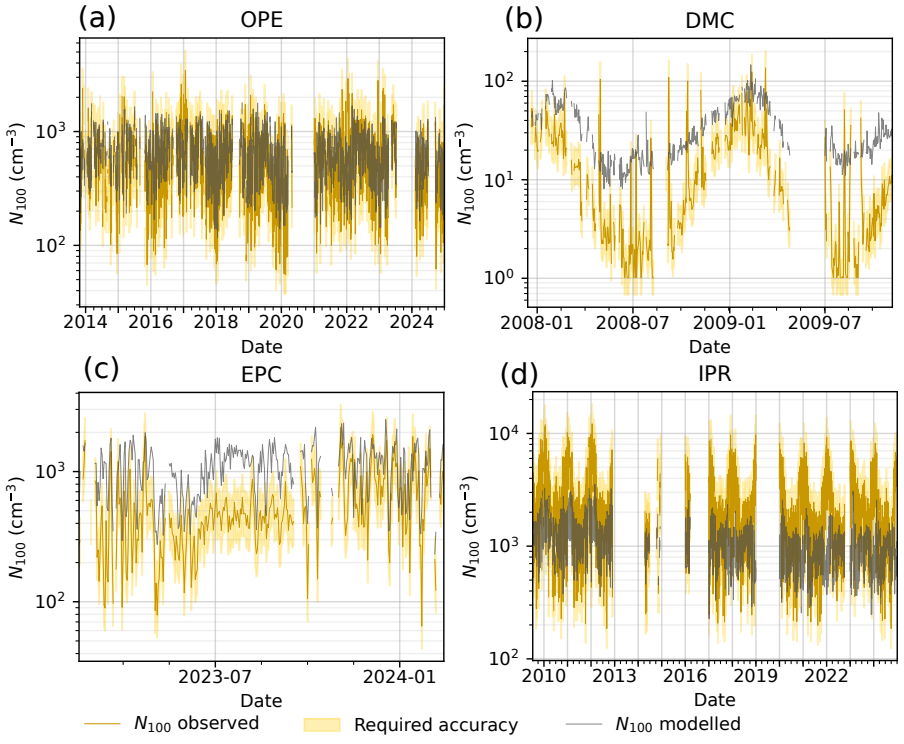

Figure 4. Examples of modelled and observed N_{100} time series at holdout stations (see Table 1 for abbreviations). The rest of the N_{100} time series and all N_{50} time series at the holdout stations are shown in the Suppelement (Figures S3-S4)

We also examined the time series of the observed and modelled number concentrations at the holdout stations. For N_{100}, these are shown in Figures 4 and S3, and the corresponding N_{50} time series are presented in Figure S4. Below, we highlight some of the features in these time series that are relevant for assessing the usability of the dataset across different environments.

At the eight of twelve holdout stations, where the median concentrations fall within the acceptable bounds, the modelled time series reproduce the main characteristics of the observations. One of these stations is Observatoire Perenne de l'Environnement (OPE) in France. OPE is a rural European site with several nearby stations (Table 2) and therefore is well represented within the training dataset. As shown in Figure 3, the median concentration at OPE is captured accurately. In the time series (Figure 4a), the model output does not fully reproduce the lowest and highest daily average values but it does follow the observations overall. This behaviour is typical also for the other well-performing holdout stations (Figures S3-S4).

Despite most of the holdout stations performing well, there are the few stations where the medians are not captured within a factor-of-1.5 from the observed medians. We next discuss a few of them as examples and interpret what they reveal about the limitations of the dataset.

At Dome-C (DMC) in Antarctica (Figure 4b), the model output systematically overestimates concentrations, likely because the training dataset contains no stations with concentrations as low as those observed at Dome-C. Even the only Antarctic

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

training station, Trollhaugen (TRO), has substantially higher concentrations and is also geographically distant (Table 2). This result illustrates how the model can perform poorly in environments that are unrepresented in the training set, particularly when the concentrations fall outside the range used to train the model. This remains the main limitation of the created dataset. It should also be noted that in environments with very low particle concentrations, such as Dome-C, cloud droplet activation may occur for particles with critical diameters smaller than 50 nm. Consequently, using N_{50} as CCN proxy can underestimate CCN concentrations in these regions.

In La Jolla (EPC) in California USA, the model output also systematically overestimates concentrations (Figures 3 and 4c), but the reason is likely different from Dome-C. La Jolla is a coastal site influenced by both clean marine air masses and polluted air masses from surrounding cities and ports (ARM, 2023). This heterogeneous environment can be difficult to capture using gridded reanalysis data with limited spatial resolution. As a result, in the grid cell that contains the measurement station, the reanalysis variable concentrations may be elevated relative to the actual concentrations. This would explain why, in Figure 4c, the modelled time series has very similar day-to-day patterns to the observation, but the summer concentrations are consistently overestimated. However, confirming this hypothesis would require a deeper analysis that is outside the scope of this study.

The model output underestimates concentrations in Ispra (ISP), Italy (Figure 3), despite its proximity to 3 training stations (Table 2). Figure 4d shows that the model output often captures summertime concentrations within the required accuracy, and the underestimation occurs mostly during winter, when the concentrations are high. The high observed winter-time particle number concentrations are related to meteorology and local topography, which promote stagnant conditions that favour both the accumulation of pollutants and the growth of sub-100 nm particles into this size range (Rodríguez et al., 2005). Because reanalysis data cannot capture this kind of local phenomenon, it is not represented in our model output, leading to underestimation. This challenge is similar to the situation in La Jolla, and together these cases suggest that there are locations where the combination of reanalysis data and machine learning cannot capture the local conditions — demonstrating another limitation of our dataset.

4 Overview and highlights of the dataset

This section presents the main characteristics of the dataset, including the overall spatial distributions of N_{100} and N_{50}, seasonal patterns, and temporal variation in the concentrations. We also further explore a few selected regions, examining long-term trends in number concentrations and comparing these trends with results from earlier studies.

4.1 Global spatiotemporal patterns in the dataset

4.1.1 Global average spatial distributions and ratios of N_{50} and N_{100}

Figure 5 shows the global distribution of the dataset averaged over 2003-2024. The spatial distributions of N_{100} (Fig. 5a) and N_{50} (Fig. 5b) have similar large-scale patterns. Concentrations are generally higher over continents compared to over oceans, with the highest concentrations occurring in densely populated regions of South and East Asia. In contrast, remote continental

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

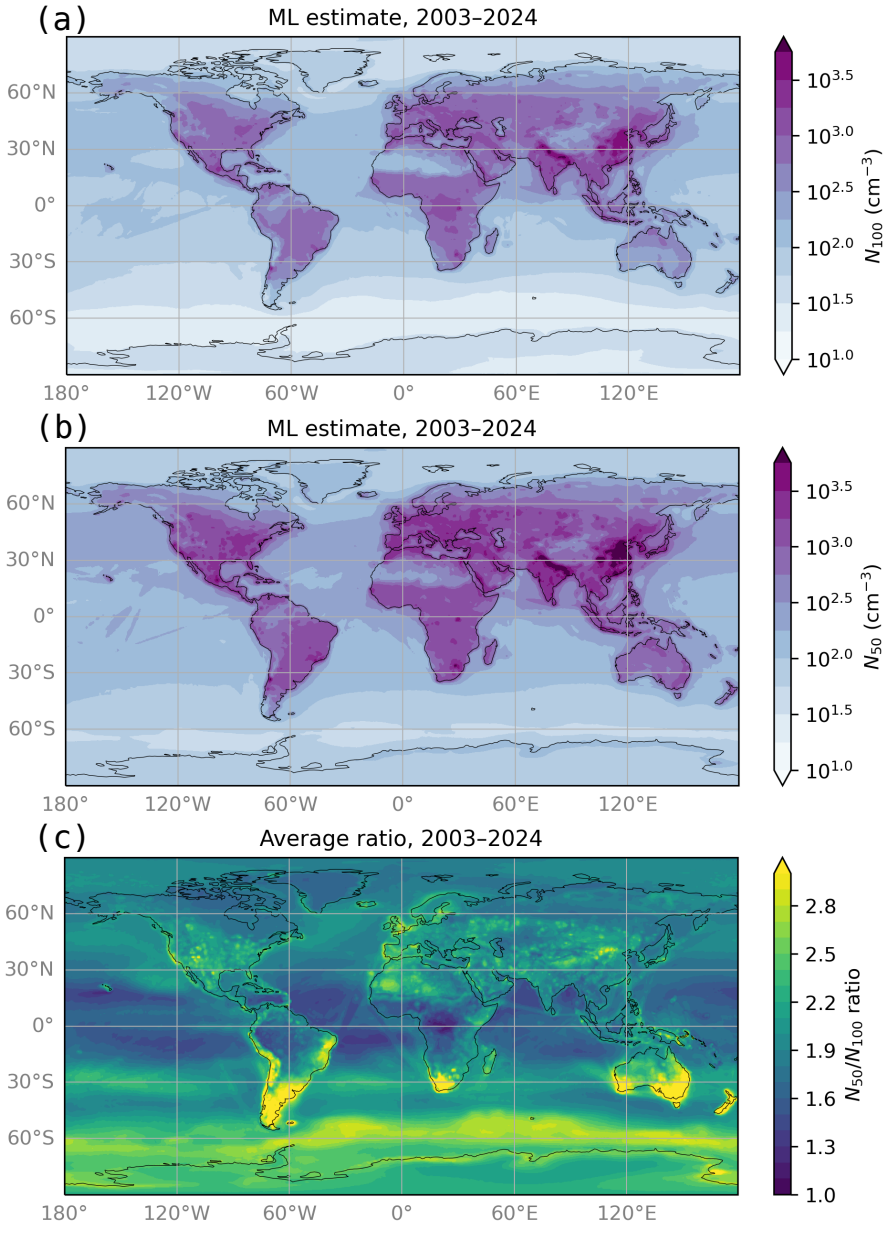

Figure 5. The global average spatial distributions of a) N_{100} and b) N_{50}, and c) the ratio of N_{50} to N_{100} for 2003-2024.

regions, including polar areas and deserts, exhibit lower concentrations, comparable to those over the oceans. Over the oceans themselves, concentrations tend to be elevated in continental outflow regions and along major shipping routes, while the lowest values appear in high-latitude marine environments.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Although the overall spatial distributions of N_{100} and N_{50} are similar, there are some differences, as indicated by the average ratio of N_{50} to N_{100} in Figure 5c. On average, the N_{50} concentrations are larger than N_{100}, consistent with their definitions, but their exact ratio varies throughout the world. Locations with high N_{50} to N_{100} ratio include Antarctica, several coastal regions of continents in the southern hemisphere and several population centres. In contrast, there are regions in the tropics where the ratio is close to one, suggesting that these locations contain proportionally fewer particles in 50-100 nm size range. In particular, tropical rain forests as well as tropical oceans, except the Indian Ocean and along major shipping routes, have a low ratio of N_{50} to N_{100}.

Furthermore, although the 21-year average N_{50} to N_{100} ratio in tropical regions is always greater than or equal to one, individual daily values can occasionally show N_{100} exceeding N_{50}. This is not physically correct, as N_{100} is by definition a subset of N_{50}. In most such cases, the two variables are nearly equal and can be considered effectively the same. However, in a few percent of grid cells on a given day, the N_{50} to N_{100} ratio falls below 0.9, which can be an issue. These inconsistencies arise because the ML models for N_{50} and N_{100} are trained independently. In such cases, we recommend using the N_{100} value for both variables.

4.1.2 Global concentrations summary and comparison to literature

To compare our results with those from other studies, we also calculated global summary metrics of the concentrations between latitudes 65°N and 65°S. For N_{100}, concentrations have a global mean of 273 cm^{-3} and a median of 118 cm^{-3}, with values ranging from 23 cm^{-3} (5th percentile) to 947 cm^{-3} (95th percentile). For N_{50}, the global mean is 506 cm^{-3} and the median 195 cm^{-3}, spanning from 59 cm^{-3} (5th percentile) to 1801 cm^{-3} (95th percentile).

For comparison, Choudhury and Tesche (2023) report the CCN concentrations between latitudes 65°N and 65°S and below 2km during June 2006 through December 2021. Their results show for the satellite-based estimate a global median of 274 cm^{-3} and a 5th–95th percentile range of 107cm^{-3}–1445cm^{-3}, and for the reanalysis-based estimate a median of 153cm^{-3} with a range of 28cm^{-3}–619cm^{-3}. These values place the N_{50} and N_{100} distributions from our dataset within a similar range of CCN-relevant particle concentrations.

4.1.3 Spatial patterns of temporal variability in N_{50} and N_{100}

The magnitude of temporal variability in N_{100} and N_{50} concentrations is shown in Figure 6. For each grid cell, we calculated the relative standard deviation over the period 2003–2024, defined as the standard deviation of daily values divided by the mean over the full time period. This metric characterises variability relative to the local concentration level. The global distributions of relative standard deviation are broadly similar for N_{100} and N_{50}, although N_{50} generally exhibits lower variability. For both variables, the highest relative standard deviations occur in northern North America and Eurasia, many coastal regions, South-East Asia, and certain tropical continental regions, including the Amazon rainforest. These regions are typically characterised either by relatively low background concentrations punctuated by episodic high-emission events (such as forest fires) or by strong seasonal variability in emissions and/or meteorological conditions. This seasonal behaviour is illustrated in Figure 7, which we discuss next.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

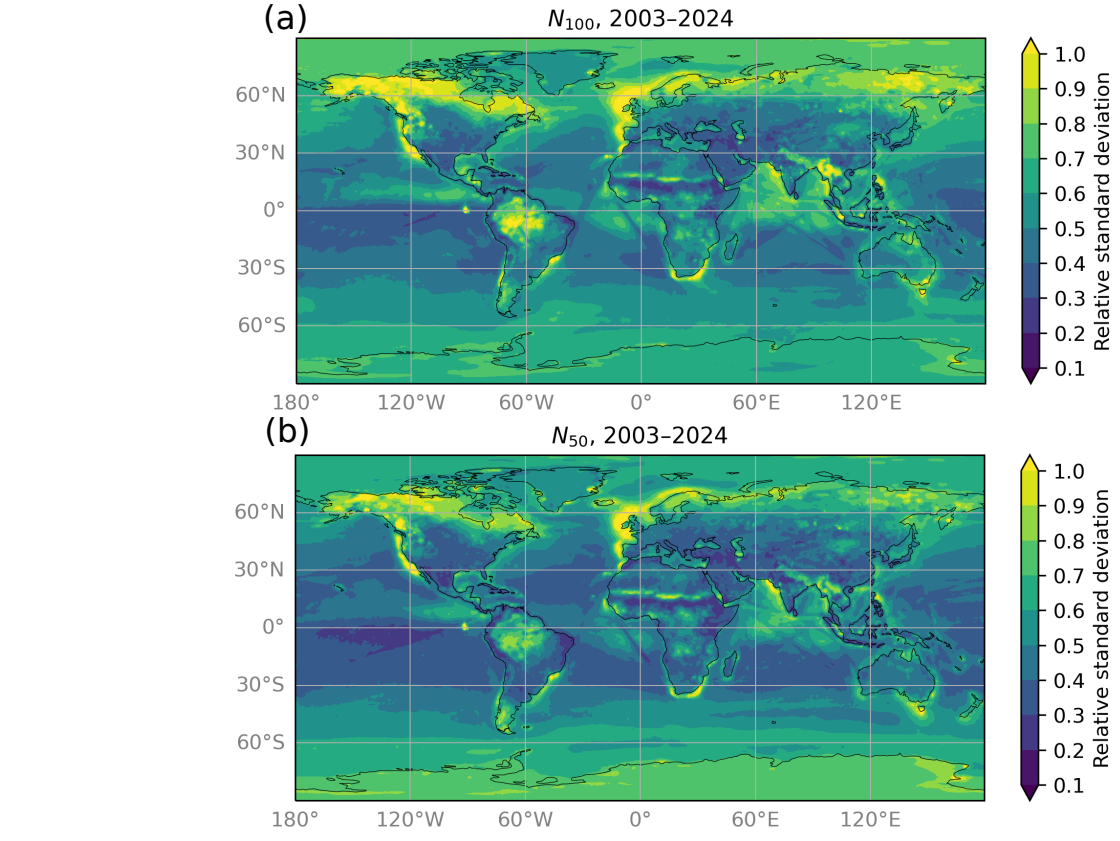

Figure 6. Relative standard deviation calculated for each grid cell for 2003-2024. Panel a) shows N_{100} result and panel b) N_{50}.

4.1.4 Global spatial distributions and ratios of N_{50} and N_{100} by season

Figure 7 shows the seasonal characteristics of the global distributions, presenting 21-year averages for December–February (Fig.7a, c) and June–August (Fig. 7b, d). The results for March–May and September–November are provided in Figure S5. There are clear differences between the seasons, reflecting the expected seasonal shifts in emissions and atmospheric processes. In most parts of the continental northern hemisphere, concentrations are higher in June-August (Fig. 7b, d) than in December-February (Fig. 7a, c), likely due to enhanced biogenic emissions during the boreal summer. The southern hemisphere shows a similar pattern, with higher concentrations in December-February during austral summer, although the seasonal contrast is weaker than in the north. In the polar regions, both the Arctic and Antarctic show peak concentrations during December–February. Although this pattern may initially appear counter-intuitive, it is likely driven by different processes in the two poles. In the Arctic, high winter concentrations tend to be associated with long-range transported anthropogenic pollution, usually called Arctic haze (e.g., Freud et al., 2017), whereas in Antarctica high concentrations are mainly due to secondary particle formation associated with strong biogenic emissions and photochemistry during local summer (e.g., Wang et al., 2023).

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

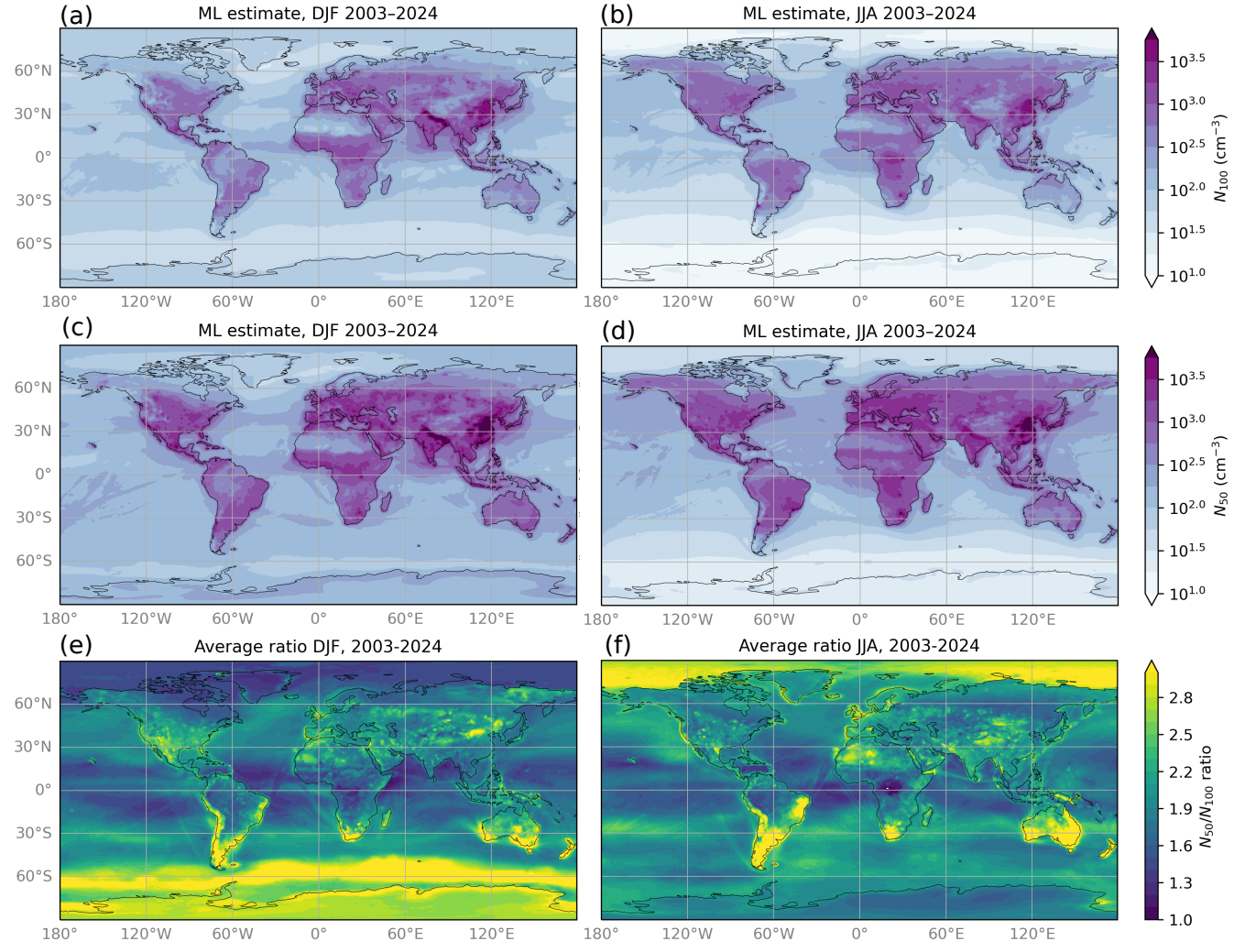

Figure 7. The global average spatial distributions of N_{100} and N_{50} and N_{50} to N_{100} ratio by season for 2003-2024. Panels a), c) and e) contain December-February result and panels b), d), f) June-August.

Overall, these patterns indicate that the dataset is able to physically meaningfully capture at least some of the key seasonal characteristics of N_{50} and N_{100}.

In addition to seasonal patterns in the average concentrations, the ratio of N_{50} to N_{100} also exhibits seasonal variation (Fig. 7e-f). Although this variation is more challenging to explain than the seasonality in the absolute concentrations, we provide a tentative interpretation here to illustrate the usability of the dataset. The ratio remains low in the tropics throughout the year. We can even see small areas where the 21-year average ratio is less than one, even though this is physically incorrect, as discussed related to Figure 5. The low ratio indicates that in the tropics N_{50} and N_{100} values are close to each other throughout the year. In high latitudes, the ratio varies and reaches maximum during the respective summer season. A more modest variation

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

with an opposite seasonal cycle is present in the mid-latitudes, where the ratio reaches a maximum during the winter months. This variation can be physically reasonable if it is related to a significant fraction of mid-latitude primary particle emissions occurring closer or below 50 nm (Paasonen et al., 2016; Rönkkö et al., 2017) and the lack of biogenic vapours growing the particles to over 100 nm during the cold season (Paasonen et al., 2013). Moving towards higher latitudes, especially in the Northern hemisphere, the increasing contributions of residential combustion and long range transport during the winter, both associated with larger particles (Paasonen et al., 2016; Pernov et al., 2022), can be speculated to become more dominant.

4.2 Regional trends

To highlight the capabilities of the dataset, we examined changes in particle number concentrations over the 21-year period in a few selected regions (Fig. 8). The corresponding average seasonal cycles for these regions are shown in the Supplement (Fig. S6).

We investigated two anthropogenically influenced regions that contain the highest particle number concentrations in our dataset. Region 1 is located in China (Fig. 8a). The selected region does not contain any of the training or holdout stations used to train and test the model, but it is located within the same broader North China Plain region. Figure 8a shows that yearly mean number concentrations have decreased in Region 1 from 2006 onwards, and a Mann-Kendall test reveals that the decreasing trend is statistically significant ($p_{N100} = 2.24 \cdot 10^{-7}$, $p_{N50} = 4.71 \cdot 10^{-7}$). Additionally, from 2003 to 2006 there is a visible increase in the number concentration. From the literature, we know that 2006 coincides with the implementation of air quality regulations leading to reduction of SO_2 (e.g., van der A et al., 2017), which is likely reflected in CAMS reanalysis data and therefore in our dataset. The decreasing trend in number concentration is also consistent with Shen et al. (2022), who reported declining accumulation-mode particle number concentrations in the Yangtze River Delta region between 2013 and 2020.

The second high-concentration region we considered in our analysis is located in the Indo-Gangetic plain, India (Fig. 8b). Region 2 was selected due to its well known air quality problem, and is positioned to avoid the measurement stations used in the training set. The yearly average number concentrations increased during 2003–2024 (Fig. 8b), and this trend is statistically significant ($p_{N100} = 1.23 \cdot 10^{-8}$, $p_{N50} = 1.70 \cdot 10^{-8}$). This increase is consistent with other aerosol measurements from the region, including aerosol optical depth and particle mass concentration observations (e.g., Namdeo et al., 2024), which, although not directly equivalent to particle number concentrations, still provide insight into overall aerosol loading. Region 2 also exhibits a very clear seasonal cycle, visible both in time series (Fig. 8b) and in Figure S6, which shows the range of monthly mean concentrations for 2003-2024. Different to North China Plain, the ratio of N_{50} to N_{100} remains quite low, which is consistent with the observed high contribution of primary aerosol emissions in this region (e.g., Bhandari et al., 2020).

Beyond these two high-concentration regions, we also included one region from Europe. Europe is likewise strongly influenced by anthropogenic emissions, but its concentration levels are notably lower and have been decreasing in recent decades (e.g., Leinonen et al., 2022). Region 3 is located in Poland to avoid training and holdout stations in Europe (Fig. 8c). Our dataset portrays a decreasing trend ($p_{N100} = 9.27 \cdot 10^{-7}$, $p_{N50} = 1.23 \cdot 10^{-6}$) for the yearly mean values (Fig. 8c), which is consistent with the observed decrease in $PM_{2.5}$ in Poland between 2005 and 2021 (Zgłobicki and Baran-Zgłobicka, 2024). Interestingly, there is also a change in the seasonal cycle: until 2016, the highest concentrations always occur during winter,

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

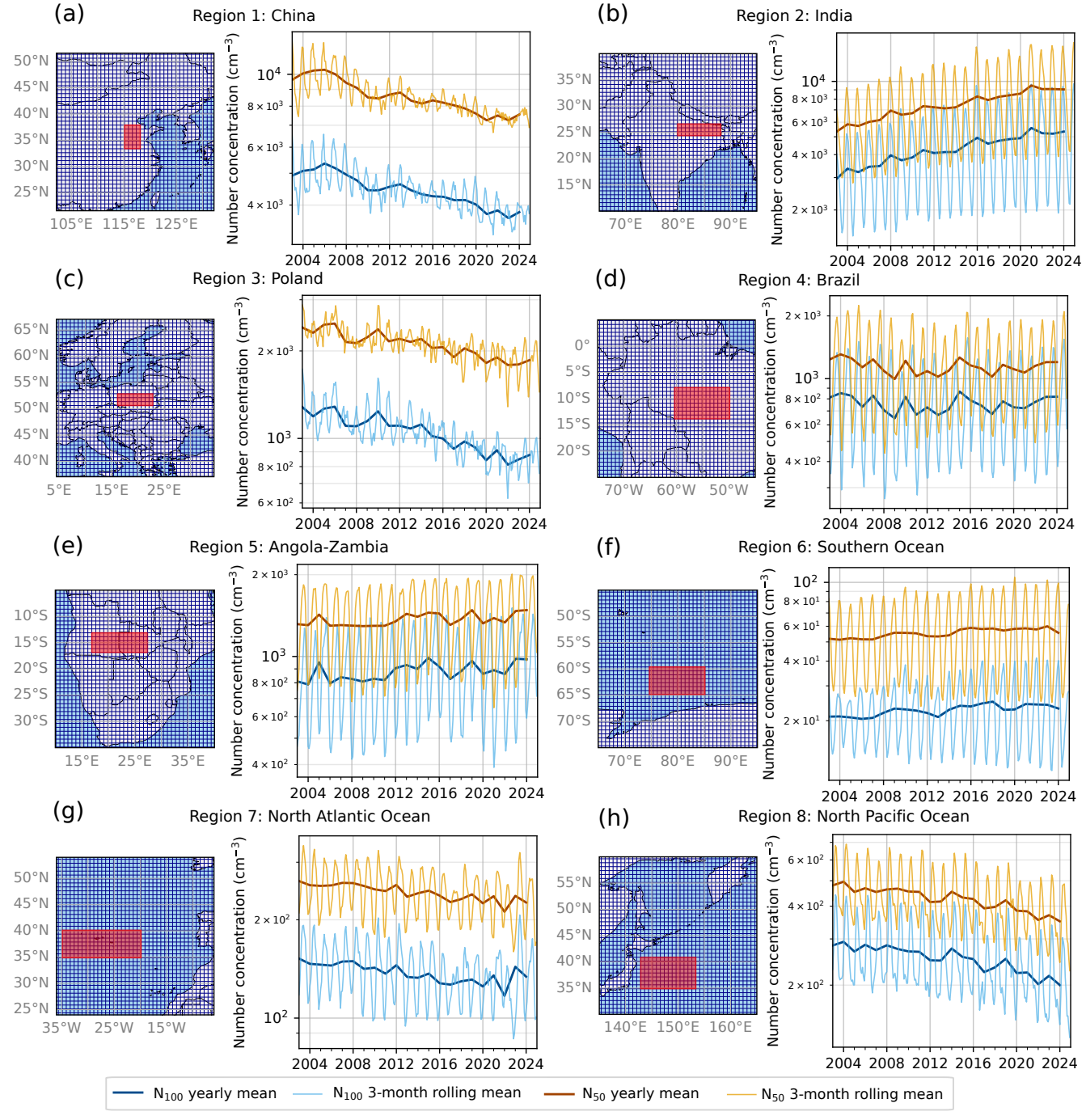

Figure 8. Number concentration time-series for selected regions.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

whereas toward the end of the time series, especially after 2021, winter becomes the season with the lowest concentrations. We hypothesize that this shift may be related to the gradual implementation of stricter air quality policies (e.g., WFOŚiGW, 2026; NFOŚiGW, 2021), targeting especially residential coal combustion for heating, which has traditionally been prevalent in Poland. Possibly driven by the decreasing emissions from this sector during winter, the seasonal cycle evolves towards a typical Western European cycle, showing maxima during summer due to enhanced biogenic growth of smaller particles to the studied size ranges (Paasonen et al., 2013; Yli-Juuti et al., 2021). This example illustrates that our dataset can be used to investigate not only long-term trends but possibly also evolving seasonal patterns in particle number concentrations.

In addition to the anthropogenically influenced regions, we investigated two regions affected by biomass burning: Region 4 in Brazil (Fig. 8d) and Region 5 in Angola and Zambia (Fig. 8e). Both of these areas are highlighted in the Global Fire Emissions Database, Version 4 (Randerson et al., 2017). In Region 4, the Mann-Kendall test found no statistically significant trends ($p_{N100} = 0.910$, $p_{N50} = 0.463$) in yearly means (Fig. 8d). In Region 5, there is a mild increasing trend ($p_{N100} = 8.03 \cdot 10^{-3}$, $p_{N50} = 9.48 \cdot 10^{-3}$), which is statistically significant (Fig. 8e). In both of these regions, there is a clear seasonal cycle and the N_{50} to N_{100} ratio is small (Figure S6), likely reflecting the dominance of biomass burning emissions, which occur mainly in sizes close to or larger than 100 nm.

Lastly, we analysed three marine regions. Region 6 is located in the Southern Ocean, south of the Indian Ocean (Fig. 8f). The concentrations exhibit a mild but statistically significant ($p_{N100} = 5.81 \cdot 10^{-4}$, $p_{N50} = 7.89 \cdot 10^{-5}$) increasing trend, driven primarily by a rise in the highest concentrations, while the lowest concentrations remain quite constant throughout the period.

Region 7 is located in the North Atlantic Ocean (Fig. 8g) and is centered around the Azores, where one of the holdout stations is located. Region 8 is in the North Pacific, near the coast of Japan (Fig. 8h). Both of these regions have a distinct seasonal cycle (Figure S6) as well as a statistically significant decreasing trend (Region 7: $p_{N100} = 1.58 \cdot 10^{-4}$, $p_{N50} = 6.43 \cdot 10^{-6}$; Region 8: $p_{N100} = 1.23 \cdot 10^{-6}$, $p_{N50} = 9.27 \cdot 10^{-7}$) in the yearly mean concentrations (Fig. 8g, h), which is also seen in the literature. For example, von Salzen et al. (2025) show a decrease in cloud droplet number concentration over the North Atlantic and North Pacific oceans during 2003-2022, caused by reduced emissions of sulphur dioxide and aerosol precursors from China and other countries in Northern Hemisphere.

5 Conclusions

Reliable information on the global distribution of particles larger than N_{100} and N_{50} is essential for understanding aerosol–cloud interactions, but observations of these particle number concentrations remain both spatially and temporally sparse. Recent years have seen substantial progress toward addressing this limitation, with several new methods and datasets emerging that estimate CCN or particle number concentrations beyond the limited in situ network. Despite these advances, there is still a need for a dataset that provides global observation-based estimates at high temporal resolution.

In this study, we produce a dataset that contains global daily-averaged ground-level N_{100} and N_{50} at 0.75 x 0.75-degree resolution from 2003 to 2024. To this end, we gathered long-term in situ measurements of particle number size distribution from extensive 74 stations and calculated N_{100} and N_{50} from them. Both datasets are published in this study.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

We used the data from 62 of the in situ stations to train an XGB machine learning model to estimate N_{100} and N_{50} based on the reanalysis data and validated the model using leave-one-station-out cross-validation. With the final global model and the global fields of 19 reanalysis variables, we generated the global model output for N_{100} and N_{50}. We evaluated these against the remaining 12 holdout stations. This analysis revealed that the final model output captures the median number concentrations within a factor of 1.5 from the observed median concentration for most holdout stations (8 out 12). However, on a daily scale, the model output tends to miss the highest and lowest daily average concentrations. The holdout stations with challenges in performance suffer from varying limitations of our dataset, which include lack of representation in the model training, or challenges in capturing the local conditions with the combination of reanalysis data and machine learning.

We describe the main characteristics of the 21-year dataset, including its global spatial distributions, seasonal patterns, and temporal variations in particle number concentrations. In our dataset, the highest concentrations occur in densely populated regions of South and East Asia, whereas the lowest values are found in remote continental areas and over the oceans. The dataset also exhibits a seasonal cycle, with higher concentrations during summer in most continental regions. The ratio of N_{50} to N_{100} varies substantially both spatially and temporally, indicating notable differences in the fraction of particles in the 50–100 nm size range. Finally, the analysis of long-term trends in several selected regions shows that, in many cases, the trends derived from the dataset qualitatively align with increases or decreases reported in the literature.

This study provides an observation-constrained framework for estimating global daily N_{50} and N_{100} concentrations over a 21-year period, effectively bridging the gap between sparse in situ measurements and computationally intensive process-based models. While model performance varies across regions, the dataset robustly captures central tendencies and large-scale spatial and temporal patterns of particle number concentrations, providing a consistent and physically plausible representation of global variability. A key strength of this approach is its ability to extend observational information into regions without direct measurements, thereby offering new insight into previously under-constrained environments.

The resulting dataset can be used for a wide range of applications, including as input for weather and climate models, as a benchmark for alternative estimation methods, and for the evaluation of purely model-driven simulations of N_{50} and N_{100}. Importantly, the methodology is inherently scalable and reproducible. As the reanalysis datasets used to train the global XGB model continue to be updated to include 2025 and later years, and may also undergo further improvements or incorporate additional variables, the workflow presented here enables the generation of regularly updated, observation-informed global datasets of particle number concentrations. By making both the dataset and the underlying code openly available, this study provides a flexible and extensible foundation for future developments in global microphysical aerosol characterization and for advancing the representation of CCN-relevant particles in atmospheric models.

6 Code and data availability

The dataset produced in this study is available at (https://doi.org/10.5281/zenodo.20202080, Ovaska et al. (2026b)). The dataset is organised into yearly NetCDF files, which each contain the global N_{100} and N_{50} in daily temporal resolution and 0.75° spatial resolution. The datasets are published under a Creative Commons Attribution 4.0 International licence.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

The in situ observations of N_{100} and N_{50} used to train the XGB model are available at (https://doi.org/10.5281/zenodo.20200333, Ovaska et al. (2026a)). These are organised into NetCDF files per measurement station. These data have daily temporal resolution. The datasets are published under a Creative Commons Attribution 4.0 International licence.

The code used to produce the results and the dataset is available at GitLab https://version.helsinki.fi/oaino/global_nconc_estimates/-/tree/e9f752af3ccd69cc2b2137f478b3386ce496fb37/ under MIT license. The code is written with Python.

The reanalysis data used for training the XGB model was retrieved from following sources: The CAMS global reanalysis (EAC4) dataset (Inness et al., 2019) was downloaded from the Copernicus Atmosphere Monitoring Service (CAMS) Atmosphere Data Store (ADS) (https://ads.atmosphere.copernicus.eu/datasets/cams-global-reanalysis-eac4?tab=overview, last access: 3 November 2025). This contains modified Copernicus Atmosphere Monitoring Service information from 2025. The "ERA5 hourly data on single levels from 1940 to present" dataset was downloaded from the Copernicus Climate Data Store (2025). This contains modified by Copernicus Climate Change Service information from 2025. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains.

Author contributions. Conceptualization: AO, PP, and TN.

Methodology: AO, PP, ER, DH, and KP. Software: AO, ER, and DH.

Formal analysis: AO, ER, and DH.

Funding acquisition: AO, PP, TP, MK, and TN.

Investigation (including experiments): all co-authors.

Data curation:AA, PA, EA, JB, BB, MBo, LB, MBu, DCDO, JACV, DCe, TCh, TCa, SCo, DCo, SCr, VD, KE, JE, EE, MF, SG, OG, FGM, RHa, AHo, RHo, AHy, TH, JJ, NK, HK, JK, AK, CK, MK, LL, AL, HL, AM, MM, SN, JO, NP, JEP, TP, MPi, CP, MPö, JPP, XQ, CR, SR, PR, JSc, KS, DS, AS, MSi, HS, MSo, TS, JSu, PT, VV, PVZ, AV, JV, KW, AW, HJY, PZ.

Project administration: PP.

Supervision: PP, VAS, and VMK.

Visualization: AO.

Writing (original draft): AO.

Writing (review and editing): AO, PP, VAS, and VMK, with all other co-authors.

Competing interests. Authors declare no competing interests.

Acknowledgements. We warmly acknowledge following people: Pasi Aalto for pioneering the PNSD measurements at Hyytiälä, Värriö and DOME-C stations; Maria Cristina Facchini, Leone Tarozzi and Stefano Decesari for measurements in San Pietro Capofiume; Steffen Freitag and Lena Große Schute for their support and collaboration related to measurements in Mülheim-Styrum; Paolo Villani for measurements in Observatoire Perenne de l'Environnement; the Environment and Climate Change stations operators (Andrew Platt, Cassie Worthy, Summer

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Hunter, Molly Farquhar, Kris Kunz, Adrienne Richards) who helped to maintain all instruments to their best working conditions over the years for generating the measurements.

Some of the data included in the analysis were collected at the Southern Great Plains (SGP), Ascension Island (ASI), Eastern North Atlantic (ENA), La Jolla (EPC), Gunnison (GUC) and Manacapuru (MAO) sites of the Atmospheric Radiation Measurement (ARM) user facility, a US Department of Energy (DOE) Office of Science user facility managed by the Biological and Environmental Research program.

Funding for the ultrafine PNSD measurements from Mülheim-Styrum (2016-2020) station was provided by the North-Rhine Westphalian State Agency for Nature, Environment and Climate (LANUK).

This publication forms part of the output of the Biogeochemistry Research Infrastructure Platform (BIOGRIP) of the Department of Science and Innovation of South Africa, which supports the Welgegund station.

The Harwell measurement station was supported by the UK Department for Environment, Food and Rural Affairs.

Measurements at Villum Research Station in Greenland were supported by the Danish Ministry of Climate, Energy, and Utilities.

For the operation of the ATTO site, we acknowledge the support by the Instituto Nacional de Pesquisas da Amazônia (INPA), the Amazon State University (UEA), the Large-Scale Biosphere-Atmosphere Experiment (LBA), FAPEAM, the Reserva de Desenvolvimento Sustentável do Uatumã (SDS/CEUC/RDS-Uatumã), and the Max Planck Society. Particularly, we would like to thank all ATTO team members involved in the technical, logistical, and scientific support. The ATTO project has been funded, among other sources, by the Max Planck Society, the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil), and the FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo).

The authors gratefully acknowledge CNRS-INSU for supporting measurements performed at the SI-SIRTA, and those within the long-term monitoring aerosol program SNO-CLAP, both of which are components of the ACTRIS French Research Instructure, and whose data is hosted at the AERIS data center (https://www.aeris-data.fr/).

The authors acknowledge the Knut and Alice Wallenberg foundation (KWA) for funding ACTRIS-Sweden.

We acknowledge the CSC – IT Centre for Science, Finland, for providing the computational resources.

Aino Ovaska acknowledges funding from the Doctoral Programme in Atmospheric Sciences at the University of Helsinki (ATM-DP).

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

References

ARM | Eastern Pacific Cloud Aerosol Precipitation Experiment, https://armgov.svcs.arm.gov/research/campaigns/amf2023epcape, last access: 12 March 2026, 2023.

Aalto, P., Hämeri, K., Becker, E., Weber, R., Salm, J., Mäkelä, J. M., Hoell, C., O'dowd, C. D., Hansson, H.-C., Väkevä, M., Koponen, I. K., Buzorius, G., and Kulmala, M.: Physical characterization of aerosol particles during nucleation events, Tellus, 53B, 344–358, https://doi.org/10.3402/tellusb.v53i4.17127, 2001.

ACTRIS: Vielsalm - ACTRIS NF labelling, https://actris-nf-labelling.out.ocp.fmi.fi/facility/6, 2024.

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

Antonovich, V. V., Antokhin, P. N., Arshinov, M. Y., Belan, B. D., Balin, Y. S., Davydov, D. K., Ivlev, G. A., Kozlov, A. V., Kozlov, V. S., Kokhanenko, G. P., Novoselov, M. M., Panchenko, M. V., Penner, I. E., Pestunov, D. A., Savkin, D. E., Simonenkov, D. V., Tolmachev, G. N., Fofonov, A. V., Chernov, D. G., Smargunov, V. P., Yausheva, E. P., Paris, J.-D., Ancellet, G., Law, K. S., Pelon, J., Machida, T., and Sasakawa, M.: Station for the comprehensive monitoring of the atmosphere at Fonovaya Observatory, West Siberia: current status and future needs, in: Proc.SPIE, vol. 10833, p. 108337Z, https://doi.org/10.1117/12.2504388, 2018.

Asbach, C., Todea, A. M., and Kaminski, H.: Evaluation of a Partector Pro for atmospheric particle number size distribution and number concentration measurements at an urban background site, Aerosol Research, 2, 1–12, https://doi.org/10.5194/ar-2-1-2024, 2024.

Asmi, A., Wiedensohler, A., Laj, P., Fjaeraa, A. M., Sellegri, K., Birmili, W., Weingartner, E., Baltensperger, U., Zdimal, V., Zikova, N., Putaud, J. P., Marinoni, A., Tunved, P., Hansson, H. C., Fiebig, M., Kivekäs, N., Lihavainen, H., Asmi, E., Ulevicius, V., Aalto, P. P., Swietlicki, E., Kristensson, A., Mihalopoulos, N., Kalivitis, N., Kalapov, I., Kiss, G., Leeuw, G. D., Henzing, B., Harrison, R. M., Beddows, D., O'Dowd, C., Jennings, S. G., Flentje, H., Weinhold, K., Meinhardt, F., Ries, L., and Kulmala, M.: Number size distributions and seasonality of submicron particles in Europe 2008-2009, Atmospheric Chemistry and Physics, 11, 5505–5538, https://doi.org/10.5194/acp-11-5505-2011, 2011.

Asmi, E., Kondratyev, V., Brus, D., Laurila, T., Lihavainen, H., Backman, J., Vakkari, V., Aurela, M., Hatakka, J., Viisanen, Y., Uttal, T., Ivakhov, V., and Makshtas, A.: Aerosol size distribution seasonal characteristics measured in Tiksi, Russian Arctic, Atmospheric Chemistry and Physics, 16, 1271–1287, https://doi.org/10.5194/acp-16-1271-2016, 2016.

Baalbaki, R., Pikridas, M., Jokinen, T., Laurila, T., Dada, L., Bezantakos, S., Ahonen, L., Neitola, K., Maisser, A., Bimenyimana, E., Christodoulou, A., Unga, F., Savvides, C., Lehtipalo, K., Kangasluoma, J., Biskos, G., Petäjä, T., Kerminen, V.-M., Sciare, J., and Kulmala, M.: Towards understanding the characteristics of new particle formation in the Eastern Mediterranean, Atmospheric Chemistry and Physics, 21, 9223–9251, https://doi.org/10.5194/acp-21-9223-2021, 2021.

Backman, J., Rizzo, L. V., Hakala, J., Nieminen, T., Manninen, H. E., Morais, F., Aalto, P. P., Siivola, E., Carbone, S., Hillamo, R., Artaxo, P., Virkkula, A., Petäjä, T., and Kulmala, M.: On the diurnal cycle of urban aerosols, black carbon and the occurrence of new particle formation events in springtime São Paulo, Brazil, Atmospheric Chemistry and Physics, 12, 11 733–11 751, https://doi.org/10.5194/acp-12-11733-2012, 2012.

Baray, J.-L., Courcoux, Y., Keckhut, P., Portafaix, T., Tulet, P., Cammas, J.-P., Hauchecorne, A., Beekmann, S. G., Mazière, M. D., Hermans, C., Desmet, F., Sellegri, K., Colomb, A., Ramonet, M., Sciare, J., Vuillemin, C., Hoareau, C., Dionisi, D., Duflot, V., Vérèmes, H., Porteneuve, J., Gabarrot, F., Gaudo, T., Metzger, J.-M., Payen, G., de Bellevue, J. L., Barthe, C., Posny, F., Ricaud, P., Abchiche, A., and

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Delmas, R.: Maïdo observatory: a new high-altitude station facility at Reunion Island (21° S, 55° E) for long-term atmospheric remote sensing and in situ measurements, Atmospheric Measurement Techniques, 6, 2865–2877, https://doi.org/10.5194/amt-6-2865-2013, 2013.

Barreiro, M., Alonso-Blanco, E., Becerril, M., Coz, E., Díaz, E., Fernández, A. J., Gómez-Moreno, F. J., Molero, F., Núñez, L., Palacios, M., Salvador, P., Pujadas, M., and Artíñano, B.: Continuous observations of chemical and physical properties of atmospheric aerosols in southern Europe, in: Proceedings of the International Conference on Aerosol Cycle Sources-Aging-Sinks-Impact (ICAC 2017), p. 77, 2017.

Beigaitė, R., Mechenich, M., and Žliobaitė, I.: Spatial Cross-Validation for Globally Distributed Data, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 13601 LNAI, pp. 127–140, Springer Science and Business Media Deutschland GmbH, ISBN 9783031188398, ISSN 16113349, https://doi.org/10.1007/978-3-031-18840-4_10, 2022.

Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., Stier, P., Watson-Parris, D., Boucher, O., Carslaw, K. S., Christensen, M., Daniau, A. L., Dufresne, J. L., Feingold, G., Fiedler, S., Forster, P., Gettelman, A., Haywood, J. M., Lohmann, U., Malavelle, F., Mauritsen, T., McCoy, D. T., Myhre, G., Mülmenstädt, J., Neubauer, D., Possner, A., Rugenstein, M., Sato, Y., Schulz, M., Schwartz, S. E., Sourdeval, O., Storelvmo, T., Toll, V., Winker, D., and Stevens, B.: Bounding Global Aerosol Radiative Forcing of Climate Change, Reviews of Geophysics, 58, 1–45, https://doi.org/10.1029/2019RG000660, 2020.

Beukes, J. P., Sehloho, T., van Zyl, P. G., Maritz, P., Josipovic, M., Jaars, K., Vakkari, V., Kulmala, M., and Laakso, L.: Organic and elemental carbon at a regional background site in South Africa, Atmospheric Environment, 362, 121 562, https://doi.org/10.1016/J.ATMOSENV.2025.121562, 2025.

Bhandari, S., Gani, S., Patel, K., Wang, D. S., Soni, P., Arub, Z., Habib, G., Apte, J. S., and Ruiz, L. H.: Sources and atmospheric dynamics of organic aerosol in New Delhi, India: insights from receptor modeling, Atmospheric Chemistry and Physics, 20, 735–752, https://doi.org/10.5194/acp-20-735-2020, 2020.

Birmili, W., Berresheim, H., Plass-Dülmer, C., Elste, T., Gilge, S., Wiedensohler, A., and Uhrner, U.: Atmospheric Chemistry and Physics The Hohenpeissenberg aerosol formation experiment (HAFEX): a long-term study including size-resolved aerosol, H2SO4 , OH, and monoterpenes measurements, Tech. rep., www.atmos-chem-phys.org/acp/3/361/, 2003.

Birmili, W., Weinhold, K., Rasch, F., Sonntag, A., Sun, J., Merkel, M., Wiedensohler, A., Bastian, S., Schladitz, A., Löschau, G., Cyrys, J., Pitz, M., Gu, J., Kusch, T., Flentje, H., Quass, U., Kaminski, H., Kuhlbusch, T. A. J., Meinhardt, F., Schwerin, A., Bath, O., Ries, L., Wirtz, K., and Fiebig, M.: Long-term observations of tropospheric particle number size distributions and equivalent black carbon mass concentrations in the German Ultrafine Aerosol Network (GUAN), Earth system science data, 8, 355–382, https://doi.org/10.5194/essd-8-355-2016, 2016.

Blichner, S. M., Sporre, M. K., Makkonen, R., and Berntsen, T. K.: Implementing a sectional scheme for early aerosol growth from new particle formation in the Norwegian Earth System Model v2: Comparison to observations and climate impacts, Geoscientific Model Development, 14, 3335–3359, https://doi.org/10.5194/gmd-14-3335-2021, 2021.

Block, K., Haghighatnasab, M., Partridge, D. G., Stier, P., and Quaas, J.: Cloud condensation nuclei concentrations derived from the CAMS reanalysis, Earth System Science Data, 16, 443–470, https://doi.org/10.5194/essd-16-443-2024, 2024.

Boyer, M., Aliaga, D., Pernov, J. B., Angot, H., Quéléver, L. L., Dada, L., Heutte, B., Dall'osto, M., Beddows, D. C., Brasseur, Z., Beck, I., Bucci, S., Duetsch, M., Stohl, A., Laurila, T., Asmi, E., Massling, A., Thomas, D. C., Nøjgaard, J. K., Chan, T., Sharma, S., Tunved, P., Krejci, R., Hansson, H. C., Bianchi, F., Lehtipalo, K., Wiedensohler, A., Weinhold, K., Kulmala, M., Petäjä, T., Sipilä, M., Schmale, J., and Jokinen, T.: A full year of aerosol size distribution data from the central Arctic under an extreme positive Arctic Oscillation: insights from

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

the Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC) expedition, Atmospheric Chemistry and Physics, 23, 389–415, https://doi.org/10.5194/ACP-23-389-2023, 2023.

Bu, J. O., Ko, H. J., Yoo, H. J., and Oh, S. M.: Multi-Aspect Analysis of Wildfire Aerosols from the 2023 Hongseong Case: Physical, Optical, Chemical, and Source Characteristics, Atmosphere, 16, https://doi.org/10.3390/atmos16091074, 2025.

Carnerero, C., Rivas, I., Reche, C., Pérez, N., Alastuey, A., and Querol, X.: Trends in primary and secondary particle number concentrations in urban and regional environments in NE Spain, Atmospheric Environment, 244, 117 982, https://doi.org/10.1016/J.ATMOSENV.2020.117982, 2021.

Casquero-Vera, J. A., Lyamani, H., Titos, G., Minguillón, M. C., Dada, L., Alastuey, A., Querol, X., Petäjä, T., Olmo, F. J., and Alados-Arboledas, L.: Quantifying traffic, biomass burning and secondary source contributions to atmospheric particle number concentrations at urban and suburban sites, Science of The Total Environment, 768, 145 282, https://doi.org/10.1016/J.SCITOTENV.2021.145282, 2021.

Charron, A., Birmili, W., and Harrison, R. M.: Factors influencing new particle formation at the rural site, Harwell, United Kingdom, Journal of Geophysical Research Atmospheres, 112, https://doi.org/10.1029/2007JD008425, 2007.

Chen, L., Qi, X., Nie, W., Wang, J., Xu, Z., Wang, T., Liu, Y., Shen, Y., Xu, Z., Kokkonen, T. V., Chi, X., Aalto, P. P., Paasonen, P., Kerminen, V. M., Petäjä, T., Kulmala, M., and Ding, A.: Cluster Analysis of Submicron Particle Number Size Distributions at the SORPES Station in the Yangtze River Delta of East China, Journal of Geophysical Research: Atmospheres, 126, https://doi.org/10.1029/2020JD034004, 2021.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, https://doi.org/10.1145/2939672.2939785, 2016.

Cho, D., Yoo, C., Im, J., and Cha, D. H.: Comparative Assessment of Various Machine Learning-Based Bias Correction Methods for Numerical Weather Prediction Model Forecasts of Extreme Air Temperatures in Urban Areas, Earth and Space Science, 7, https://doi.org/10.1029/2019EA000740, 2020.

Choudhury, G. and Tesche, M.: Estimating cloud condensation nuclei concentrations from CALIPSO lidar measurements, Atmospheric Measurement Techniques, 15, 639–654, https://doi.org/10.5194/amt-15-639-2022, 2022.

Choudhury, G. and Tesche, M.: A first global height-resolved cloud condensation nuclei data set derived from spaceborne lidar measurements, Earth System Science Data, 15, 3747–3760, https://doi.org/10.5194/essd-15-3747-2023, 2023.

Choudhury, G., Block, K., Haghighatnasab, M., Quaas, J., Goren, T., and Tesche, M.: Pristine oceans are a significant source of uncertainty in quantifying global cloud condensation nuclei, Atmospheric Chemistry and Physics, 25, 3841–3856, https://doi.org/10.5194/acp-25-3841-2025, 2025.

Conil, S., Helle, J., Langrene, L., Laurent, O., Delmotte, M., and Ramonet, M.: Continuous atmospheric CO2 , CH4 and CO measurements at the Observatoire Pérenne de l'Environnement (OPE) station in France from 2011 to 2018, Atmospheric Measurement Techniques, 12, 6361–6383, https://doi.org/10.5194/amt-12-6361-2019, 2019.

Copernicus Climate Change Service: ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), DOI:10.24381/cds.adbb2d47, last access: 3 November 2025, 2025.

Crumeyrolle, S., Kontkanen, J. S. S., Rose, C., Garcia, A. V., Bourrianne, E., Catalfamo, M., Riffault, V., Tison, E., de Brito, J. F., Visez, N., Ferlay, N., Auriol, F., and Chiapello, I.: Measurement report: Atmospheric new particle formation at a peri-urban site in Lille, northern France, Atmospheric Chemistry and Physics, 23, 183–201, https://doi.org/10.5194/acp-23-183-2023, 2023.

Dinoi, A., Gulli, D., Weinhold, K., Ammoscato, I., Calidonna, C. R., Wiedensohler, A., and Contini, D.: Characterization of ultrafine particles and the occurrence of new particle formation events in an urban and coastal site of the Mediterranean area, Atmospheric Chemistry and Physics, 23, 2167–2181, https://doi.org/10.5194/acp-23-2167-2023, 2023.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Dusek, U., Frank, G. P., Hildebrandt, L., Curtius, J., Schneider, J., Walter, S., Chand, D., Drewnick, F., Hings, S., Jung, D., Borrmann, S., and Andreae, M. O.: Size Matters More Than Chemistry for Cloud-Nucleating Ability of Aerosol Particles, Science, 312, 1375–1378, https://doi.org/10.1126/science.1125261, 2006.

Eckhardt, S., Breivik, K., Li, Y. F., Manø, S., and Stohl, A.: Source regions of some persistent organic pollutants measured in the atmosphere at Birkenes, Norway, Atmospheric Chemistry and Physics, 9, 6597–6610, https://doi.org/10.5194/acp-9-6597-2009, 2009.

Eleftheriadis, K., Gini, M. I., Diapouli, E., Vratolis, S., Vasilatou, V., Fetfatzis, P., and Manousakas, M. I.: Aerosol microphysics and chemistry reveal the COVID19 lockdown impact on urban air quality, Scientific Reports, 11, 14 477, https://doi.org/10.1038/s41598-021-93650-6, 2021.

Engler, C., Rose, D., Wehner, B., Wiedensohler, A., Brüggemann, E., Brüggemann, B., Gnauk, T., Spindler, G., Tuch, T., and Birmili, W.: Size distributions of non-volatile particle residuals (D p <800 nm) at a rural site in Germany and relation to air mass origin, Tech. rep., www.atmos-chem-phys.net/7/5785/2007/, 2007.

Fanourgakis, G. S., Kanakidou, M., Nenes, A., Bauer, S. E., Bergman, T., Carslaw, K. S., Grini, A., Hamilton, D. S., Johnson, J. S., Karydis, V. A., Kirkevåg, A., Kodros, J. K., Lohmann, U., Luo, G., Makkonen, R., Matsui, H., Neubauer, D., Pierce, J. R., Schmale, J., Stier, P., Tsigaridis, K., van Noije, T., Wang, H., Watson-Parris, D., delivery date: 11.05.2026 Daniel M. Westervelt, Yang, Y., Yoshioka, M., Daskalakis, N., Decesari, S., Gysel-Beer, M., Kalivitis, N., Liu, X., Mahowald, N. M., Myriokefalitakis, S., Schrödner, R., Sfakianaki, M., Tsimpidi, A. P., Wu, M., and Yu, F.: Evaluation of global simulations of aerosol particle and cloud condensation nuclei number, with implications for cloud droplet formation, Atmospheric Chemistry and Physics, 19, 8591–8617, https://doi.org/10.5194/acp-19-8591-2019, 2019.

Farah, A., Villani, P., Rose, C., Conil, S., Langrene, L., Laj, P., and Sellegri, K.: Characterization of Aerosol Physical and Optical Properties at the Observatoire Pérenne de l'Environnement (OPE) Site, Atmosphere 2020, Vol. 11, Page 172, 11, 172, https://doi.org/10.3390/ATMOS11020172, 2020.

Feldman, D. R., Aiken, A. C., Boos, W. R., Carroll, R. W., Chandrasekar, V., Collis, S., Creamean, J. M., de Boer, G., Deems, J., DeMott, P. J., Fan, J., Flores, A. N., Gochis, D., Grover, M., Hill, T. C., Hodshire, A., Hulm, E., Hume, C. C., Jackson, R., Junyent, F., Kennedy, A., Kumjian, M., Levin, E. J., Lundquist, J. D., O'Brien, J., Raleigh, M. S., Reithel, J., Rhoades, A., Rittger, K., Rudisill, W., Sherman, Z., Siirila-Woodburn, E., Skiles, S. M., Smith, J. N., Sullivan, R. C., Theisen, A., Tuftedal, M., Varble, A. C., Wiedlea, A., Wielandt, S., Williams, K., and Xu, Z.: The Surface Atmosphere Integrated Field Laboratory (SAIL) Campaign, Bulletin of the American Meteorological Society, 104, E2192–E2222, https://doi.org/10.1175/BAMS-D-22-0049.1, 2023.

Fiebig, M., Hirdman, D., Lunder, C. R., Ogren, J. A., Solberg, S., Stohl, A., and Thompson, R. L.: Annual cycle of Antarctic baseline aerosol: controlled by photooxidation-limited aerosol formation, Atmospheric Chemistry and Physics, 14, 3083–3093, https://doi.org/10.5194/acp-14-3083-2014, 2014.

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., Lunt, D., Mauritsen, T., Palmer, M., Watanabe, M., Wild, M., and Zhang, H.: Earth's energy budget, climate feedbacks, and climate sensitivity, p. 923–1054, Cambridge University Press, ISBN 9781009157896, https://doi.org/doi:10.1017/9781009157896.009, 2021.

Foucart, B., Sellegri, K., Tulet, P., Rose, C., Metzger, J.-M., and Picard, D.: High occurrence of new particle formation events at the Maïdo high-altitude observatory (2150 m), Réunion (Indian Ocean), Atmospheric Chemistry and Physics, 18, 9243–9261, https://doi.org/10.5194/acp-18-9243-2018, 2018.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Freud, E., Krejci, R., Tunved, P., Leaitch, R., Nguyen, Q. T., Massling, A., Skov, H., and Barrie, L.: Pan-Arctic aerosol number size distributions: Seasonality and transport patterns, Atmospheric Chemistry and Physics, 17, 8101–8128, https://doi.org/10.5194/acp-17-8101-2017, 2017.

Gani, S., Bhandari, S., Seraj, S., Wang, D. S., Patel, K., Soni, P., Arub, Z., Habib, G., Ruiz, L. H., and Apte, J. S.: Submicron aerosol composition in the world's most polluted megacity: the Delhi Aerosol Supersite study, Atmospheric Chemistry and Physics, 19, 6843–6859, https://doi.org/10.5194/acp-19-6843-2019, 2019.

Gani, S., Bhandari, S., Patel, K., Seraj, S., Soni, P., Arub, Z., Habib, G., Ruiz, L. H., and Apte, J. S.: Particle number concentrations and size distribution in a polluted megacity: The Delhi Aerosol Supersite study, Atmospheric Chemistry and Physics, 20, 8533–8549, https://doi.org/10.5194/acp-20-8533-2020, 2020.

García, M. I., Rodríguez, S., González, Y., and García, R. D.: Climatology of new particle formation at Izaña mountain GAW observatory in the subtropical North Atlantic, Atmospheric Chemistry and Physics, 14, 3865–3881, https://doi.org/10.5194/acp-14-3865-2014, 2014.

Georgiades, P., Kohl, M., Nicolaou, M. A., Christoudias, T., Pozzer, A., Dovrolis, C., and Lelieveld, J.: Global high-resolution ultrafine particle number concentrations through data fusion with machine learning, Scientific Data, 12, https://doi.org/10.1038/s41597-025-06055-9, 2025.

Gómez-Moreno, F. J., Alonso-Blanco, E., Mediavilla, P., and Pujadas, M.: Evolution of the number of nucleation events in an urban background site, in: Proceedings of the 7th Iberian Meeting on Aerosol Science and Technology (RICTA 2024), p. 105, http://hdl.handle.net/2183/36875, 2024.

Hamed, A., Joutsensaari, J., Mikkonen, S., Sogacheva, L., Maso, M. D., Kulmala, M., Cavalli, F., Fuzzi, S., Facchini, M. C., Decesari, S., Mircea, M., Lehtinen, K. E. J., and Laaksonen, A.: Atmospheric Chemistry and Physics Nucleation and growth of new particles in Po Valley, Italy, Tech. rep., www.atmos-chem-phys.net/7/355/2007/, 2007.

Hansen, G., Aspmo, K., Berg, T., Edvardsen, K., Fiebig, M., Kallenborn, R., Krognes, T., Lunder, C., Stebel, K., Schmidbauer, N., Solberg, S., and Yttri, K. E.: Atmospheric monitoring at the Norwegian Antarctic station Troll: measurement programme and first results, Polar Research, 28, 353–363, https://doi.org/10.1111/j.1751-8369.2009.00134.x, 2009.

Hari, P. and Kulmala, M.: Station for Measuring Ecosystem-Atmosphere Relations (SMEAR II), Boreal Env. Res, pp. 315–322, 2005.

Hari, P., Kulmala, M., Pohja, T., Lahti, T., Siivola, E., Palva, E., Aalto, P., Hameri, K., Vesala, T., Luoma, S., and Pulliainen, E.: Air pollution in eastern Lapland: challenge for an environmental measurement station, Silva Fennica, 28, 29–39, https://doi.org/https://doi.org/10.14214/sf.a9160, 1994.

Heintzenberg, J., Birmili, W., Otto, R., Andreae, M. O., Mayer, J.-C., Chi, X., and Panov, A.: Aerosol particle number size distributions and particulate light absorption at the ZOTTO tall tower (Siberia), 2006–2009, Atmospheric Chemistry and Physics, 11, 8703–8719, https://doi.org/10.5194/acp-11-8703-2011, 2011.

Hooda, R. K., Kivekäs, N., O'Connor, E. J., Coen, M. C., Pietikäinen, J. P., Vakkari, V., Backman, J., Henriksson, S. V., Asmi, E., Komppula, M., Korhonen, H., Hyvärinen, A. P., and Lihavainen, H.: Driving Factors of Aerosol Properties Over the Foothills of Central Himalayas Based on 8.5 Years Continuous Measurements, Journal of Geophysical Research: Atmospheres, 123, 13,421–13,442, https://doi.org/10.1029/2018JD029744, 2018.

Hussein, T., Dada, L., Hakala, S., Petäjä, T., and Kulmala, M.: Urban aerosol particle size characterization in Eastern Mediterranean Conditions, Atmosphere, 10, https://doi.org/10.3390/atmos10110710, 2019.

Hussein, T., Atashi, N., Sogacheva, L., Hakala, S., Dada, L., Petäjä, T., and Kulmala, M.: Characterization of urban new particle formation in Amman-Jordan, Atmosphere, 11, https://doi.org/10.3390/atmos11010079, 2020.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Hyvärinen, A.-P., Lihavainen, H., Komppula, M., Panwar, T. S., Sharma, V. P., Hooda, R. K., and Viisanen, Y.: Aerosol measurements at the Gual Pahari EUCAARI station: preliminary results from in-situ measurements, Atmospheric Chemistry and Physics, 10, 7241–7252, https://doi.org/10.5194/acp-10-7241-2010, 2010.

in 't Veld, M., Alastuey, A., Pandolfi, M., Amato, F., Pérez, N., Reche, C., Via, M., Minguillón, M. C., Escudero, M., and Querol, X.: Compositional changes of PM2.5 in NE Spain during 2009–2018: A trend analysis of the chemical composition and source apportionment, Science of The Total Environment, 795, 148 728, https://doi.org/10.1016/J.SCITOTENV.2021.148728, 2021.

Inness, A., Ades, M., Agusti-Panareda, A., Barre, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., and Suttie, M.: The CAMS reanalysis of atmospheric composition, Atmospheric chemistry and physics, 19, 3515–3556, https://doi.org/10.5194/acp-19-3515-2019, 2019a.

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A., Dominguez, J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., M, R., Remy, S., Schulz, M., and Suttie, M.: CAMS global reanalysis (EAC4). Copernicus atmosphere monitoring service (CAMS) atmosphere data store (ADS), https://ads.atmosphere.copernicus.eu/datasets/cams-global-reanalysis-eac4?tab=overview, last access: 3 November 2025, 2019b.

Isenor, B. H., Varnaite, G., Jeong, C.-H., Subramanian, P. S. G., Duruisseau-Kuntz, R., Zaherddine, V., Chan, T. W., Al-Jabiri, M. H., Askari, A., Rivellini, L.-H., Olfert, J. S., Sharma, S., Evans, G. J., Verma, V., Abbatt, J. P. D., and Chan, A. W. H.: Impacts of the 2023 Canadian Wildfires on the Oxidative Potential of Particulate Matter, ACS ES&T Air, https://doi.org/10.1021/ACSESTAIR.5C00182/SUPPL_FILE/EA5C00182_SI_001.PDF, 2025.

Järvi, L., Hannuniemi, H., Hussein, T., Junninen, H., Aalto, P. P., Hillamo, R., Mäkelä, T., Keronen, P., Siivola, E., Vesala, T., and Kulmala, M.: The urban measurement station SMEAR III: Continuous monitoring of air pollution and surface-atmosphere interactions in Helsinki, Finland, Tech. rep., 2009.

Järvinen, E., Virkkula, A., Nieminen, T., Aalto, P. P., Asmi, E., Lanconelli, C., Busetto, M., Lupi, A., Schioppo, R., Vitale, V., Mazzola, M., Petäjä, T., Kerminen, V.-M., and Kulmala, M.: Seasonal cycle and modal structure of particle number size distribution at Dome C, Antarctica, Atmospheric Chemistry and Physics, 13, 7473–7487, https://doi.org/10.5194/acp-13-7473-2013, 2013.

Kalivitis, N., Kerminen, V. M., Kouvarakis, G., Stavroulas, I., Tzitzikalaki, E., Kalkavouras, P., Daskalakis, N., Myriokefalitakis, S., Bougiatioti, A., Manninen, H. E., Roldin, P., Petäjä, T., Boy, M., Kulmala, M., Kanakidou, M., and Mihalopoulos, N.: Formation and growth of atmospheric nanoparticles in the eastern Mediterranean: Results from long-term measurements and process simulations, Atmospheric Chemistry and Physics, 19, 2671–2686, https://doi.org/10.5194/acp-19-2671-2019, 2019.

Karlsson, L., Krejci, R., Koike, M., Ebell, K., and Zieger, P.: A long-term study of cloud residuals from low-level Arctic clouds, Atmospheric Chemistry and Physics, 21, 8933–8959, https://doi.org/10.5194/acp-21-8933-2021, 2021.

Kecorius, S., Madueño, L., Lovric, M., Racic, N., Schwarz, M., Cyrys, J., Casquero-Vera, J. A., Alados-Arboledas, L., Conil, S., Sciare, J., Ondracek, J., Hallar, A. G., Gómez-Moreno, F. J., Ellul, R., Kristensson, A., Sorribas, M., Kalivitis, N., Mihalopoulos, N., Peters, A., Gini, M., Eleftheriadis, K., Vratolis, S., Jeongeun, K., Birmili, W., Bergmans, B., Nikolova, N., Dinoi, A., Contini, D., Marinoni, A., Alastuey, A., Petäjä, T., Rodriguez, S., Picard, D., Brem, B., Priestman, M., Green, D. C., Beddows, D. C. S., Harrison, R. M., O'Dowd, C., Ceburnis, D., Hyvärinen, A., Henzing, B., Crumeyrolle, S., Putaud, J.-P., Laj, P., Weinhold, K., Plauškaitė, K., and Byčenkienė, S.: Atmospheric new particle formation identifier using longitudinal global particle number size distribution data, Scientific Data, 11, 1239, https://doi.org/10.1038/s41597-024-04079-1, 2024.

Kendall, M. G.: A NEW MEASURE OF RANK CORRELATION, Biometrika, 30, 81–93, https://doi.org/10.1093/biomet/30.1-2.81, 1938.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Kerminen, V. M., Paramonov, M., Anttila, T., Riipinen, I., Fountoukis, C., Korhonen, H., Asmi, E., Laakso, L., Lihavainen, H., Swietlicki, E., Svenningsson, B., Asmi, A., Pandis, S. N., Kulmala, M., and Petäjä, T.: Cloud condensation nuclei production associated with atmospheric nucleation: A synthesis based on existing literature and new results, Atmospheric Chemistry and Physics, 12, 12 037–12 059, https://doi.org/10.5194/acp-12-12037-2012, 2012.

Kesti, J., Asmi, E., O'Connor, E. J., Backman, J., Budhavant, K., Andersson, A., Dasari, S., Praveen, P. S., Zahid, H., and Örjan Gustafsson: Changes in aerosol size distributions over the Indian Ocean during different meteorological conditions, Tellus, Series B: Chemical and Physical Meteorology, 72, 1–14, https://doi.org/10.1080/16000889.2020.1792756, 2020.

Kesti, J., Backman, J., O'Connor, E. J., Hirsikko, A., Asmi, E., Aurela, M., Makkonen, U., Filioglou, M., Komppula, M., Korhonen, H., and Lihavainen, H.: Aerosol particle characteristics measured in the United Arab Emirates and their response to mixing in the boundary layer, Atmospheric Chemistry and Physics, 22, 481–503, https://doi.org/10.5194/acp-22-481-2022, 2022.

Kesti, J., O'Connor, E. J., Hirsikko, A., Backman, J., Filioglou, M., Sundström, A.-M., Tonttila, J., Lihavainen, H., Korhonen, H., and Asmi, E.: How horizontal transport and turbulent mixing impact aerosol particle and precursor concentrations at a background site in the UAE, Atmospheric Chemistry and Physics, 24, 9369–9386, https://doi.org/10.5194/acp-24-9369-2024, 2024.

Kivekäs, N., Sun, J., Zhan, M., Kerminen, V.-M., Hyvärinen, A., Komppula, M., Viisanen, Y., Hong, N., Zhang, Y., Kulmala, M., Zhang, X.-C., and Lihavainen, H.: Long term particle size distribution measurements at Mount Waliguan, a high-altitude site in inland China, Atmospheric Chemistry and Physics, 9, 5461–5474, https://doi.org/10.5194/acp-9-5461-2009, 2009.

Kristensson, A., Maso, M. D., Swietlicki, E., Hussein, T., Zhou, J., Kerminen, V. M., and Kulmala, M.: Characterization of new particle formation events at a background site in southern Sweden: Relation to air mass history, Tellus, Series B: Chemical and Physical Meteorology, 60 B, 330–344, https://doi.org/10.1111/j.1600-0889.2008.00345.x, 2008.

Kubelová, L., Vodička, P., Schwarz, J., Cusack, M., Makeš, O., Ondráček, J., and Ždímal, V.: A study of summer and winter highly time-resolved submicron aerosol composition measured at a suburban site in Prague, Atmospheric Environment, 118, 45–57, https://doi.org/10.1016/J.ATMOSENV.2015.07.030, 2015.

Laakso, L., Laakso, H., Aalto, P. P., Keronen, P., Petäjä, T., Nieminen, T., Pohja, T., Siivola, E., Kulmala, M., Kgabi, N., Molefe, M., Mabaso, D., Phalatse, D., Pienaar, K., and Kerminen, V.-M.: Basic characteristics of atmospheric particles, trace gases and meteorology in a relatively clean Southern African Savannah environment, Atmos. Chem. Phys, 8, 4823–4839, www.atmos-chem-phys.net/8/4823/2008/, 2008.

Lampilahti, A., Garmash, O., Aliaga, D., Arshinov, M., Davydov, D., Belan, B., Lampilahti, J., Kerminen, V.-M., Petäjä, T., Kulmala, M., and Ezhova, E.: Insights into new particle formation in a Siberian boreal forest from nanoparticle ranking analysis, Aerosol Research, 3, 441–459, https://doi.org/10.5194/ar-3-441-2025, 2025.

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

Leinonen, V., Kokkola, H., Yli-Juuti, T., Mielonen, T., Kühn, T., Nieminen, T., Heikkinen, S., Miinalainen, T., Bergman, T., Carslaw, K., Decesari, S., Fiebig, M., Hussein, T., Kivekäs, N., Krejci, R., Kulmala, M., Leskinen, A., Massling, A., Mihalopoulos, N., Mulcahy, J. P., Noe, S. M., Noije, T. V., O'connor, F. M., O'dowd, C., Olivie, D., Pernov, J. B., Petäjä, T., Øyvind Seland, Schulz, M., Scott, C. E., Skov, H., Swietlicki, E., Tuch, T., Wiedensohler, A., Virtanen, A., and Mikkonen, S.: Comparison of particle number size distribution trends in ground measurements and climate models, Atmospheric Chemistry and Physics, 22, 12 873–12 905, https://doi.org/10.5194/acp-22-12873-2022, 2022.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Leskinen, A., Portin, H., Komppula, M., Miettinen, P., Arola, A., Lihavainen, H., Hatakka, J., Laaksonen, A., Lehtinen, K. E. J., Leskinen, A., Komppula, M., Arola, P., Lihavainen, A., Laaksonen, J., and Lehtinen, A. .: Overview of the research activities and results at Puijo semi-urban measurement station, Boreal environment research, 14, 576–590, https://doi.org/https://doi.org/10.60910/zr01-u9e6, 2009.

Lihavainen, H., Alghamdi, M. A., Hyvärinen, A. P., Hussein, T., Aaltonen, V., Abdelmaksoud, A. S., Al-Jeelani, H., Almazroui, M., Almehmadi, F. M., Zawad, F. M. A., Hakala, J., Khoder, M., Neitola, K., Petäjä, T., Shabbaj, I. I., and Hämeri, K.: Aerosols physical properties at Hada Al Sham, western Saudi Arabia, Atmospheric Environment, 135, 109–117, https://doi.org/10.1016/j.atmosenv.2016.04.001, 2016.

Liu, X., Gouveia, D. A., Henzing, B., Apituley, A., Hensen, A., van Dinther, D., Huang, R., and Dusek, U.: Aerosol optical properties within the atmospheric boundary layer predicted from ground-based observations compared to Raman lidar retrievals during RITA-2021, Atmospheric Chemistry and Physics, 24, 9597–9614, https://doi.org/10.5194/acp-24-9597-2024, 2024.

Liu, Y., Yan, C., Feng, Z., Zheng, F., Fan, X., Zhang, Y., Li, C., Zhou, Y., Lin, Z., Guo, Y., Zhang, Y., Ma, L., Zhou, W., Liu, Z., Dada, L., Dällenbach, K., Kontkanen, J., Cai, R., Chan, T., Chu, B., Du, W., Yao, L., Wang, Y., Cai, J., Kangasluoma, J., Kokkonen, T., Kujansuu, J., Rusanen, A., Deng, C., Fu, Y., Yin, R., Li, X., Lu, Y., Liu, Y., Lian, C., Yang, D., Wang, W., Ge, M., Wang, Y., Worsnop, D. R., Junninen, H., He, H., Kerminen, V. M., Zheng, J., Wang, L., Jiang, J., Petäjä, T., Bianchi, F., and Kulmala, M.: Continuous and comprehensive atmospheric observations in Beijing: a station to understand the complex urban atmospheric environment, Big Earth Data, 4, 295–321, https://doi.org/10.1080/20964471.2020.1798707, 2020.

Mann, H. B.: Nonparametric Tests Against Trend, Econometrica, 13, 245–259, https://doi.org/10.2307/1907187, 1945.

Marinescu, P. J., Levin, E. J., Collins, D., Kreidenweis, S. M., and Heever, S. C. V. D.: Quantifying aerosol size distributions and their temporal variability in the Southern Great Plains, USA, Atmospheric Chemistry and Physics, 19, 11 985–12 006, https://doi.org/10.5194/acp-19-11985-2019, 2019.

Martin, S. T., Artaxo, P., Machado, L., Manzi, A. O., Souza, R. A., Schumacher, C., Wang, J., Biscaro, T., Brito, J., Calheiros, A., Jardine, K., Medeiros, A., Portela, B., Sá, S. S. D., Adachi, K., Aiken, A. C., Alblbrecht, R., Alexander, L., Andreae, M. O., Barbosa, H. M., Buseck, P., Chand, D., Comstmstmstock, J. M., Day, D. A., Dubey, M., Fan, J., Faststt, J., Fisch, G., Fortner, E., Giangrande, S., Gilllles, M., Goldststein, A. H., Guenther, A., Hubbbbe, J., Jensen, M., Jimenez, J. L., Keutstsch, F. N., Kim, S., Kuang, C., Laskskin, A., McKinney, K., Mei, F., Millller, M., Nascimento, R., Pauliquevis, T., Pekour, M., Peres, J., Petäjä, T., Pöhlklker, C., Pöschl, U., Rizzo, L., Schmid, B., Shilllling, J. E., Dias, M. A. S., Smith, J. N., Tomlmlinson, J. M., Tóta, J., and Wendisch, M.: The Green Ocean Amazon Experiment (GoAmazon2014/5) Observes Pollution Affecting Gases, Aerosols, Clouds, and Rainfall over the Rain Forest, Bulletin of the American Meteorological Society, 98, 981–997, https://doi.org/10.1175/BAMS-D-15-00221.1, 2017.

Merico, E., Cesari, D., Dinoi, A., Potì, S., Pennetta, A., Bloise, E., and Contini, D.: Long-term analysis of carbonaceous fractions of particulate at a Central Mediterranean site in Italy, Atmospheric Pollution Research, 16, 102 668, https://doi.org/10.1016/J.APR.2025.102668, 2025.

Mihalopoulos, N., Stephanou, E., Kanakidou, M., Pilitsidis, S., and Bousquet, P.: Tropospheric aerosol ionic composition in the Eastern Mediterranean region, Tellus, Series B: Chemical and Physical Meteorology, 49, 314–326, https://doi.org/10.3402/tellusb.v49i3.15970, 1997.

Morawska, L., Thomas, S., Jamriska, M., and Johnson, G.: The modality of particle size distributions of environmental aerosols, Atmospheric Environment, 33, 4401–4411, 1999.

Nair, A. A. and Yu, F.: Using machine learning to derive cloud condensation nuclei number concentrations from commonly available measurements, Atmospheric chemistry and physics, 20, 12 853–12 869, https://doi.org/10.5194/acp-20-12853-2020, 2020.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Namdeo, P., Chakraborty, T. C., and Chakraborty, A.: Two decades of aerosol trends over India: seasonal characteristics and urban-rural dynamics, Environmental Research Letters, 19, https://doi.org/10.1088/1748-9326/ad9291, 2024.

NFOŚiGW: „Mój Prąd" już działa! Pierwszy nabór wniosków rozpoczął się 30 sierpnia, https://www.gov.pl/web/nfosigw/moj-prad-juz-dziala-pierwszy-nabor-wnioskow-rozpoczal-sie-30-sierpnia, last access: 30 March 2026, 2021.

Nieminen, T., Kerminen, V.-M., Petaja, T., Aalto, P. P., Arshinov, M., Asmi, E., Baltensperger, U., Beddows, D. C. S., Beukes, J. P., Collins, D., Ding, A., Harrison, R. M., Henzing, B., Hooda, R., Hu, M., Horrak, U., Kivekas, N., Komsaare, K., Krejci, R., Kristensson, A., Laakso, L., Laaksonen, A., Leaitch, W. R., Lihavainen, H., Mihalopoulos, N., Nemeth, Z., Nie, W., O'Dowd, C., Salma, I., Sellegri, K., Svenningsson, B., Swietlicki, E., Tunved, P., Ulevicius, V., Vakkari, V., Vana, M., Wiedensohler, A., Wu, Z., Virtanen, A., and Kulmala, M.: Global analysis of continental boundary layer new particle formation based on long-term measurements, Atmospheric chemistry and physics, 18, 14 737–14 756, https://doi.org/10.5194/acp-18-14737-2018, 2018.

Noe, S. M., Ülo Niinemets, Krasnova, A., Krasnov, D., Motallebi, A., Kängsepp, V., Jõgiste, K., Hõrrak, U., Komsaare, K., Mirme, S., Vana, M., Tammet, H., Bäck, J., Vesala, T., Kulmala, M., Petäjä, T., and Kangur, A.: SMEAR Estonia: Perspectives of a large-scale forest ecosystem– Atmosphere research infrastructure, Forestry Studies, 63, 56–84, https://doi.org/10.1515/fsmu-2015-0009, 2015.

O'Connor, T. C., Jennings, S. G., and O'Dowd, C. D.: Highlights of fifty years of atmospheric aerosol research at Mace Head, Atmospheric Research, 90, 338–355, https://doi.org/10.1016/J.ATMOSRES.2008.08.014, 2008.

Ovaska, A., Rauth, E., Holmberg, D., Artaxo, P., Backman, J., Bergmans, B., Collins, D., Franco, M. A., Gani, S., Harrison, R. M., Hooda, R. K., Hussein, T., Hyvärinen, A. P., Jaars, K., Kristensson, A., Kulmala, M., Laakso, L., Laaksonen, A., Mihalopoulos, N., O'Dowd, C., Ondracek, J., Petäjä, T., Plauškaitė, K., Pöhlker, M., Qi, X., Tunved, P., Vakkari, V., Wiedensohler, A., Puolamäki, K., Nieminen, T., Kerminen, V. M., Sinclair, V. A., and Paasonen, P.: Global fields of daily accumulation-mode particle number concentrations using in situ observations, reanalysis data, and machine learning, Aerosol Research, 3, 589–618, https://doi.org/10.5194/ar-3-589-2025, 2025.

Ovaska, A., Holmberg, D., Rauth, E., Alghamdi, M., Artaxo, P., Asmi, E., Backman, J., Bergmans, B., Boyer, M., Bredenkamp, L., Busetto, M., Campos De Oliveir, D., Casquero Vera, J. A., Ceburnis, D., Chan, T., Chan, T., Conil, S., Contini, D., Crumeyrolle, S., Duflot, V., Eleftheriadis, K., Esveld, J., Ezhova, E., Fiebig, M., Gani, S., Garmash, O., Gómez-Moreno, F. J., Harrison, R. M., Hoffer, A., Hooda, R. K., Hyvärinen, A., Hussein, T., Joutsensaari, J., Kalivitis, N., Kaminski, H., Kesti, J., Krejci, R., Kristensson, A., Kuang, C., Kulmala, M., Laakso, L., Leskinen, A., Lihavainen, H., Massling, A., Merkel, M., Noe, S., Ondracek, J., Perez, N., Petit, J.-E., Petäjä, T., Pikridas, M., Pöhlker, C., Pöhlker, M., Putaud, J. P., Qi, X., Reche, C., Rodríguez, S., Roztocil, P., Sciare, J., Sellegri, K., Shang, D., Singh, A., Sipilä, M., Skov, H., Sorribas, M., Subba, T., Sun, J., Tunved, P., Vakkari, V., van Zyl, P. G., Virkkula, A., Voigtländer, J., Weinhold, K., Wiedensohler, A., Yoo, H.-J., Zhou, P., Puolamäki, K., Nieminen, T., Kerminen, V.-M., Sinclair, V., and Paasonen, P.: Daily Measurements of Number Concentrations (N100 and N50) at 74 stations (2003-2024), https://doi.org/10.5281/zenodo.20200333, 2026a.

Ovaska, A., Holmberg, D., Rauth, E., Mansour, A., Artaxo, P., Asmi, E., Backman, J., Bergmans, B., Boyer, M., Bredenkamp, L., Busetto, M., Campos De Oliveir, D., Casquero Vera, J. A., Ceburnis, D., Chan, T., Chan, T., Conil, S., Contini, D., Crumeyrolle, S., Duflot, V., Eleftheriadis, K., Esveld, J., Ezhova, E., Fiebig, M., Gani, S., Garmash, O., Gómez-Moreno, F. J., Harrison, R. M., Hoffer, A., Hooda, R. K., Hyvärinen, A., Hussein, T., Joutsensaari, J., Kalivitis, N., Kaminski, H., Kesti, J., Krejci, R., Kristensson, A., Kuang, C., Kulmala, M., Laakso, L., Leskinen, A., Lihavainen, H., Massling, A., Merkel, M., Noe, S., Ondracek, J., Perez, N., Petit, J.-E., Petäjä, T., Pikridas, M., Pöhlker, C., Pöhlker, M., Putaud, J. P., Qi, X., Reche, C., Rodríguez, S., Roztocil, P., Sciare, J., Sellegri, K., Shang, D., Singh, A., Sipilä, M., Skov, H., Sorribas, M., Subba, T., Sun, J., Tunved, P., Vakkari, V., van Zyl, P., Virkkula, A., Voigtländer, J., Weinhold, K., Wiedensohler, A., Yoo, H.-J., Zhou, P., Puolamäki, K., Nieminen, T., Kerminen, V.-M., Sinclair, V., and Paasonen, P.: A 21-year global

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

dataset of particle number concentrations (N100 and N50) for aerosol-cloud interaction studies, https://doi.org/10.5281/zenodo.20202080, 2026b.

Paasonen, P., Asmi, A., Petäjä, T., Kajos, M. K., Äijälä, M., Junninen, H., Holst, T., Abbatt, J. P. D., Arneth, A., Birmili, W., Gon, H. D. V. D., Hamed, A., Hoffer, A., Laakso, L., Laaksonen, A., Leaitch, W. R., Plass-dülmer, C., Pryor, S. C., Räisänen, P., Swietlicki, E., Wiedensohler, A., Worsnop, D. R., matti Kerminen, V., and Kulmala, M.: Warming-induced increase in aerosol number concentration likely to moderate climate change, Nature Geoscience, 6, 438–442, https://doi.org/http://dx.doi.org.libproxy.helsinki.fi/10.1038/ngeo1800, 2013.

Paasonen, P., Kupiainen, K., Klimont, Z., Visschedijk, A., Gon, H. A. V. D., and Amann, M.: Continental anthropogenic primary particle number emissions, Atmospheric Chemistry and Physics, 16, 6823–6840, https://doi.org/10.5194/acp-16-6823-2016, 2016.

Pernov, J. B., Beddows, D., Thomas, D. C., Dall´Osto, M., Harrison, R. M., Schmale, J., Skov, H., and Massling, A.: Increased aerosol concentrations in the High Arctic attributable to changing atmospheric transport patterns, npj Climate and Atmospheric Science, 5, https://doi.org/10.1038/s41612-022-00286-y, 2022.

Petit, J.-E., Favez, O., Sciare, J., Crenn, V., Sarda-Estève, R., Bonnaire, N., Močnik, G., Dupont, J.-C., Haeffelin, M., and Leoz-Garziandia, E.: Two years of near real-time chemical composition of submicron aerosols in the region of Paris using an Aerosol Chemical Speciation Monitor (ACSM) and a multi-wavelength Aethalometer, Atmospheric Chemistry and Physics, 15, 2985–3005, https://doi.org/10.5194/acp-15-2985-2015, 2015.

Pierce, J. R., Westervelt, D. M., Atwood, S. A., Barnes, E. A., and Leaitch, W. R.: New-particle formation, growth and climate-relevant particle production in egbert, canada: Analysis from 1 year of size-distribution observations, Atmospheric Chemistry and Physics, 14, 8647–8663, https://doi.org/10.5194/acp-14-8647-2014, 2014.

Pikridas, M., Vrekoussis, M., Sciare, J., Kleanthous, S., Vasiliadou, E., Kizas, C., Savvides, C., and Mihalopoulos, N.: Spatial and temporal (short and long-term) variability of submicron, fine and sub-10μm particulate matter (PM1, PM2.5, PM10) in Cyprus, Atmospheric Environment, 191, 79–93, https://doi.org/10.1016/J.ATMOSENV.2018.07.048, 2018.

Platt, S. M., Øystein Hov, Berg, T., Breivik, K., Eckhardt, S., Eleftheriadis, K., Evangeliou, N., Fiebig, M., Fisher, R., Hansen, G., Hansson, H.-C., Heintzenberg, J., Hermansen, O., Heslin-Rees, D., Holmén, K., Hudson, S., Kallenborn, R., Krejci, R., Krognes, T., Larssen, S., Lowry, D., Myhre, C. L., Lunder, C., Nisbet, E., Nizzetto, P. B., Park, K.-T., Pedersen, C. A., Pfaffhuber, K. A., Röckmann, T., Schmidbauer, N., Solberg, S., Stohl, A., Ström, J., Svendby, T., Tunved, P., Tørnkvist, K., van der Veen, C., Vratolis, S., Yoon, Y. J., Yttri, K. E., Zieger, P., Aas, W., and Tørseth, K.: Atmospheric composition in the European Arctic and 30 years of the Zeppelin Observatory, Ny-Ålesund, Atmospheric Chemistry and Physics, 22, 3321–3369, https://doi.org/10.5194/acp-22-3321-2022, 2022.

Putaud, J.-P., Pozzoli, L., Pisoni, E., Santos, S. M. D., Lagler, F., Lanzani, G., Santo, U. D., and Colette, A.: Impacts of the COVID-19 lockdown on air pollution at regional and urban background sites in northern Italy, Atmospheric Chemistry and Physics, 21, 7597–7609, https://doi.org/10.5194/acp-21-7597-2021, 2021.

Pöhlker, M. L., Zhang, M., Braga, R. C., Krüger, O. O., Pöschl, U., and Ervens, B.: Aitken mode particles as CCN in aerosol- And updraft-sensitive regimes of cloud droplet formation, Atmospheric Chemistry and Physics, 21, 11 723–11 740, https://doi.org/10.5194/acp-21-11723-2021, 2021.

Qi, X. M., Ding, A. J., Nie, W., Petäjä, T., Kerminen, V. M., Herrmann, E., Xie, Y. N., Zheng, L. F., Manninen, H., Aalto, P., Sun, J. N., Xu, Z. N., Chi, X. G., Huang, X., Boy, M., Virkkula, A., Yang, X. Q., Fu, C. B., and Kulmala, M.: Aerosol size distribution and new particle formation in the western Yangtze River Delta of China: 2 years of measurements at the SORPES station, Atmospheric Chemistry and Physics, 15, 12 445–12 464, https://doi.org/10.5194/acp-15-12445-2015, 2015.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Quaas, J., Arola, A., Cairns, B., Christensen, M., Deneke, H., Ekman, A. M., Feingold, G., Fridlind, A., Gryspeerdt, E., Hasekamp, O., Li, Z., Lipponen, A., Mülmenstädt, J., Nenes, A., Penner, J. E., Rosenfeld, D., Schrödner, R., Sinclair, K., Sourdeval, O., Stier, P., Tesche, M., Diedenhoven, B. V., and Wendisch, M.: Constraining the Twomey effect from satellite observations: Issues and perspectives, Atmospheric Chemistry and Physics, 20, 15 079–15 099, https://doi.org/10.5194/acp-20-15079-2020, 2020.

Randerson, J. T., van der Werf, G. R., Giglio, L., Collatz, G. J., and Kasibhatla, P. S.: Global Fire Emissions Database, Version 4.1 (GFEDv4), https://doi.org/https://doi.org/10.3334/ORNLDAAC/1293, 2017.

Rodríguez, S., Dingenen, R. V., Putaud, J. P., Santos, S. M.-D., and Roselli, D.: Nucleation and growth of new particles in the rural atmosphere of Northern Italy - Relationship to air quality monitoring, Atmospheric Environment, 39, 6734–6746, https://doi.org/10.1016/j.atmosenv.2005.07.036, 2005.

Rosenfeld, D., Andreae, M. O., Asmi, A., Chin, M., Leeuw, G., Donovan, D. P., Kahn, R., Kinne, S., Kivekäs, N., Kulmala, M., Lau, W., Schmidt, K. S., Suni, T., Wagner, T., Wild, M., and Quaas, J.: Global observations of aerosol-cloud-precipitation- climate interactions, Rev. Geophys., 52, 750–808, https://doi.org/10.1002/2013RG000441.Received, 2014.

Rosenfeld, D., Kokhanovsky, A., Goren, T., Gryspeerdt, E., Hasekamp, O., Jia, H., Lopatin, A., Quaas, J., Pan, Z., and Sourdeval, O.: Frontiers in Satellite-Based Estimates of Cloud-Mediated Aerosol Forcing, Reviews of Geophysics, 61, https://doi.org/10.1029/2022RG000799, 2023.

Russell, L. M., Lubin, D., Silber, I., Eloranta, E., Muelmenstaedt, J., Burrows, S., Aiken, A., Wang, D., Petters, M., Miller, M., Ackerman, A., Fridlind, A., Witte, M., Lebsock, M., Painemal, D., Chang, R., Liggio, J., and Wheeler, M.: EPCAPE Field: Final Campaign Report, Tech. rep., https://doi.org/10.2172/3015487, 2025.

Rönkkö, T., Kuuluvainen, H., Karjalainen, P., Keskinen, J., Hillamo, R., Niemi, J. V., Pirjola, L., Timonen, H. J., Saarikoski, S., Saukko, E., Järvinen, A., Silvennoinen, H., Rostedt, A., Olin, M., Yli-Ojanperä, J., Nousiainen, P., Kousa, A., and Maso, M. D.: Traffic is a major source of atmospheric nanocluster aerosol, Proceedings of the National Academy of Sciences of the United States of America, 114, 7549–7554, https://doi.org/10.1073/pnas.1700830114, 2017.

Schladitz, A., Leníček, J., Beneš, I., Kováč, M., Skorkovský, J., Soukup, A., Jandlová, J., Poulain, L., Plachá, H., Löschau, G., and Wiedensohler, A.: Air quality in the German-Czech border region: A focus on harmful fractions of PM and ultrafine particles, Atmospheric Environment, 122, 236–249, https://doi.org/10.1016/j.atmosenv.2015.09.044, 2015.

Schmale, J., Henning, S., Decesari, S., Henzing, B., Keskinen, H., Sellegri, K., Ovadnevaite, J., Pöhlker, M., Brito, J., Bougiatioti, A., Kristensson, A., Kalivitis, N., Stavroulas, I., Carbone, S., Jefferson, A., Park, M., Schlag, P., Iwamoto, Y., Aalto, P., Äijälä, M., Bukowiecki, N., Ehn, M., Fröhlich, R., Frumau, A., Herrmann, E., Herrmann, H., Holzinger, R., Kos, G., Kulmala, M., Mihalopoulos, N., Nenes, A., O'Dowd, C., Petäjä, T., Picard, D., Pöhlker, C., Pöschl, U., Poulain, L., Swietlicki, E., Aneae, M., Artaxo, P., Wiedensohler, A., Ogren, J., Matsuki, A., Yum, S. S., Stratmann, F., Baltensperger, U., and Gysel, M.: Long-term cloud condensation nuclei number concentration, particle number size distribution and chemical composition measurements at regionally representative observatories, Atmospheric chemistry and physics, 18, 2853–2881, https://doi.org/10.5194/acp-18-2853-2018, 2018.

Shen, X., Sun, J., Ma, Q., Zhang, Y., Zhong, J., Yue, Y., Xia, C., Hu, X., Zhang, S., and Zhang, X.: Long-term trend of new particle formation events in the Yangtze River Delta, China and its influencing factors: 7-year dataset analysis, Science of the Total Environment, 807, https://doi.org/10.1016/j.scitotenv.2021.150783, 2022.

Singh, A., Kuang, C., Howie, J., Salwen, C., and Hayes, C.: Scanning mobility particle sizer (AOSSMPS), 2014-01-29 to 2015-12-01, ARM Mobile Facility (MAO), Manacapuru, Amazonas, Brazil; MAOS (S1), https://doi.org/10.5439/1476898, 2014.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Singh, A., Kuang, C., Howie, J., Salwen, C., and Hayes, C.: Scanning mobility particle sizer (AOSSMPS), 2016-05-20 to 2017-10-20, ARM Mobile Facility (ASI), Ascension Island, South Atlantic Ocean; AMF1 (M1), https://doi.org/10.5439/1476898, 2016a.

Singh, A., Kuang, C., Howie, J., Salwen, C., and Hayes, C.: Scanning mobility particle sizer (AOSSMPS), 2016-11-15 to 2026-05-14, Southern Great Plains (SGP), Lamont, OK (Extended and Co-located with C1) (E13), https://doi.org/10.5439/1476898, 2016b.

Singh, A., Kuang, C., Howie, J., Salwen, C., and Hayes, C.: Scanning mobility particle sizer (AOSSMPS), 2021-10-27 to 2023-06-16, ARM Mobile Facility (GUC), Gunnison, CO; Supplemental Facility 2 (S2), https://doi.org/10.5439/1476898, 2021.

Singh, A., Kuang, C., Howie, J., Salwen, C., and Hayes, C.: Scanning mobility particle sizer (AOSSMPS), 2022-11-01 to 2026-05-14, Eastern North Atlantic (ENA), Graciosa Island, Azores, Portugal (C1), https://doi.org/10.5439/1476898, 2022.

Singh, A., Kuang, C., Howie, J., Salwen, C., and Hayes, C.: Scanning mobility particle sizer (AOSSMPS), 2023-02-15 to 2024-02-14, ARM Mobile Facility (EPC), La Jolla, CA; AMF1 (main site for EPCAPE on Scripps Pier) (M1), https://doi.org/10.5439/1476898, 2023.

Slowik, J. G., Stroud, C., Bottenheim, J. W., Brickell, P. C., Chang, R. Y.-W., Liggio, J., Makar, P. A., Martin, R. V., Moran, M. D., Shantz, N. C., Sjostedt, S. J., van Donkelaar, A., Vlasenko, A., Wiebe, H. A., Xia, A. G., Zhang, J., Leaitch, W. R., and Abbatt, J. P. D.: Characterization of a large biogenic secondary organic aerosol event from eastern Canadian forests, Atmospheric Chemistry and Physics, 10, 2825–2845, https://doi.org/10.5194/acp-10-2825-2010, 2010.

SmartSMEAR: https://smear.avaa.csc.fi/, last access: 19 May 2026.

Sorribas, M., Adame, J. A., Olmo, F. J., Vilaplana, J. M., Gil-Ojeda, M., and Alados-Arboledas, L.: A long-term study of new particle formation in a coastal environment: Meteorology, gas phase and solar radiation implications, Science of The Total Environment, 511, 723–737, https://doi.org/10.1016/J.SCITOTENV.2014.12.011, 2015.

Stier, P., van den Heever, S. C., Christensen, M. W., Gryspeerdt, E., Dagan, G., Saleeby, S. M., Bollasina, M., Donner, L., Emanuel, K., Ekman, A. M., Feingold, G., Field, P., Forster, P., Haywood, J., Kahn, R., Koren, I., Kummerow, C., L'Ecuyer, T., Lohmann, U., Ming, Y., Myhre, G., Quaas, J., Rosenfeld, D., Samset, B., Seifert, A., Stephens, G., and Tao, W. K.: Multifaceted aerosol effects on precipitation, https://doi.org/10.1038/s41561-024-01482-6, 2024.

Suchánková, L., Crumeyrolle, S., Bourrianne, E., Prokeš, R., Holoubek, I., Ždímal, V., and Chiapello, I.: An insight into recent PM1 aerosol light scattering properties and particle number concentration variabilities at the suburban site ATOLL in Northern France, Science of The Total Environment, 959, 178 190, https://doi.org/10.1016/J.SCITOTENV.2024.178190, 2025.

Tammet, H., Mirme, A., and Tamm, E.: Electrical aerosol spectrometer of Tartu University, Atmospheric Research, 62, https://doi.org/https://doi.org/10.1016/S0169-8095(02)00017-0, 2002.

Tunved, P. and Ström, J.: On the seasonal variation in observed size distributions in northern Europe and their changes with decreasing anthropogenic emissions in Europe: Climatology and trend analysis based on 17 years of data from Aspvreten, Sweden, Atmospheric Chemistry and Physics, 19, 14 849–14 873, https://doi.org/10.5194/acp-19-14849-2019, 2019.

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

UBA: Das Luftmessnetz des Umweltbundesamtes: Langzeitmessungen, Prozessverständnis und Wirkungen ferntransportierter Luftverunreinigungen, Tech. Rep., edited by Schleyer R and Bieber E and Wallasch M, Umweltbundesamt (UBA), 2013.

Vakkari, V., Beukes, J. P., Laakso, H., Mabaso, D., Pienaar, J. J., Kulmala, M., and Laakso, L.: Long-term observations of aerosol size distributions in semi-clean and polluted savannah in South Africa, Atmospheric Chemistry and Physics, 13, 1751–1770, https://doi.org/10.5194/acp-13-1751-2013, 2013.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

van der A, R. J., Mijling, B., Ding, J., Koukouli, M. E., Liu, F., Li, Q., Mao, H., and Theys, N.: Cleaning up the air: Effectiveness of air quality policy for SO2 and NOx emissions in China, Atmospheric Chemistry and Physics, 17, 1775–1789, https://doi.org/10.5194/acp-17-1775-2017, 2017.

von Salzen, K., Akingunola, A., Cole, J. N., Digby, R. A., Doherty, S. J., Fraser-Leach, L., Gryspeerdt, E., Sigmond, M., and Wood, R.: Reduced aerosol pollution diminished cloud reflectivity over the North Atlantic and Northeast Pacific, Nature Communications, 16, https://doi.org/10.1038/s41467-025-65127-x, 2025.

Wang, J., Wood, R., Jensen, M. P., Chiu, J. C., Liu, Y., Lamer, K., Desai, N., Giangrande, S. E., Knopf, D. A., Kollias, P., Laskin, A., Liu, X., Lu, C., Mechem, D., Mei, F., Starzec, M., Tomlinson, J., Wang, Y., Yum, S. S., Zheng, G., Aiken, A. C., Azevedo, E. B., Blanchard, Y., China, S., Dong, X., Gallo, F., Gao, S., Ghate, V. P., Glienke, S., Goldberger, L., Hardin, J. C., Kuang, C., Luke, E. P., Matthews, A. A., Miller, M. A., Moffet, R., Pekour, M., Schmid, B., Sedlacek, A. J., Shaw, R. A., Shilling, J. E., Sullivan, A., Suski, K., Veghte, D. P., Weber, R., Wyant, M., Yeom, J., Zawadowicz, M., and Zhang, Z.: Aerosol and Cloud Experiments in the Eastern North Atlantic (ACE-ENA), Bulletin of the American Meteorological Society, 103, E619–E641, https://doi.org/10.1175/BAMS-D-19-0220.1, 2022.

Wang, J., Xu, G., Chen, L., and Chen, K.: Atmospheric Particle Number Concentrations and New Particle Formation over the Southern Ocean and Antarctica: A Critical Review, https://doi.org/10.3390/atmos14020402, 2023.

WFOŚiGW: Czyste Powietrze, https://wfosigw.pl/czyste-powietrze/, last access: 30 March 2026, 2026.

Wiedensohler, A., Birmili, W., Nowak, A., Sonntag, A., Weinhold, K., Merkel, M., Wehner, B., Tuch, T., Pfeifer, S., Fiebig, M., M., A. F., Asmi, E., Sellegri, K., Depuy, R., Venzac, H., Villani, P., Laj, P., Aalto, P., Ogren, J. A., Swietlicki, E., Williams, P., Roldin, P., Quincey, P., Hüglin, C., Fierz-Schmidhauser, R., Gysel, M., Weingartner, E., Riccobono, F., Santos, S., Grüning, C., Faloon, K., Beddows, D., Harrison, R., Monahan, C., Jennings, S. G., D., C. O., Marinoni, A., Horn, H. G., Keck, L., Jiang, J., Scheckman, J., McMurry, P. H., Deng, Z., Zhao, C. S., Moerman, M., Henzing, B., de Leeuw, G., Löschau, G., and Bastian, S.: Mobility particle size spectrometers: harmonization of technical standards and data structure to facilitate high quality long-term observations of atmospheric particle number size distributions, Atmospheric measurement techniques, 5, 657–685, https://doi.org/10.5194/amt-5-657-2012, 2012.

Wu, Z., Hu, M., Liu, S., Wehner, B., Bauer, S., ßLing, A. M., Wiedensohler, A., Petäjä, T., Maso, M. D., and Kulmala, M.: New particle formation in Beijing, China: Statistical analysis of a 1-year data set, Journal of Geophysical Research Atmospheres, 112, https://doi.org/10.1029/2006JD007406, 2007.

Yli-Juuti, T., Riipinen, I., Aalto, P. P., Nieminen, T., Maenhaut, W., Janssens, I. A., Claeys, M., Salma, I., Ocskay, R., Hoffer, A., Imre, K., and Kulmala, M.: Characteristics of new particle formation events and cluster ions at K-puszta, Hungary, Hungary. Boreal Env. Res, 14, 683–698, www.nilu.no/projects/ccc/sitedescriptions/hu/, 2009.

Yli-Juuti, T., Mielonen, T., Heikkinen, L., Arola, A., Ehn, M., Isokääntä, S., Keskinen, H. M., Kulmala, M., Laakso, A., Lipponen, A., Luoma, K., Mikkonen, S., Nieminen, T., Paasonen, P., Petäjä, T., Romakkaniemi, S., Tonttila, J., Kokkola, H., and Virtanen, A.: Significance of the organic aerosol driven climate feedback in the boreal area, Nature Communications, 12, https://doi.org/10.1038/s41467-021-25850-7, 2021.

Yttri, K. E., Canonaco, F., Eckhardt, S., Evangeliou, N., Fiebig, M., Gundersen, H., Hjellbrekke, A.-G., Myhre, C. L., Platt, S. M., Prévôt, A. S. H., Simpson, D., Solberg, S., Surratt, J., Tørseth, K., Uggerud, H., Vadset, M., Wan, X., and Aas, W.: Trends, composition, and sources of carbonaceous aerosol at the Birkenes Observatory, northern Europe, 2001–2018, Atmospheric Chemistry and Physics, 21, 7149–7170, https://doi.org/10.5194/acp-21-7149-2021, 2021.

https://doi.org/10.5194/essd-2026-415
Preprint. Discussion started: 20 July 2026
© Author(s) 2026. CC BY 4.0 License.

Yu, F., Luo, G., Nair, A. A., Tsigaridis, K., and Bauer, S. E.: Use of Machine Learning to Reduce Uncertainties in Particle Number Concentration and Aerosol Indirect Radiative Forcing Predicted by Climate Models, Geophysical Research Letters, 49, https://doi.org/10.1029/2022GL098551, 2022.

Zgłobicki, W. and Baran-Zgłobicka, B.: Air pollution in major Polish cities in the period 2005–2021: Intensity, effects and attempts to reduce it, Environmental Research, 240, https://doi.org/10.1016/j.envres.2023.117497, 2024.

Zuidema, P., Sedlacek, A. J., Flynn, C., Springston, S., Delgadillo, R., Zhang, J., Aiken, A. C., Koontz, A., and Muradyan, P.: The Ascension Island Boundary Layer in the Remote Southeast Atlantic is Often Smoky, Geophysical Research Letters, 45, 4456–4465, https://doi.org/10.1002/2017GL076926, 2018.

Zíková, N. and Ždímal, V.: Long-term measurement of aerosol number size distributions at rural background station Košetice, Aerosol and Air Quality Research, 13, 1464–1474, https://doi.org/10.4209/aaqr.2013.02.0056, 2013.