

Automated Analysis and Quality Assurance of Ice-Nucleating Particle Data: The PINE INP Analysis Software PIA

Nicole Büttner¹, Romy Fösig¹, Alexander Böhmmländer¹, Larissa Lacher¹, Franziska Vogel^{1a}, Mark Tarn², Pia Bogert¹, Jens Nadolny¹, Benjamin Murray², and Ottmar Möhler¹

¹Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany

^aNow at: Institute of Atmospheric Sciences and Climate (ISAC), National Research Council (CNR), Bologna, Italy

²School of Earth and Environment, University of Leeds, Leeds, UK

Correspondence: Nicole Büttner (nicole.buettner@kit.edu)

Abstract. The presence of ice-nucleating particles (INPs) in the atmosphere plays a crucial role in shaping cloud radiative properties, influencing their lifespan, and affecting precipitation and storm dynamics. To enable continuous and high-resolution monitoring of INP concentrations, the Portable Ice Nucleation Experiment (PINE) was developed. Complementing this, the PINE INP Analysis (PIA) software was created to ensure a standardised and reproducible data processing workflow. This work presents the setup of software version 3.0.0 and the structure of the processed data. The two main components of the software - the automated quality control of the data and the algorithm to distinguish between aerosols and droplets versus ice crystals based on their optical size - are described in detail. The second part of this study provides recommendations for quality assurance of PINE measurements. It outlines procedures for conducting background checks to detect potential contamination within the chamber, evaluates the consistency between adjacent temperature sensors, and discusses how large aerosol particles can impact measurement uncertainty.

1 Introduction

Ice-nucleating particles (INPs; Vali et al., 2015) catalyse the first formation of ice crystals in clouds and can thereby influence their radiative properties (e.g., Murray et al., 2021). An increase in the INP concentration can lead to faster cloud glaciation and associated dissipation, which in turn reduces the cloud lifetime (e.g., Lohmann and Feichter, 2005). Moreover, precipitation occurrence and rates, especially over land and in extratropical regions, are strongly influenced by ice processes (Mülmenstädt et al., 2015; Field and Heymsfield, 2015; Heymsfield et al., 2020). In addition to other cloud microphysical processes, INPs have been shown to impact the formation of precipitation (e.g., Muhlbauer and Lohmann, 2009; Diehl and Grützun, 2018; Fan et al., 2017; Lin et al., 2022), and to impact storm development and dynamics (e.g., Yang et al., 2024; Chen et al., 2019; Hazra et al., 2022). The scarcity of INPs, which are a very minor subset of atmospheric aerosol particles, challenges measurement techniques (Cziczo et al., 2017; DeMott et al., 2011). Ambient INP concentration measurements relevant for mixed-phase cloud conditions were started in the 1950s (e.g., Aufm Kampe and Weickmann, 1951; Bigg, 1957), but only in recent years long-term measurements using automated instruments became available. Continuous-flow diffusion chambers (CFDS) have been widely

used for short term INP field measurements over recent decades (e.g., Rogers, 1988; DeMott et al., 2010; Chou et al., 2011; Lacher et al., 2017). CFDCs work by creating a region of supersaturation between two ice coated temperature controllable walls. A temperature gradient results in a heat and water flux from the warm to cold wall that produces a region of defined supersaturation. More recently, automated CFDC instruments were developed (Bi et al., 2019; Brunner and Kanji, 2021). Since their initial development, CFDCs have been extensively utilised and continuously refined and validated.

More recently, a novel mobile and automated cloud expansion-type chamber was developed, the Portable Ice Nucleation Experiment PINE (Möhler et al., 2021). Based on the design and working principle of the AIDA (Aerosol Interaction and Dynamics in the Atmosphere; Möhler et al., 2003) simulation chamber, PINE establishes cloud-like conditions by rapidly expanding the air inside the chamber, thus inducing the formation of supercooled water droplets and ice crystals at temperatures relevant to mixed-phase clouds. PINE was used in several short- and long-term field campaigns to measure INPs that are active at mixed-phase cloud temperatures (e.g., Lacher et al., 2024; Vogel et al., 2024; Canzi et al., 2025; Böhmländer et al., 2025; Herbert et al., 2026). Due to its automated operation and minimal requirements for user input, PINE is well suited for INP monitoring. For example, PINE is one of the instruments used by the ACTRIS (Aerosol, Clouds and Trace Gases Research Infrastructure; Laj et al., 2024) Topical Centre for Cloud In Situ Measurements (CIS).

An important part of making PINE a reliable monitoring instrument is good data management following the FAIR principles (Findability, Accessibility, Interoperability, and Reusability; Wilkinson et al., 2016). To establish a FAIR life cycle of the data, the Python software PIA (PINE INP Analysis; Büttner and Fösig, 2024) was developed to ensure a standardised and reproducible data analysis. This paper describes the individual steps of data processing for the different data levels. The most important features of the software are the automated data quality control and the algorithm that automatically determines the size threshold for distinguishing between aerosols and droplets versus ice crystals. The functionality of this algorithm, the so-called ice threshold finder, is described in detail. Also included is a detailed overview of the quality control tests applied to the data and the resulting handling of the flagged data. The conditions for the tests are based on instrument specifications, empirical observations by the instrument operators, and statistical analysis. This study also investigates the characteristics of the temperature sensors of the PINE instrument and provides recommendations for quality assurance. To show results created by the software, this study includes a data set of a 4-month measurement period that was conducted close to Karlsruhe (Germany) from 2020 to 2021.

2 PINE Setup and Description of Raw Data

2.1 PINE Setup

This section describes the setup and the working principles of the currently manufactured PINE instrument by the company Bilfinger Nuclear & Energy Transition GmbH (Würzburg, Germany). For a more detailed description see Möhler et al. (2021). The abbreviations for the instrument sensors used by the control and the analysis software are shown in brackets. The PINE instrument is composed of several main components, including an inlet system, a temperature-controlled cloud chamber, an optical particle counter (OPC), and a control system. Upon entering PINE via the aerosol inlet, the sampled air is guided through

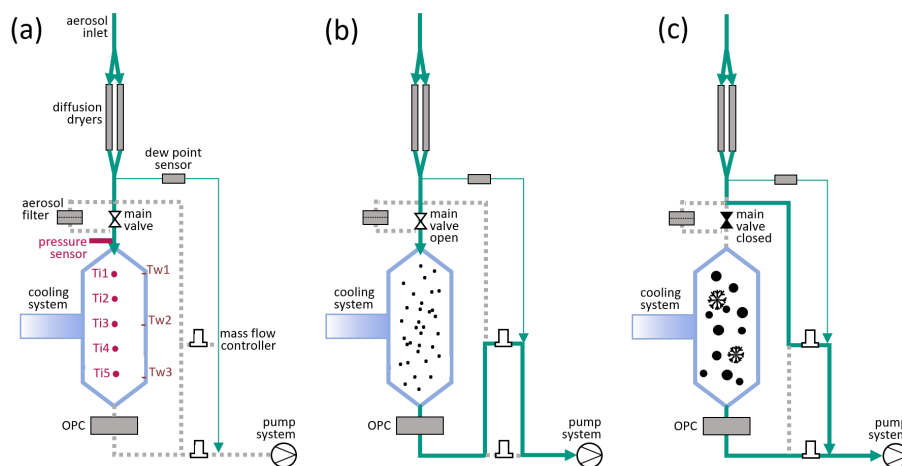


Figure 1. (a) Setup of PINE with the following main components: diffusion dryers, dew point sensor, main valve to the chamber, aerosol filter for background measurements, pressure sensor attached to the chamber, five gas temperature sensors (Ti1 – Ti5), three wall temperature sensors (Tw1 – Tw3), an optical particle counter (OPC) at the exit of the chamber, and a pump system controlling the different flows. The green lines indicate the active sample flow, while the grey dotted lines indicate no sample flow. The thin green line indicates the flow over the dew point sensor. Panel (b) shows the chamber during flush mode, the main valve is open and aerosol particles are flushed through the chamber. Panel (c) shows the expansion mode, the main valve is closed and cloud droplets and ice crystals can form. Figure adapted from Möhler et al. (2021).

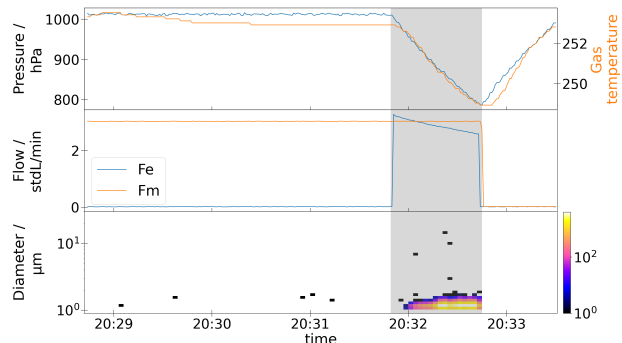


Figure 2. Example measurement of one cycle of flush mode, expansion mode, and refill mode. The shaded area shows the expansion mode.

two parallel NafionTM membrane diffusion dryers (Perma Pure, MD-700-24S-1; Fig. 1(a)). By controlling the water vapour pressure gradient across the membrane, the humidity of the sampled air is reduced to the required measurement conditions. The humidity of the sampled air is monitored with a dew point sensor (Vaisala, DRYCAP® DMT143) located downstream of the dryers and is expressed as dew point temperature (DP). Although the sensor reports the dew point above 0°C and the frost point below, this study will always use the term dew point temperature or DP for simplicity. For proper operation, the DP has to exceed the chamber temperature during expansion to enable the formation of cloud droplets, while still being sufficiently

low to prevent increased frost formation that eventually can cause an ice background in the chamber. The conditioned sampled air enters the chamber via the main valve. The cloud chamber has a volume of 10 L and is located within a vacuum chamber and cooled by a Stirling system. Five thermocouples (Thermoelement Typ K IKT 10/070 D2.5x15/2KK06/09 IEC 583-3 class 1, ES Electronic Sensor GmbH) located inside the cloud chamber (Ti1 – Ti5) measure the gas temperature, and three pt-100 temperature sensors (Pt100 RS 1/5 class B, Sensorshop 24; Tw1 – Tw3) measure the temperature of the chamber wall. A pressure sensor attached to the inlet tubes upstream of the cloud chamber monitors the chamber pressure (Pch). The OPC (fidas-pine, Palas GmbH) positioned directly below the cloud chamber determines the number and size of the particles within the sample flow.

Operation of the PINE instrument follows a cyclic sequence consisting of flushing the aerosol sample through the chamber (flush mode; Fig. 1(b)), expanding the air to create cloud droplets and ice crystals (expansion mode; Fig. 1(c); Fig. 2 shaded area), and refilling the chamber with filtered and dried air to balance the pressure to ambient conditions (refill mode). During the flush mode, the dried sample air is guided through the cooled cloud chamber at a constant mass flow rate to exchange the air inside the chamber. The photomultiplier (PM) voltage of the OPC is configured to detect relatively large particles such as large droplets and ice crystals. Therefore, it can not detect smaller aerosol particles. The expansion mode is started by closing the main valve upstream of the cloud chamber and the expansion flow (Fe) is regulated to a constant volume flow rate by a mass flow controller (MFC; ELFLOW Select F-201CV 10 l, Bronkhorst Instruments GmbH; reference conditions are $T = 273.15$ K and $p = 101\,325$ Pa) according to the current temperature and pressure conditions inside the chamber while the air is expanding. The INP concentration is therefore reported in number per standard litre (stdL^{-1}). The cooling resulting from the decrease in pressure induces an increase in relative humidity that exceeds water saturation, allowing cloud droplets and ice crystals to form in the presence of cloud condensation nuclei and INPs, respectively. As aspherical ice crystals are optically larger than spherical cloud droplets when measured with the OPC in PINE, ice crystals can later be identified on the basis of their optical size (see Sect. 4.2). During expansion, the wall temperatures remain approximately constant and therefore cause an increasing heat flux from the wall to the gas. Due to this heat source, which leads to the evaporation of the cloud droplets inside the chamber, the cloud formation and ice crystal detection time is limited to approximately 40 seconds, depending on the expansion flow rate. To maintain a stable inlet flow during the expansion a bypass flow is applied (see 1(c)). In the refill mode, ambient pressure conditions are established by refilling the chamber with filtered and dried air. One cycle of these three modes is called a run and takes about 5 to 6 minutes, depending on the duration of the flush, the expansion and refill flow, the minimum pressure, as well as ambient conditions.

The number of formed ice crystals is related to the air volume analysed during the expansion. This allows the calculation of the ice crystal number concentration, which represents the INP concentration. The temperature assigned to the measured INP concentration is the minimum temperature (Ti5) recorded at the end of the expansion. Different variables are recorded during the operations (Table 1), not only for determining the INP concentration, but also for data quality control measures (see Sect. 4.3).

The uncertainty budget of the PINE is discussed in Böhmländer et al. (2025). It can be concluded that the uncertainties are

Table 1. Overview of measurement variables.

sensor	variable
gas temperature sensors	Ti1 - Ti5
wall temperature sensors	Tw1 - Tw3
main flow	Fm
expansion flow	Fe
dew point sensor	DP
pressure sensor	Pch

dominated by the uncertainty of the OPC, which is considered to be 10% of the INP concentration. For low INP concentrations statistical uncertainties given by

$$\frac{n_{\text{INP}}}{\sqrt{N_{\text{INP}}}}, \tag{1}$$

100 where n_{INP} is the INP concentration and N_{INP} the INP count, become also relevant.

2.2 Raw Data

The PINE instrument is operated using an in-house developed LabVIEW control software that also manages the data acquisition. The control software creates a directory with the campaign name and stores multiple .txt files containing both measurement data and metadata. If the campaign name changes, a new directory is generated automatically. Figure A1 shows the structure of the raw data. The PINE measurements are structured into runs and operations. One run is composed of one cycle of flush, expansion, and refill. One operation consists of multiple runs and is linked to certain measurement settings (e.g., temperature settings, filter for background measurements).

105 The files in the *operation setup* sub-directory contain the metadata from each operation. The relevant information for the analysis software is the ID of the OPC (*opc-id*) used in this setup and the name of the file used to map the OPC bin number to the particle diameter (*calib-file*).

110 The sub-directory *raw_Data* contains the files with the run-time information per operation. The names of these files start with *pfr* and will be referred to as *run files*. They also contain the pre-set flow rates and the end pressure during expansion. The sub-directory *opc-id* contains the raw data of the OPC per run. The file names end with *opc.txt* and will be referred to as *opc files*.

115 The sub-directory *housekeeping* contains files with the endings *valve.txt* and *instrument.txt* for each operation. The *valve.txt* files contain information about the opening and closing times of the valves and are not used for the analysis, but can be useful for troubleshooting. The *instrument.txt* files, referred to as *instrument files*, contain the data from the sensors inside the chamber at a 1 second resolution. The instrument files report the measurements of the following sensors: gas temperature 1 - 5

(Ti1 - Ti5), wall temperature 1 - 3 (Tw1 - Tw3), chamber pressure (Pch), main mass flow rate through the inlet (Fm), mass flow rate during expansion (Fe), and dew point temperature (DP).

The file starting with *pfo* will be referred to as *operation file*. It gives an overview of all operations in one campaign including the start and end times.

The data presented in this work were created with an older LabVIEW control software version that stored the run files in the *L0_Data* directory instead of *raw_Data*. The structure described here is that of the LabVIEW control software version 3.12.

125

An electronic logbook is needed for the analysis of the data. At the moment this logbook needs to be created manually as an Excel sheet. A template can be found within the software and should be stored under the campaign directory named *Logbook_<pine-id>_<campaign>.xlsx*. It needs to contain the operation type and the aerosol type (e.g. ambient air) for each operation. The operation type can be chosen from the following: background, constant temperature, temperature ramp, cirrus and test.

130

3 Data Set

The data set used in this study was created during the CORONA (Characterization Of Rural aerOsol and ice Nucleating pArticles) campaign that took place at Karlsruhe Institute of Technology (KIT) Campus North, close to Karlsruhe (Germany), from 2020 to 2022. The campaign was designed to monitor the influence of the pandemic and the corresponding measures on the atmospheric aerosol properties. The campus is located in a forest approximately 10 km north of Karlsruhe at 49° 5' 42.61" N, 8° 25' 45.91" E, 114 m above sea level. During this campaign, multiple aerosol instruments (e.g. Condensation Particle Counter (CPC), Scanning Mobility Particle Sizer (SMPS), and Aerodynamic Particle Sizer (APS)) were deployed and filter samples were collected for offline analysis with INSEKT (Ice Nucleation Spectrometer of the Karlsruhe Institute of Technology; described in Schneider et al. (2021)). All instruments were connected to the same inlet for sampling ambient air through a PM10 sampling head. Some results from the INSEKT measurements are described in Vogel (2022). Throughout the whole campaign period, different PINE instruments were connected to the inlet. The data used here were measured by the PINE version 04-02 (PINE-04-02) during the periods 16 November 2020 - 25 February 2021 and 18 March 2021 - 22 April 2021, providing roughly four months of data in total. The first period will be referred to as CORONA campaign and the second part will be referred to as CORONA_new campaign. During these time periods, PINE was mainly operated at a constant temperature between -19 °C and -20 °C. Occasionally, temperature ramps were performed in a temperature range between -16 °C and -27 °C to extend the measured INP data towards higher and lower temperatures. The mass flow rate during flush (Fm) was varying between 2 and 3 stdL min⁻¹ and the expansion volume flow rate was set to 3 L min⁻¹. The pressure inside the chamber was reduced by 150 - 200 mbar during the expansion.

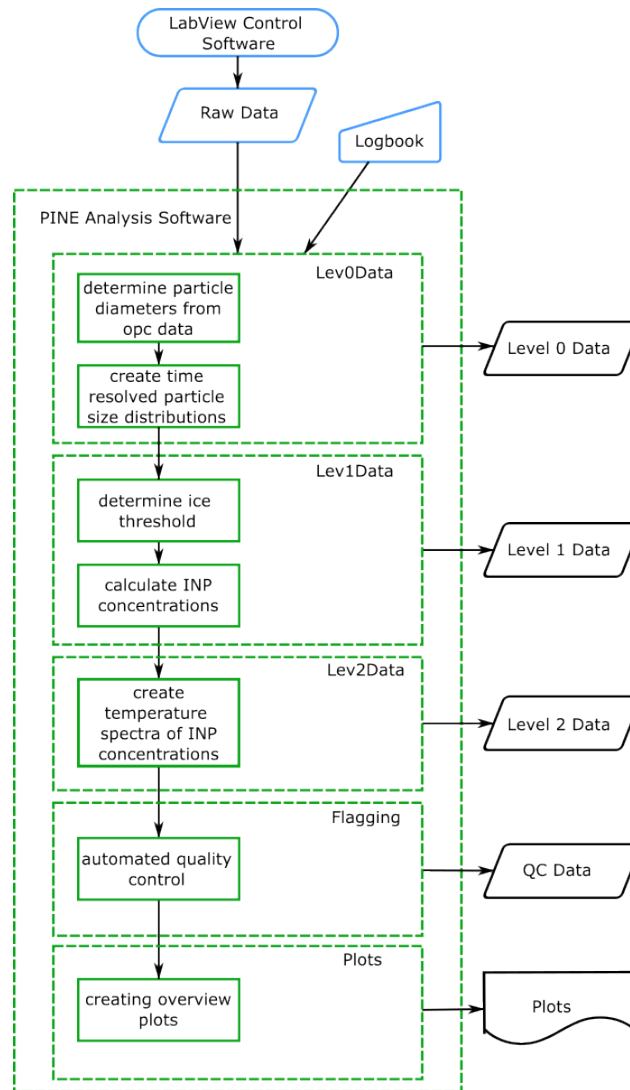


Figure 3. Dataflow of Software

4 PIA Software

150 4.1 Structure and Data Flow

The PINE INP Analysis (PIA) software is a tool written in Python, that allows for standardised analysis of PINE measurements. This work describes the structure and functionality of PIA v3.0.0. It consists of five sections, represented as classes (see Fig. 3) and are based on a data structure with processing levels 0 - 2. The first class deals with Level 0 data, stored in *level0_data.py*. The second and third classes are processing Level 1 and Level 2 data, respectively, and are stored in the *level_data.py* file. After

155 that, the data undergo automated quality control tests, and finally, overview and analysis plots are created. The respective python files are *flagging.py* and *plotting.py*. There is one additional class which is called *ExtractData* and stored in the *extract_data.py* file. It contains functions to collect raw data and processed data files. The software analyses one operation at a time. The software tool itself is executed by using the *main.py* file. In the *settings_<pine-id>_<campaign>.toml* file, data paths and metadata have to be specified.

160 In the following paragraphs, the five processing parts are described in detail.

Lev0Data:

In this class the raw OPC data are mapped to the respective particle diameters given by the calibration tables and saved as Level 0 data. It is important to note that these diameters do not necessarily represent the real diameters of the particles. As described in Sec. 2.1, the PM voltage of the OPC is adjusted to see cloud droplets and ice crystals, therefore the representation of the calibration table is not always accurate. As the measured variable of the PINE instrument is the INP concentration, this does not change anything in the results. The particle diameters are then distributed into time and size bins to allow for easier handling. The default value for the time bins is 3 seconds, if another time resolution is required this can be changed in the *settings_analysis_flagging.toml* file. The size bins follow a logarithmically equidistant distribution between 0.1 and 1000 μm divided into 100 bins.

Lev1Data:

The Level-1-class determines the diameter, above which detected particles are considered to be ice crystals (denoted as ice threshold, see Sect. 4.2). Afterwards, the particle concentration n_p is calculated as described in Möhler et al. (2021) via

$$n_p = \frac{c_p}{F}, \quad (2)$$

with the particle count rate c_p and the mass flow rate F through the OPC. The INP concentration can be calculated with the same equation (Eq. 2) but using only the particles above the ice threshold. This is done for each time bin and for the whole expansion.

Lev2Data:

The Level-2-class creates freezing spectra of the INP concentration, i.e. the INP concentration as a function of their freezing temperature, with a resolution defined as *del_temp* for each expansion. The default temperature bin width is 0.5 K.

Flagging:

This part of the code performs the automated quality control of the data. It creates log-files with errors and warnings for every operation. The applied quality checks are listed in Sec. 4.3.

Plots:

The plotting routine is separable from the rest of the code and is independent from the other classes. Three functions for plotting are integrated.

1. Time series: Overview plot for one operation including temperatures, pressure, particle size distribution and INP concentration. Data that are associated with an *error* or *warning* flag are marked in the time series plot.
2. Diagnostics: Creates comparison plots with theoretical dry adiabatic temperatures and saturation ratio for one operation.

3. Histogram: For specified runs, a histogram of the size distribution including the ice threshold is plotted.

190 **Optional parameters:**

The *main.py* file has two optional parameters. The first one is the "manual_ice_threshold". If it is set to True, the ice threshold will not be calculated, but taken from a file called *<pine-id>_<campaign>_op_id_<operation-id>_ice_threshold_manually.txt*. An example file is shown in the *files*-folder of the software, which also contains the conversion table (*bins_ice_threshold.txt*) from diameter to bin number for the ice threshold. The manual ice threshold files need to be stored in the

195 */L1_Data/exportdata/ice_threshold/*-folder. The second parameter is "cirrus_mode". If this is set to True, the software will not calculate the ice threshold, but the total particle concentration during expansion subtracted by the particle concentration during flush is used as ice concentration. However, measurements at cirrus conditions are not fully tested and applicable for the standard PINE instrument. This feature is mainly implemented for the PINEair prototype (Bogert, 2024).

200 The output of the analysis software is either stored inside the same folder as the raw data or, if specified in the *settings_<pine-id>_<campaign>.toml* file, at a user-defined path. Table A1 in the appendix summarises the content of the files. All output files start with the prefix *<pine-name>_<campaign>_op_id_<op-id>_*.

4.2 Ice Threshold Finder

4.2.1 Approach

205 The ice threshold diameter (in the following referred to as the ice threshold) is defined as the particle diameter that distinguishes between ice crystals versus cloud droplets, and aerosol particles, whereby the former are optically larger than the latter. The optical discrimination of particles is based on the size-dependent scattering behaviour of particles, as measured by the OPC. It is of paramount importance to determine the ice threshold accurately in order to calculate the INP number concentration from the aforementioned measurements. Two principal methods may be employed to determine this threshold diameter. A)

210 From the time series, or B) from the discrete and typically bimodal particle size distribution. On the timescale of PINE expansions, cloud droplets grow by diffusion. The maximum size of cloud droplets, and thus the ice threshold, is impacted by the experimental conditions regarding temperature and supersaturation, as well as the concentration of aerosol particles that serve as cloud condensation nuclei. Due to this, it is preferable to determine this threshold on a run-by-run basis rather than on an operation-wide scale. In addition to the potential for bias at the individual level, the sheer number of expansions renders

215 a manual threshold determination impractical. To address this challenge, a tool has been developed for the PIA software that automatically, objectively, and reproducibly identifies the ice threshold. The particle number size distribution (method B) was selected as the basis for determining the ice threshold, which corresponds to the minimum between the two modes. Three options for finding this minimum will be discussed in the following sections.

1. Fitting:

220 The objective is to fit a discrete distribution in order to obtain a continuous probability distribution function (PDF). The automatic fitting of such a function is not a straightforward process. It typically requires either machine learning

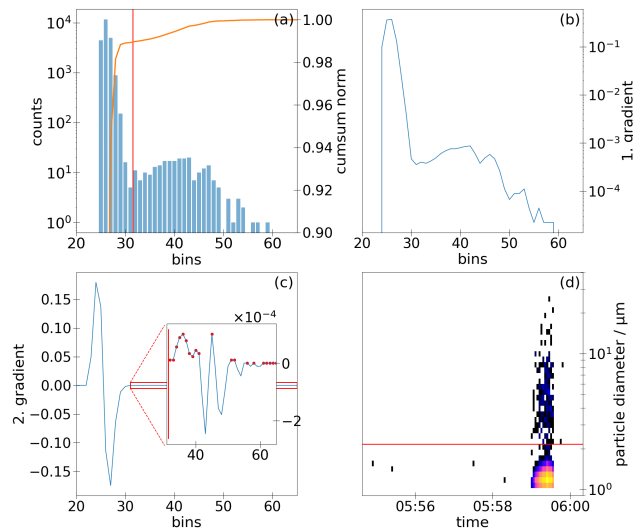


Figure 4. Overview plots of the ice threshold finder algorithm. Panel (a) shows the single particle size distribution of one run (in counts on left y-axis) and, in orange, the accumulated and normalised sum of all detected particles (cumsum). Panels (b) and (c) show the first and second gradient of the accumulated sum, respectively. The zoomed-in window in panel (c) shows the part of the second gradient when it becomes positive after the minimum. The red dots mark the values ≥ 0 . Panel (d) shows the time series of the size distributions. The red lines in (a), (c) and (d) indicate the ice threshold diameter. The graph was created using the data of the CORONA campaign, operation 61, run 126.

algorithms or manual adjustments, and it introduces additional uncertainties into the subsequent step of calculating the minimum by differentiating the PDF. This is particularly at odds with the concept of an automated software solution.

2. Discrete calculus:

In the context of discrete distributions, the analogue of the derivative function is the forward differences method. This involves calculating the difference $\Delta f(n)$ between the next and the current bin n , as follows:

$$\Delta f(n) = f(n+1) - f(n). \quad (3)$$

A change in sign, from negative to positive, indicates the occurrence of a local minimum. However, this is also the method's most significant drawback. This method will yield multiple local minima, but no single global minimum. This is also due to two factors:

- (a) The OPC size detection range is much larger than the typical size range of aerosols and cloud particles.
- (b) The typically few ice particles are spread over a wide size range, resulting in the creation of empty bins.

One potential solution would be to select a smaller size range for the ice threshold finder, where the ice threshold diameter is typically identified. However, the diameter is contingent upon the experimental conditions (temperature, aerosol type,

235 etc.), rendering the method non-scalable. Moreover, the question of how to distinguish between artificially empty bins in the ice particle mode and the minimum remains unresolved.

3. Cumulative distribution function (CDF):

For each probability distribution, whether discrete or continuous, a continuous monotonic increasing function can be determined, known as the CDF. In the case of a discrete distribution, the CDF is the bin-wise sum of the distribution, with
240 each bin normalised to a sum of one. The crucial advantage of this approach is its monotone increasing character, which effectively addresses the issue of multiple minima present in option 2. In the case of a bimodal log-normal distribution, the CDF would exhibit a single distinct inflection point, which would directly define the threshold.

Consequently, the automated ice threshold diameter finder tool was developed on the basis of option 3.

4.2.2 Algorithm

245 Based on the aforementioned discussion, the ice threshold finder tool employs the following algorithm:

1. The CDF is calculated by accumulating all size bins over all time bins during flush and expansion mode and normalising it (Fig. 4a orange line), based on the particle number size distribution created in ExtractData for the selected run.
2. From this accumulated and normalised distribution, the first and second gradients (based on the central differences method) are calculated, providing an equivalent to the first and second derivatives of a continuous function (Fig. 4b and
250 4c).
3. In theoretical terms, the inflection point is defined as the point at which the curvature undergoes a change in sign. Consequently, the second gradient curve crosses the zero line. The algorithm is designed to identify the bins (i.e., the bins above the first minimum in the second gradient CDF) where the second gradient CDF becomes greater than or equal to zero (indicated by the red dots in Fig. 4c).
- 255 4. The ice threshold diameter bin is set to the first bin for which the second gradient CDF becomes greater than or equal to zero (Fig. 4d), except in the following cases:
 - No distribution was found.
 - No bins for which the second gradient CDF becomes greater than or equal to zero were identified.
 - The bin of the maximum of the second gradient CDF is larger than the very first bin for which the second gradient
260 CDF becomes greater than or equal to zero (e.g., due to a tri-modal distribution).

In such cases, the ice threshold bin is set to its default value of 0 (ice threshold = 0.1 μm).

It should be noted that this algorithm is only applicable to mixed-phase clouds. However, if the liquid cloud is not visible in the detection area (e.g., due to an incorrect PM voltage at the OPC), the threshold value cannot be determined. In cases of pure ice clouds (e.g., homogeneous freezing), the algorithm may identify a valid but erroneous ice threshold.

To estimate the uncertainties introduced by the ice threshold finder, we compared the INP concentrations calculated from manually setting the ice threshold and from the automatically determined ice threshold. The manually set ice threshold is represented by one value for one operation if the temperature is constant. For temperature scans the ice threshold varies slightly within one operation due to the change in temperature. We therefore randomly selected operations conducted at constant temperature and temperature scans.

For comparison, we also included some operations from a laboratory campaign that was conducted under controlled conditions regarding the sampled aerosol population with the model PINE model 05-02 (PINE-05-02) and also operations from a field campaign called CountIce (Tarn and Murray, 2025) that was conducted by the University of Leeds with their PINE model 04-03 (PINE-04-03), which is from the same production series as the PINE-04-02. Measurements of a laboratory campaign differ from those of a field campaign in the sense that the aerosol type and concentration are controlled in the laboratory. Therefore, much more particles and INPs can be involved. This can be reflected in the bimodal particle size distribution and must be considered when verifying the ice threshold finder.

The CountIce campaign took place from August 2021 to February 2022 and from March 2023 to August 2023 at the University of Leeds campus (53°48'18.1"N 1°33'10.5"W, 66 m above sea level). These measurements are included for comparison to another PINE model and a different operator setting the manual ice threshold. The comparison was performed for 12 operations from CORONA, 4 operations of the laboratory campaign and 9 operations from CountIce, resulting in a total of more than 4000 analysed runs. For CountIce data above 273 K was neglected.

Figure 5 shows an example of a CORONA operation for which the comparison was carried out. The operation was conducted with a constant wall temperature of about 257 K. It can be seen that the INP concentrations are generally underestimated when using the ice threshold finder. However, there are also some runs for which the INP concentration is overestimated when using the manual threshold (runs at about 15:00).

To quantitatively investigate the uncertainty of the INP concentration, the ratio between the INP concentration calculated from the automatically determined ice threshold and the INP concentration calculated from the manual ice threshold was calculated. This ratio is close to 1 for most of the runs ($\approx 60\%$ have a ratio between 0.98 and 1.02). About 25 % of the runs have a ratio above 1.02, and 15 % fall below a ratio of 0.98.

Figure 6 shows that there is a strong dependency on the total INP concentration with a higher bias at lower INP concentrations. In such cases, when there are only a few ice crystals, the margin of error is larger, as even a slight shift in the ice threshold could result in these low numbers not being counted. It is noticeable that those concentrations derived from the ice threshold finder that are higher than the ones derived from the manual ice threshold are mainly from the CountIce campaign. This difference highlights the need of an objective, operator-independent tool, as the manual ice threshold was set by different responsible persons for each campaign.

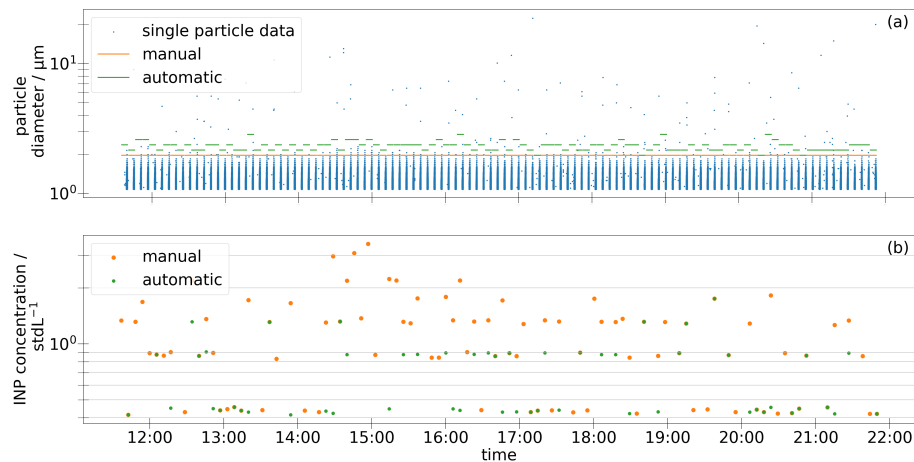


Figure 5. Comparison between the manually set ice threshold and the ice threshold determined by the ice threshold finder for operation 140 of the CORONA campaign. The operation was conducted at a constant temperature of about 257 K. Panel (a) shows the single particle data measured by the OPC, the horizontal lines indicate the diameter which represents the ice threshold. Panel (b) shows the INP concentration per run.

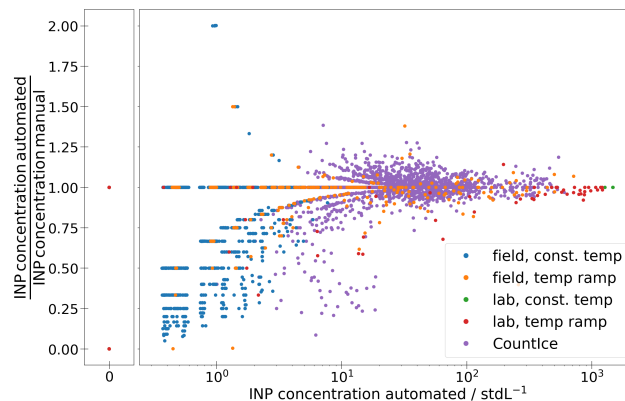


Figure 6. Ratio between the INP concentration calculated from the ice threshold determined by the ice threshold finder and the INP concentration calculated from the manually set ice threshold plotted against the total INP concentration derived from the ice threshold finder. The colours show whether the data are taken from a lab or a field study and if the PINE was measuring at constant temperature or performing a temperature scan. Runs with an INP concentration of zero for both manually and automatically determined ice thresholds are displayed with a ratio of one.

4.3 Quality Control

All raw and processed data undergo an automated quality control (QC). For the quality control the Python tool System for automated Quality Control (SaQC; Schmidt et al., 2023) is used. SaQC is a tool for automated quality control of time series

300 data that was developed within the Helmholtz Earth and Environment community. The tool has predefined flagging schemes, but also allows users to design a scheme tailored to their own needs.

In case of the PIA software, a scheme returning 3-digit codes for the analysed data was created. Within the software, these codes are mapped to a message containing the run and information about the failed test. The messages including the run numbers are saved in a log-file.

305 The flags are reported as failure levels *info*, *warning*, and *error*. The flags, messages, and the respective failure levels are defined in the *settings_analysis_flagging.toml* file. It should be noted, that no data is removed during the quality control. It lies within the responsibilities of the data user to handle the flagged data. If any data points within a run are flagged with an *error* the run should be excluded from further analysis. If the data are flagged with *info*, a threshold is set in the settings file. This threshold is indicated in the following test descriptions inside brackets behind the flag. If the number of flagged data points
310 within one run exceed the threshold, the data points are flagged as *warning*.

Data with *warning* and *info* flags should be inspected manually, as it may indicate an instrumental malfunction. Therefore, a manual data reviewing and flagging tool is currently in development. This will help the PINE users to make their data fully quality controlled.

The tests are split into two categories: tests for single data points and tests for a whole run. The latter category describes tests
315 on parameters that influence the result of the whole run: e.g., if the ice threshold is not set correctly, the whole run is flagged accordingly. In contrast, an outlier in one of the instrument sensors does not necessarily invalidate the entire run. The zero test on instrument data is performed on the raw data with a 1 second frequency, all other tests for data points are performed on the 3 second averaged data. The detailed information of the test results can be found under */Quality_Control/quality_flags/* in the *flags_instrument.txt* file for the tests per data point and in the *flags_metadata.txt* file for the run-wise tests. Table A2 summarises
320 the QC tests including the internally used flag code, the error type, which test category they belong to, if possible the parameters used for the test, and the short description used in the file headers. In the following all tests, where the parameters could not easily be summarised in the table, or require additional explanation are listed.

Unless stated otherwise, the test conditions are based on empirical observations and have been adapted in an iterative process to get stable test results or are based on instrument specifications.

325 **4.3.1 Zero test on instrument data**

The raw data from the instrument file are checked for zero values. If all parameters (temperature, pressure, flow, dew point) have zero values simultaneously, a connection error has occurred. If this is the case, the zero values are replaced with NaNs and handled as missing values for the further analysis.

4.3.2 Flatline test on instrument data

330 This test looks for sensor data, which don't change their value over a certain amount of data point. This could be a result of incomplete data retrieval. For the gas temperature only the data during expansion is tested and therefore five consecutive data points with the same value are already enough to detect a malfunction of the sensors. For pressure and dew point the

test is performed for whole time span und therefore checking for 10 constant data points. An example of flatline behaviour in temperature data is shown in Fig. A2.

335 4.3.3 Outlier test on flow

The flow is not constant over all run modes, therefore this test first filters the data for the relevant run modes (for Fe expansion, for Fm flush and expansion) and performs a Z-test on the flow rate. If the first or last data points of a run mode are flagged, they are ignored, as the flow typically stabilises after a few seconds. Other outliers are flagged as *warning*.

4.3.4 Particles during refill

340 During refill, no flow is actively passing the OPC, so no particles should be detected. The first two time bins of the refill mode are ignored, as it is possible that there are still particles in the line after the flow through the OPC is turned off. If particles are detected in any other bins, the data are flagged as *warning*. This could be a sign for a contaminated OPC or icing inside the chamber. Figure A3 shows an example run where particles were detected during refill.

4.3.5 Test on run times

345 This test checks that the run times (start, expansion, refill, end) are valid and strictly increasing. If this test fails, the run is flagged with *error*.

4.3.6 Variability of ice threshold

For this test, the mean and standard deviation of all ice thresholds in one operation are calculated. If a single ice threshold deviates more than two standard deviations from the mean, the run is flagged with a *warning*.

350 4.3.7 Large aerosol particles during flush mode

If, during flush mode, more particles above the ice threshold than 50 % of the INP concentration during the expansion are detected, the run is flagged with *warning*. This indicates a high ice background or other contamination inside the chamber.

4.3.8 Temperatures at the end of expansion

The temperatures sensors inside the chamber should have the following order at the end of the expansion: $Ti1 > Ti2 > Ti3 > Ti4 > Ti5$. If this is not the case for the temperature sensors $Ti1$ to $Ti4$, the data are flagged with *warning*. If $Ti5$ is higher than $Ti4$, it results in an *error*, as $Ti5$ is used as the reference temperature for the INP concentration. If this is the case it may be due to a malfunction of the temperature sensors or a wrong positioning of the sensors inside the chamber. Figure A4 shows an example of expansions where $Ti5$ measures a higher value than $Ti4$.

4.3.9 Global range test for duration of expansion

360 The duration of the expansions must lie within the following ranges:

- Ti5 below -35°C : 9 sec to 2 min
- Ti5 above -35°C : 20 sec to 2 min

Below -35°C PINEair can measure homogeneous freezing by conducting a rapid expansion. In this case 9 sec are enough time to detect the ice crystals with the OPC. For heterogeneous freezing it takes a longer time for the ice crystals to form and to be
365 guided through the OPC. If this test fails the run is flagged with *error*.

4.3.10 Check for supersaturation

If water supersaturated conditions are not given inside the chamber, a liquid cloud cannot form and the ice threshold finder would not function efficiently. If Ti5 is higher than DP, supersaturation is likely not achieved, and the run is flagged. As the DP sensor is positioned in front of the chamber, it is still possible to have supersaturation, even if Ti5 is higher than DP, namely
370 when an ice layer formed on the chamber wall after a longer time of operation. Thus this check is only flagged as *warning*.

4.3.11 Icing of chamber inlet

If the sampled air is highly humid, ice may form on the chamber inlet. This reduces the flow rate into the chamber and, consequently, the pressure inside the chamber during flush. The pressure sensor is positioned outside the chamber, in front of the iced inlet, so it does not detect this reduced pressure. Once the valve closes, the pressure equilibrates, which is reflected
375 by a faster-than-expected pressure drop at the start of expansion. To detect such cases, the theoretical expansion duration is calculated from the initial pressure at the beginning of the expansion P_0 , the pre-set end pressure P_{end} , the pre-set expansion flow Fe_{set} , and the volume of the chamber V_{ch} :

$$\Delta t_{\text{exp}} = \frac{\frac{P_0}{P_{\text{end}}} \cdot V_{\text{ch}} - V_{\text{ch}}}{Fe_{\text{set}}} \cdot 60. \quad (4)$$

As the expansion flow is given in L min^{-1} , the whole term needs to be multiplied by 60 to get the result in seconds. If the actual
380 expansion duration is shorter than the theoretical value, the run is flagged with *error*. This test only works if the inlet is already heavily iced. Runs that are only slightly affected by the iced inlet are not detected. To ensure no faulty runs are missed, all runs following three consecutively flagged runs, as well as the five preceding them, are also flagged. If five or fewer unflagged runs remain in an operation, the entire operation is flagged. Figure A5 shows an example of pressure measurements during inlet icing.

385 4.3.12 Gaps in OPC data

This tests checks for missing data in the OPC data during expansion. This is first applied to the 3 second binned data. If a bin without OPC data is found, the test is repeated with 1 second bins. This test is only performed if the total particle concentration

of the run exceeds 180 stdL^{-1} . If the particle concentration is lower, gaps seen in the OPC data are most likely real gaps. If this test fails, the run is flagged with *error*.

390 4.4 Maintenance of Software

The PIA software is managed in a publicly available GitLab repository¹ containing multiple branches. The main branch is used for releases and the development is done in the development branch, or in branches dedicated to a specific task. The issues-system is used to mark needed improvements, bug reports, or new feature requests. The issues can be filed from code developers or PINE users. The repository has a Continuous Integration and Continuous Delivery (CI/CD) pipeline implemented, which
395 runs multiple tests for every push event to the repository. These tests include: linting tests, check that all licensing information is conform with the REUSE (Free Software Foundation Europe, 2025) recommendations, verifying that the software is running for different PINE and Python versions. For the latter, test data from different PINE and LabVIEW control software versions are included in the repository to ensure their compatibility with new PIA software or Python versions. Since PIA version 3.1.0, the CI/CD pipeline includes a workflow for automated releases on Zenodo, based on the FACILE-RS tool (openCARP consortium
400 et al., 2025). This workflow is triggered by creating a *pre*-tag on the main branch. It updates the metadata, the changelog-file, and the version number. Afterwards, the real version tag and a release on GitLab is created, followed by pushing the code and the metadata to Zenodo, where it can be released after manual inspection. The documentation for the PIA software² is hosted on Read the Docs (Read the Docs, Inc & contributors, 2025). The GitLab repository is connected to Read the Docs via a webhook, and whenever a new tag in the repository is created, the documentation will be updated.
405 With regard to the above mentioned workflow, the PIA software follows the FAIR4RS (FAIR for Research Software; Barker et al., 2022) principles. The fulfilment of these principles is summarised with the FAIR software checklist³.

5 Quality Assurance of PINE Measurements

5.1 Characterisation of Temperature Sensors

The gas temperature inside the chamber is influenced by the ambient air entering from the top. Therefore, a temperature gradient
410 exists, depending on the temperature of the incoming air and the wall temperature, with the highest temperatures at the top and the lowest temperatures at the bottom of the chamber. This general temperature gradient is also observed during expansion: while the gas temperature decreases at all sensor locations, the lowest temperature is consistently recorded by the sensor at the lowest position (Ti5). Evaluation of the behaviour of the temperature sensors can help to validate the functionality of the sensors and identify possible malfunctions. To verify the stability and comparability of this behaviour over time and across different
415 instruments, the temperature differences between neighbouring sensors were analysed for three different PINE models and 22 measurement campaigns. These differences were inspected for correlations with the pressure change during expansion and

¹https://codebase.helmholtz.cloud/pine/pia_software

²<https://pia.readthedocs.io/en/stable/>

³<https://fairsoftwarechecklist.net/v0.2?f=31&a=32113&i=22101&r=123>

the end temperature of Ti5 during expansion. The run-wise data did not show any significant correlations. Nevertheless, it is interesting to compare the different PINE models in order to identify any systematic deviations in the reported temperatures. Figure 7 shows the temperature difference between sensors Ti3 and Ti4, and between sensors Ti4 and Ti5 as a function of the temperature measured at the end of the expansion. Ti1 and Ti2 usually show high fluctuations, which can be expected as they are the those mostly impacted by the entering ambient air, and are therefore not further analysed. The data is temperature binned per PINE model and the markers show the mean values, while the error bars show the standard deviations. The data of the PINE-05-02 is split into two, because the temperature sensors had to be repositioned by the manufacturer, as Ti5 was positioned to close to the wall and therefore reporting higher values than Ti4 due to the wall temperature. The values from before this repositioning are marked in red, the values marked in green are results from afterwards. While the differences between Ti3 and Ti4, and Ti4 and Ti5 agree well for the PINE models 04-01 and 04-02, the results for PINE-05-02 (after repositioning) show a different behaviour. The difference between Ti4 and Ti5 is about 1 K higher than for the other instruments and the difference between Ti3 and Ti4 is slightly lower than for the other instruments. This could indicate that the sensors Ti4 and Ti5 are at different positions in PINE-05-02 compared to the positions in the other PINE models. It is possible that the position of the Ti5 sensor is located in a colder region of the chamber compared to the Ti5 positions in other instruments and that the Ti4 sensor is positioned in a warmer region.

In the future, a better control and documentation of the sensor positioning is required. This would enable the use of temperature differences between sensors as a quality assurance measure. This test would be particularly useful for the initial validation of new instruments.

5.2 Background Measurements

Ice can form on the chamber walls during continuous operation of PINE, if the air inside the chamber is too humid. To ensure that this does not influence the INP concentration, regular ice background measurements must be performed. Therefore, ambient air is guided through a HEPA filter to ensure that no external particles enter the chamber, meaning that any detected cloud droplets or ice crystals are only due to particles already inside the chamber. These operations must be marked as background measurements in the electronic logbook.

Analysis of background measurements from existing campaigns shows that it takes approximately 80 to 100 L to completely flush out the aerosol inside the chamber. This includes both the air that is guided through the chamber during flush and the air filled into the chamber during refill.

We recommend performing background measurements once per day using the following settings (or similar, according to the settings during measurement operations) to achieve an air exchange of at least 80 L: 9 runs with 5 min flush time at a mass flow rate of 2 stdL min^{-1} , and a pressure reduction during expansion to approximately 15 % below ambient conditions.

If the ninth run still shows significant particle counts, the chamber is likely iced or otherwise contaminated. Ice-sized particles should not be visible after the sixth run.

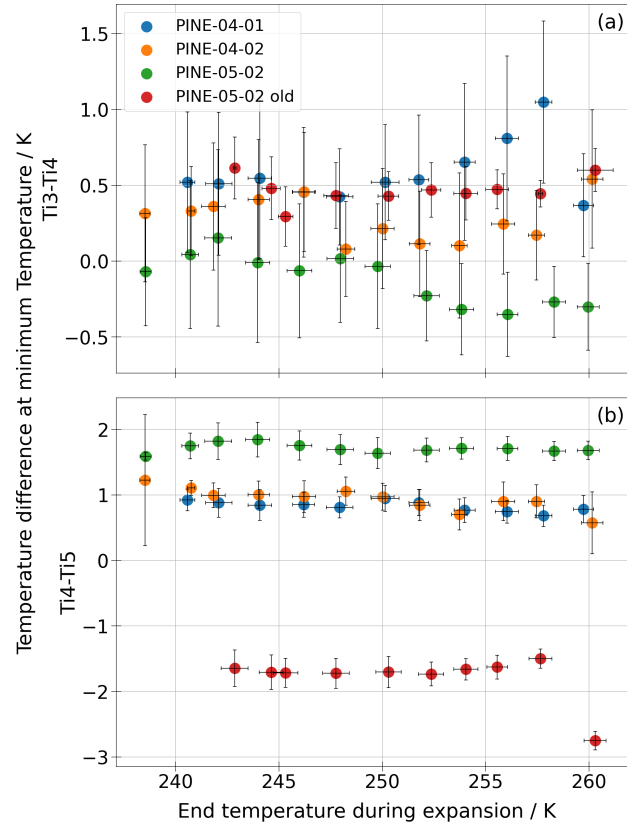


Figure 7. Difference between two neighbouring temperature sensors as a function of the temperature measured at the end of the expansion. The values are temperature binned and the error bars show the standard deviations of the x- and y- values. Different colours indicate different instruments.

5.3 Influence from Large Aerosol Particles

450 Aerosols with diameters larger than the ice threshold are occasionally observed during flush mode. For the CORONA campaign, this occurred in more than 30 % of the runs. If such large aerosols are present during flush, it is possible that they are also present during expansion and contribute falsely to the INP concentration. Figure 8 shows the concentration of ice-sized aerosols during flush as a function of the INP concentration during expansion. There is only a slight dependency ($r = 0.273$) between large aerosols during flush and the INP concentration during expansion, therefore the relative influence of large aerosols becomes

455 relevant for concentrations below 10 stdL^{-1} . This influence may in fact be smaller, as the size of aspherical particles appears larger to the OPC. If water condenses on these particles, they become spherical droplets and appear optically smaller. More systematic measurements need to be performed to understand this potential influence from large aerosol particles. For now, large aerosol particles should be accounted for by including the ice-sized aerosol concentration during flush in the measurement

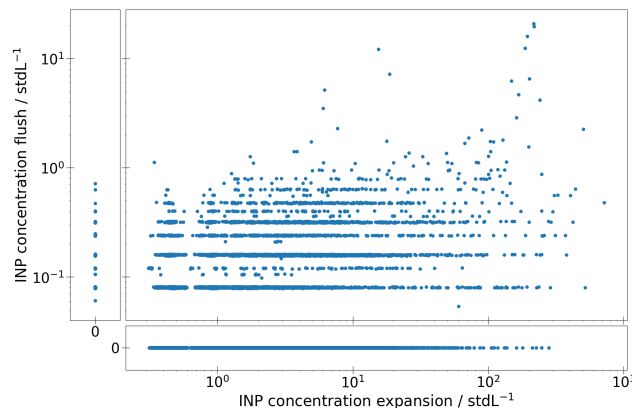


Figure 8. Ice-sized aerosol concentration during flush as a function of the INP concentration during expansion.

uncertainties, if they exceed the statistical uncertainty. These values are also stored in the *ice.txt*-files generated by the PIA
 460 software.

6 Community Aspects

6.1 Aerosol, Clouds, and Trace Gases Research Infrastructure (ACTRIS)

The ACTRIS (Laj et al., 2024) Topical Centre CIS is one of the six Topical Centres, supporting the observational platforms in
 producing high-quality data and information on the abundance, variability, and trends of short-lived atmospheric constituents
 465 and related processes. Among the variables for CIS are the concentration and freezing spectrum of INPs. CIS provides mea-
 surement guidelines for operating the online instruments PINE and the automated CFDC Horizontal Ice Nucleation Chamber
 (HINC-Auto; Brunner and Kanji, 2021), as well as the offline INP measurement method INSEKT. The PINE instrument serves
 for laboratory-based validation and for field-based intercomparisons, aiming to achieve accurate and quality-controlled long-
 term INP data at ACTRIS observational platforms, atmospheric simulation chambers, and mobile platforms.

470 6.2 PINE Community Portal

A growing worldwide community of PINE users will also be connected to the future INP monitoring network and CIS activities
 in Europe. To strengthen the INP community and enhance collaborations, we developed the PINE community portal⁴. The
 portal is publicly accessible and maintained by KIT. Any user is able to submit additional datasets to the portal. Existing data
 can be viewed online and up to three measurement sites can be compared across different nucleation temperatures.

⁴<https://pcp.imk.kit.edu/pine-dashboard>

475 6.3 Application for other PINE Versions

Since PIA software version 3.0.0, the compatibility with the further developed prototype PINEair and the laboratory-based instrument AIDAm (Vogel et al., 2022) is given. Thanks to its well-defined data structure, comprehensive data acquisition software and modular architecture, the PIA software can easily be adapted for future PINE versions.

7 Conclusions

480 A standardised workflow for handling data from the PINE instrument was established. Therefore, the Python software package PIA (PINE INP Analysis) was developed and is presented here. The software analyses the raw data acquired by an in-house developed LabVIEW control software and converts it into time series of INP concentrations and freezing spectra. The required ice threshold for distinguishing between aerosols and droplets versus ice crystals is automatically determined by the ice threshold finder described in Sect. 4.2. This allows for a standardised data set that is independent of the data analyst. All data undergoes
485 several automated quality control tests, which are described in detail in Sect. 4.3. These tests are continuously reviewed by the PINE users and will be further developed as needed. The same applies for the ice threshold finder. As the number of data sets increases, statistical robustness will improve, allowing the current algorithm to be verified or further improved. The modular structure of the software allows for easy adaption to future PINE versions.

For further improving the software, the possibility to manually flag the data will be implemented in the future. It is also desirable that the LabVIEW control software automatically generates the metadata currently provided by the manual electronic
490 logbook (e.g., sample aerosol, operation type) and stores it in the raw data to ensure compatibility with the PIA software and further increase the automatisisation of the data workflow.

Section 5 provides recommendations on how to treat the data and conduct measurements to ensure reliable and standardised results. The offset between temperature sensors should be checked as part of the initial validation and possibly after longer
495 transport of the instrument, as an indicator of sensor positioning inside the chamber. Background measurements should be performed regularly (approximately once per day) to check for icing and contamination inside the chamber. The presence of large aerosol particles during flush mode should be considered when calculating measurement uncertainties, if they exceed the statistical uncertainty given by Eq. 1.

Code and data availability. The PIA software is available on Zenodo (<https://doi.org/10.5281/zenodo.15592883>).

500 The CORONA data set is available on RADAR4KIT (<https://dx.doi.org/10.35097/sqmdyj7ckbccq9zy>).

The CORONA_new data set is available on RADAR4KIT (<https://doi.org/10.35097/c78mxhyjyd269pr9>).

The CountIce data set is available on zenodo (<https://doi.org/10.5281/zenodo.17451019>). For this work the data sets 211015, 211110, 211122, 211214, 220110, 230316, 230414, 230524, and 230609 have been used.

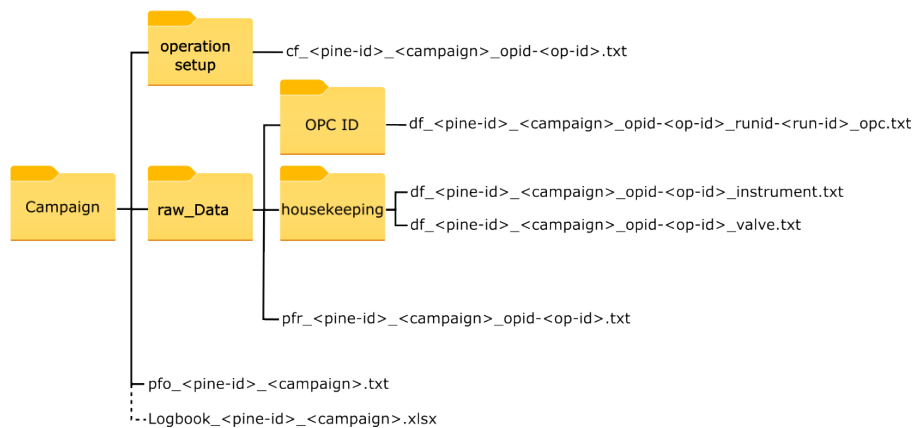


Figure A1. Raw data structure as produced by the LabVIEW control software. The dashed line indicates the location of the manual electronic logbook, which is not created by the LabVIEW control software.

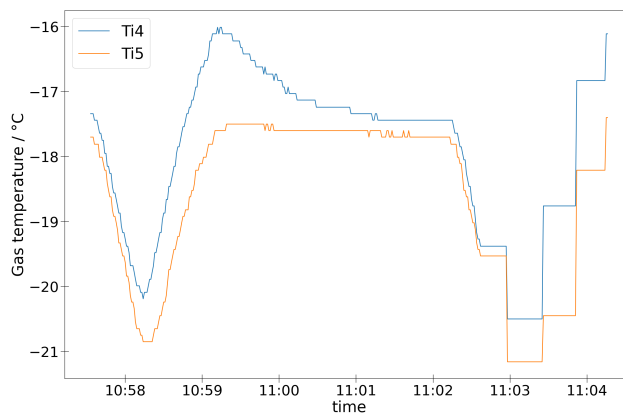


Figure A2. Example of flatlines in the instrument data. The plot shows operation 76, run 16 from the CORONA_new campaign.

Appendix A

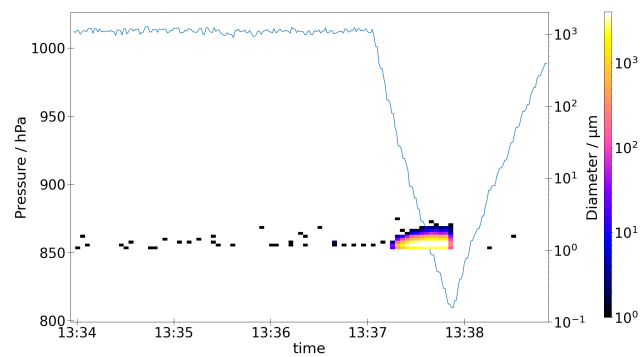


Figure A3. Example of particles detected during refill mode. The plot shows operation 28, run 61 from the CORONA_new campaign.

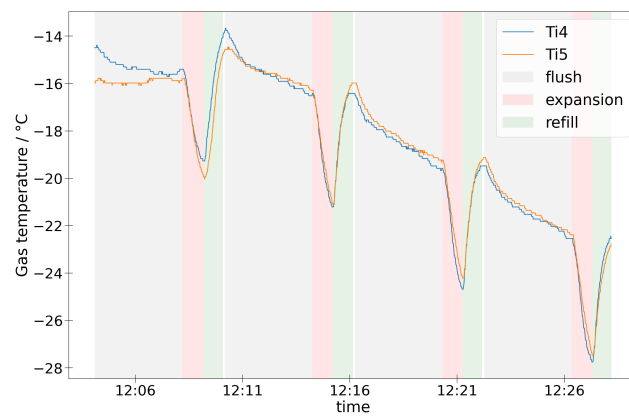


Figure A4. Example of Ti4 measuring a lower temperature than Ti5. The plot shows a temperature scan of operation 205, runs 4 to 7 from the CORONA campaign.

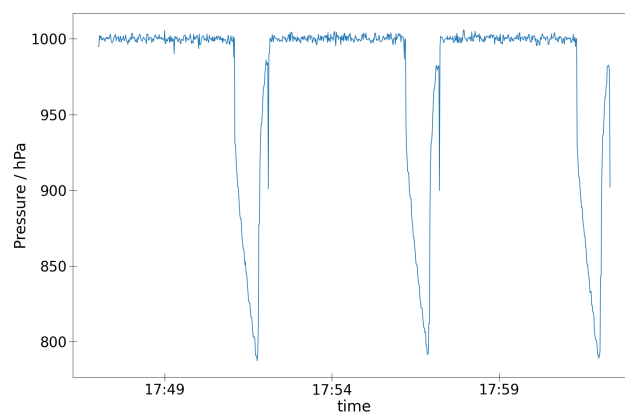


Figure A5. Example of detected inlet icing. At the start of the expansion and the end of refill a steep drop can be seen in the pressure. This is due to the reduced pressure inside the chamber resulting from the iced inlet. The pressure drop results from the equilibrating gas once the valve switches. The plot shows operation 198, runs 10 to 12 from the CORONA campaign.

Table A1. Output files

file ending	location	content	metadata
opc_spd.txt	/L0_Data/exportdata/ exportdata_opc_spd/ OP<op-id>/	The single particle data from the raw OPC data with an additional column containing the diameter representing the bin number.	software version, description of variables
cn.txt	/L1_Data/exportdata/ exportdata_cn/ OP<op-id>/	3 second binned data: time relative to start of run, particle concentration, particle count, INP concentration, INP count, run mode: flush (1), expansion (2), refill (3), mean flow rate, mean chamber pressure, mean Ti1, mean Ti2, mean Ti3, mean Ti4, mean Ti5, particle counts per bin size	software version, station, latitude, longitude, altitude, PI, institution, aerosol, operation type, OPC number, start time, flush duration, size of time bins, ice threshold, description of variables
ice.txt	/L1_Data/exportdata/ exportdata_ice/	run-wise data: run ID, time at the middle of the flush, minimal temperature (Ti5) reached in expansion, pressure at the end of expansion, mean flow rate during expansion, INP concentration during expansion, INP concentration during flush, particle concentration during expansion, particle concentration during flush	software version, station, latitude, longitude, altitude, PI, institution, aerosol, operation type, OPC number, description of variables
ice_threshold.txt	/L1_Data/exportdata/ ice_threshold/	run-wise data: run ID, bin number, ice threshold diameter	software version
temp.txt	/L2_Data/Temp_Spec/	temperature-binned data: start temperature, end temperature, run ID, time at the middle of the flush, mean time relative to start of expansion, mean INP concentration, lowest chamber pressure	software version, station, latitude, longitude, altitude, PI, institution, aerosol, operation type, OPC number, description of variables
temp_mean.txt		operation mean of temperature-binned data: start temperature, end temperature, mean INP concentration in temperature bin over all runs, standard deviation of INP concentration	software version, station, latitude, longitude, altitude, PI, institution, aerosol, operation type, OPC number
flags_instrument.txt	/Quality_Control/ quality_flags/	flags for data points, for internal use of the software	software version
flags_metadata.txt		flags for tests per run, for internal use of the software	software version
flags.log	/Quality_Control/	Log file with warnings and errors for all runs	

Author contributions. NB prepared the manuscript with contributions from RF, OM, LL and FV. The software was designed and written by NB and RF. LL, FV, and MT reviewed the data for the manual ice threshold analysis. AB, LL, FV, PB, and JN contributed to the design of the quality control tests and the output data. FV, LL, OM, MT, and BM were responsible for collecting and processing the observational data.

Competing interests. OM, LL, and BM were involved in a prior technology transfer project between KIT, University of Leeds, and Bilfinger
510 Nuclear & Energy Transition GmbH (Würzburg, Germany) related to the development and commercialization of the underlying PINE instrument. The software described in this work is independent of this project, open source and freely available. The authors declare no competing financial interests.

Acknowledgements. AI has been used to check the manuscript for grammatical errors and improve its formulation.
CountIce was supported by the European Research Council Proof of Concept grant (862565). The development of the PINE instrument was
515 supported by the Karlsruhe Institute of Technology through the technology transfer project "A Portable expansion chamber for Ice Nucleating particle measurements - PINE (project number N059)" and by the University of Leeds International Strategy Fund. The development of the automated quality control was supported by the Digital Earth project of the Helmholtz Association.

References

- Aufm Kampe, H. and Weickmann, H.: The effectiveness of natural and artificial aerosols as freezing nuclei, *Journal of Atmospheric Sciences*, 8, 283–288, [https://doi.org/10.1175/1520-0469\(1951\)008%3C0283:TEONAA%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(1951)008%3C0283:TEONAA%3E2.0.CO;2), 1951.
- 520 Barker, M., Chue Hong, N. P., Katz, D. S., Lamprecht, A.-L., Martinez-Ortiz, C., Psomopoulos, F., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., et al.: Introducing the FAIR Principles for research software, *Scientific Data*, 9, 622, <https://doi.org/10.1038/s41597-022-01710-x>, 2022.
- Bi, K., McMeeking, G., Ding, D., Levin, E., DeMott, P., Zhao, D., Wang, F., Liu, Q., Tian, P., Ma, X., et al.: Measurements of ice nucleating particles in Beijing, China, *Journal of Geophysical Research: Atmospheres*, 124, 8065–8075, <https://doi.org/10.1029/2019JD030609>, 2019.
- 525 Bigg, E.: A new technique for counting ice-forming nuclei in aerosols, *Tellus*, 9, 394–400, <https://doi.org/10.1111/j.2153-3490.1957.tb01895.x>, 1957.
- Bogert, P.: Ice-nucleating particles in the free troposphere: long-term observation and first measurements at cirrus formation temperatures using the novel Portable Ice Nucleation Experiment PINEair, <https://doi.org/10.5445/IR/1000174265>, 2024.
- 530 Böhmländer, A., Lacher, L., Fösig, R., Büttner, N., Nadolny, J., Brus, D., Doulgeris, K.-M., and Möhler, O.: Measurement of the ice-nucleating particle concentration with the Portable Ice Nucleation Experiment during the Pallas Cloud Experiment 2022, *Earth System Science Data Discussions*, pp. 1–10, <https://doi.org/10.5194/essd-2025-89>, 2025.
- Brunner, C. and Kanji, Z. A.: Continuous online monitoring of ice-nucleating particles: development of the automated Horizontal Ice Nucleation Chamber (HINC-Auto), *Atmospheric Measurement Techniques*, 14, 269–293, <https://doi.org/10.5194/amt-14-269-2021>, 2021.
- 535 Büttner, N. and Fösig, R.: PIA Software, <https://doi.org/10.5281/zenodo.15592883>, 2024.
- Canzi, A., Freney, E., Grzegorzczak, P., Baray, J., Lacher, L., and Planche, C.: Unraveling ice nucleating particle concentration variability: Insights into source emissions origin and parameterizations, *Journal of Geophysical Research: Atmospheres*, 130, e2024JD041258, <https://doi.org/10.1029/2024JD041258>, 2025.
- 540 Chen, Q., Yin, Y., Jiang, H., Chu, Z., Xue, L., Shi, R., Zhang, X., and Chen, J.: The roles of mineral dust as cloud condensation nuclei and ice nuclei during the evolution of a hail storm, *Journal of Geophysical Research: Atmospheres*, 124, 14 262–14 284, <https://doi.org/10.1029/2019JD031403>, 2019.
- Chou, C., Stetzer, O., Weingartner, E., Jurányi, Z., Kanji, Z. A., and Lohmann, U.: Ice nuclei properties within a Saharan dust event at the Jungfraujoch in the Swiss Alps, *Atmospheric Chemistry and Physics*, 11, 4725–4738, <https://doi.org/10.5194/acp-11-4725-2011>, 2011.
- 545 Cziczo, D. J., Ladino, L., Boose, Y., Kanji, Z. A., Kupiszewski, P., Lance, S., Mertes, S., and Wex, H.: Measurements of ice nucleating particles and ice residuals, *Meteorological Monographs*, 58, 8–1, <https://doi.org/10.1175/AMSMONOGRAPHS-D-16-0008.1>, 2017.
- DeMott, P. J., Prenni, A. J., Liu, X., Kreidenweis, S. M., Petters, M. D., Twohy, C. H., Richardson, M., Eidhammer, T., and Rogers, D.: Predicting global atmospheric ice nuclei distributions and their impacts on climate, *Proceedings of the National Academy of Sciences*, 107, 11 217–11 222, <https://doi.org/10.1073/pnas.0910818107>, 2010.
- 550 DeMott, P. J., Möhler, O., Stetzer, O., Vali, G., Levin, Z., Petters, M. D., Murakami, M., Leisner, T., Bundke, U., Klein, H., Kanji, Z. A., Cotton, R., Jones, H., Benz, S., Brinkmann, M., Rzesanke, D., Saathoff, H., Nicolet, M., Saito, A., Nillius, B., Bingemer, H., Abbatt, J., Ardon, K., Ganor, E., Georgakopoulos, D. G., and Saunders, C.: Resurgence in Ice Nuclei Measurement Research, *Bulletin of the American Meteorological Society*, 92, 1623–1635, <https://doi.org/10.1175/2011bams3119.1>, 2011.

- Diehl, K. and Grützun, V.: Model simulations with COSMO-SPECS: impact of heterogeneous freezing modes and ice nucleating particle types on ice formation and precipitation in a deep convective cloud, *Atmospheric Chemistry and Physics*, 18, 3619–3639, <https://doi.org/10.5194/acp-18-3619-2018>, 2018.
- Fan, J., Leung, L. R., Rosenfeld, D., and DeMott, P. J.: Effects of cloud condensation nuclei and ice nucleating particles on precipitation processes and supercooled liquid in mixed-phase orographic clouds, *Atmospheric Chemistry and Physics*, 17, 1017–1035, <https://doi.org/10.5194/acp-17-1017-2017>, 2017.
- Field, P. and Heymsfield, A.: Importance of snow to global precipitation, *Geophysical Research Letters*, 42, 9512–9520, <https://doi.org/10.1002/2015GL065497>, 2015.
- Free Software Foundation Europe: REUSE Software, <https://reuse.software/>, 2025.
- Hazra, A., Mallick, C., Mohan, G. M., Vani, K. G., Kumar, V. A., Chowdhury, R., Chaudhari, H. S., Das, S. K., Konwar, M., Pokhrel, S., et al.: Understanding the impact of ice nucleation on lightning and rainfall: A case study, *Atmospheric Research*, 278, 106350, <https://doi.org/10.1016/j.atmosres.2022.106350>, 2022.
- Herbert, R. J., Lacher, L., Böhmüller, A., Tarn, M. D., Canzi, A., Pantoya, A., Freney, E., Höhler, K., Bogert, P., Planche, C., Tian, P., Adams, M., Barr, S., Brus, D., Büttner, N., Daily, M., Doulgieris, K., Eleftheriadis, K., Forster, G., Fösig, R., Georgakopoulos, D., Gini, M., Hallar, A. G., Krejci, R., Ludewig, E., Mazzola, M., McCubbin, I. B., Petäjä, T., Robinson, J., Vogel, F., Zieger, P., Arnold, S. R., Carslaw, K. S., Hiranuma, N., Möhler, O., and Murray, B. J.: GloPINE dataset: model-ready measurements of INP concentrations using PINE instruments, *Earth System Science Data Discussions*, 2026, 1–24, <https://doi.org/10.5194/essd-2026-41>, 2026.
- Heymsfield, A. J., Schmitt, C., Bansemer, A., Gettelman, A., Field, P. R., Liu, C., et al.: Contributions of the liquid and ice phases to global surface precipitation: Observations and global climate modeling, *Journal of the Atmospheric Sciences*, 77, 2629–2648, <https://doi.org/10.1175/JAS-D-19-0352.1>, 2020.
- Lacher, L., Lohmann, U., Boose, Y., Zipori, A., Herrmann, E., Bukowiecki, N., Steinbacher, M., and Kanji, Z. A.: The Horizontal Ice Nucleation Chamber (HINC): INP measurements at conditions relevant for mixed-phase clouds at the High Altitude Research Station Jungfraujoch, *Atmospheric Chemistry and Physics*, 17, 15 199–15 224, <https://doi.org/10.5194/acp-17-15199-2017>, 2017.
- Lacher, L., Adams, M. P., Barry, K., Bertozzi, B., Bingemer, H., Boffo, C., Bras, Y., Büttner, N., Castarede, D., Cziczó, D. J., DeMott, P. J., Fösig, R., Goodell, M., Höhler, K., Hill, T. C. J., Jentzsch, C., Ladino, L. A., Levin, E. J. T., Mertes, S., Möhler, O., Moore, K. A., Murray, B. J., Nadolny, J., Pfeuffer, T., Picard, D., Ramírez-Romero, C., Ribeiro, M., Richter, S., Schrod, J., Sellegri, K., Stratmann, F., Swanson, B. E., Thomson, E. S., Wex, H., Wolf, M. J., and Freney, E.: The Puy de Dôme ICe Nucleation Intercomparison Campaign (PICNIC): comparison between online and offline methods in ambient air, *Atmospheric Chemistry and Physics*, pp. 2651–2678, <https://doi.org/10.5194/acp-24-2651-2024>, 2024.
- Laj, P., Myhre, C. L., Riffault, V., Amiridis, V., Fuchs, H., Eleftheriadis, K., Petäjä, T., Salameh, T., Kivekäs, N., Juurola, E., Saponaro, G., Philippin, S., Cornacchia, C., Arboledas, L. A., Baars, H., Claude, A., Mazière, M. D., Dils, B., Dufresne, M., Evangeliou, N., Favez, O., Fiebig, M., Haeffelin, M., Herrmann, H., Höhler, K., Illmann, N., Kreuter, A., Ludewig, E., Marinou, E., Möhler, O., Mona, L., Murberg, L. E., Nicolae, D., Novelli, A., O'Connor, E., Ohneiser, K., Altieri, R. M. P., Picquet-Varrault, B., van Pinxteren, D., Pospichal, B., Putaud, J.-P., Reimann, S., Siomos, N., Stachlewska, I., Tillmann, R., Voudouri, K. A., Wandinger, U., Wiedensohler, A., Apituley, A., Comerón, A., Gysel-Beer, M., Mihalopoulos, N., Nikolova, N., Pietruczuk, A., Sauvage, S., Sciare, J., Skov, H., Svendby, T., Swietlicki, E., Tonev, D., Vaughan, G., Zdimal, V., Baltensperger, U., Doussin, J.-F., Kulmala, M., Pappalardo, G., Sundet, S. S., and Vana, M.: Aerosol, Clouds and Trace Gases Research Infrastructure (ACTRIS): The European Research Infrastructure Supporting Atmospheric Science, *Bulletin of the American Meteorological Society*, 105, E1098 – E1136, <https://doi.org/10.1175/BAMS-D-23-0064.1>, 2024.

- Lin, Y., Fan, J., Li, P., Leung, L.-y. R., DeMott, P. J., Goldberger, L., Comstock, J., Liu, Y., Jeong, J.-H., and Tomlinson, J.: Modeling impacts of ice-nucleating particles from marine aerosols on mixed-phase orographic clouds during 2015 ACAPEX field campaign, *Atmospheric Chemistry and Physics*, 22, 6749–6771, <https://doi.org/10.5194/acp-22-6749-2022>, 2022.
- 595 Lohmann, U. and Feichter, J.: Global indirect aerosol effects: a review, *Atmospheric Chemistry and Physics*, 5, 715–737, <https://doi.org/10.5194/acp-5-715-2005>, 2005.
- Möhler, O., Stetzer, O., Schaefers, S., Linke, C., Schnaiter, M., Tiede, R., Saathoff, H., Krämer, M., Mangold, A., Budz, P., et al.: Experimental investigation of homogeneous freezing of sulphuric acid particles in the aerosol chamber AIDA, *Atmospheric Chemistry and Physics*, 3, 211–223, <https://doi.org/10.5194/acp-3-211-2003>, 2003.
- 600 Muhlbauer, A. and Lohmann, U.: Sensitivity studies of aerosol–cloud interactions in mixed-phase orographic precipitation, *Journal of the Atmospheric Sciences*, 66, 2517–2538, <https://doi.org/10.1175/2009JAS3001.1>, 2009.
- Mülmenstädt, J., Sourdeval, O., Delanoë, J., and Quaas, J.: Frequency of occurrence of rain from liquid-, mixed-, and ice-phase clouds derived from A-Train satellite retrievals, *Geophysical Research Letters*, 42, 6502–6509, <https://doi.org/10.1002/2015GL064604>, 2015.
- Murray, B. J., Carslaw, K. S., and Field, P. R.: Opinion: Cloud-phase climate feedback and the importance of ice-nucleating particles, *Atmospheric Chemistry and Physics*, 21, 665–679, <https://doi.org/10.5194/acp-21-665-2021>, 2021.
- 605 Möhler, O., Adams, M., Lacher, L., Vogel, F., Nadolny, J., Ullrich, R., Boffo, C., Pfeuffer, T., Hobl, A., Weiß, M., Vepuri, H. S. K., Hiranuma, N., and Murray, B. J.: The Portable Ice Nucleation Experiment (PINE): a new online instrument for laboratory studies and automated long-term field observations of ice-nucleating particles, *Atmos. Meas. Tech.*, 14, 1143–1166, <https://doi.org/10.5194/amt-14-1143-2021>, 2021.
- openCARP consortium, Houillon, M., Klar, J., Loewe, A., and Stary, T.: FACILE-RS (v3.3.0), <https://doi.org/10.35097/15htk3b7605e87yp>, 610 2025.
- Read the Docs, Inc & contributors: Read the Docs, <https://about.readthedocs.com/>, 2025.
- Rogers, D. C.: Development of a continuous flow thermal gradient diffusion chamber for ice nucleation studies, *Atmospheric research*, 22, 149–181, [https://doi.org/10.1016/0169-8095\(88\)90005-1](https://doi.org/10.1016/0169-8095(88)90005-1), 1988.
- Schmidt, L., Schäfer, D., Geller, J., Lünenschloss, P., Palm, B., Rinke, K., Rebmann, C., Rode, M., and Bumberger, J.: System 615 for automated Quality Control (SaQC) to enable traceable and reproducible data streams in environmental science, 169, 105 809, <https://doi.org/10.1016/j.envsoft.2023.105809>, 2023.
- Schneider, J., Höhler, K., Heikkilä, P., Keskinen, J., Bertozzi, B., Bogert, P., Schorr, T., Umo, N. S., Vogel, F., Brasseur, Z., Wu, Y., Hakala, S., Duplissy, J., Moiseev, D., Kulmala, M., Adams, M. P., Murray, B. J., Korhonen, K., Hao, L., Thomson, E. S., Castarède, D., Leisner, T., Petäjä, T., and Möhler, O.: The seasonal cycle of ice-nucleating particles linked to the abundance of biogenic aerosol in boreal forests, 620 *Atmospheric Chemistry and Physics*, 21, 3899–3918, <https://doi.org/10.5194/acp-21-3899-2021>, 2021.
- Tarn, M. and Murray, B.: Data from the Portable Ice Nucleation Experiment (PINE) during the CountIce (part 1) 2021–2022 campaign, <https://doi.org/10.5281/ZENODO.17451020>, 2025.
- Vali, G., DeMott, P., Möhler, O., and Whale, T.: A proposal for ice nucleation terminology, *Atmospheric Chemistry and Physics*, 15, 10 263–10 270, <https://doi.org/10.5194/acp-15-10263-2015>, 2015.
- 625 Vogel, F.: Short-term Variation in Measurements of Atmospheric Ice-Nucleating Particle Concentrations, <https://doi.org/10.5445/IR/1000151147>, 2022.
- Vogel, F., Lacher, L., Nadolny, J., Saathoff, H., Leisner, T., and Möhler, O.: Development and validation of a new cloud simulation experiment for lab-based aerosol–cloud studies, *Review of Scientific Instruments*, 93, <https://doi.org/10.1063/5.0098777>, 2022.

- Vogel, F., Adams, M. P., Lacher, L., Foster, P., Porter, G. C. E., Bertozzi, B., Höhler, K., Schneider, J., Schorr, T., Umo, N. S., Nadolny, J., Brasseur, Z., Heikkilä, P., Thomson, E. S., Büttner, N., Daily, M. I., Fösig, R., Harrison, A. D., Keskinen, J., Proske, U., Duplissy, J., Kulmala, M., Petäjä, T., Möhler, O., and Murray, B. J.: Ice-nucleating particles active below -24°C in a Finnish boreal forest and their relationship to bioaerosols, *EGUsphere*, 2024, 1–25, <https://doi.org/10.5194/egusphere-2024-853>, 2024.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., et al.: The FAIR Guiding Principles for scientific data management and stewardship, *Scientific data*, 3, 1–9, <https://doi.org/10.1038/sdata.2016.18>, 2016.
- Yang, Y., Deng, W., Li, F., and Zhang, T.: The impact of aerosols as ice nucleating particle on microphysics and electrification in the cumulus model, *Journal of Aerosol Science*, 180, 106 401, <https://doi.org/10.1016/j.jaerosci.2024.106401>, 2024.

Table A2. Overview of qc tests, including the software internal flags, the error type, whether the test is applied to a single data point or the whole run, the parameters of the test, and the short description used in the headers of the *flags_instrument.txt* and *flags_metadata.txt* files.

test	flag	type	run or data point	parameters	short
Fe outlier	459	info/warning	data point	see Sec. 4.3.3	Fe_outlier
Fm outlier	460	info/warning	data point	see Sec. 4.3.3	Fm_outlier
particle concentration range	461	warning	data point	0 to 700'000 stdL ⁻¹	concentration_range
variability ice threshold	463	warning	run	see Sec 4.3.6	ice_threshold_2_sigma
supersaturation	549	warning	run	DP higher than Ti5	supersaturation
difference between Ti1 and Ti2	635	warning	run	Ti1 higher than Ti2	Ti1/Ti2_diff
difference between Ti2 and Ti3	636	warning	run	Ti2 higher than Ti3	Ti2/Ti3_diff
difference between Ti3 and Ti4	637	warning	run	Ti3 higher than Ti4	Ti3/Ti4_diff
difference between Ti4 and Ti5	638	error	run	Ti4 higher than Ti5	Ti4/Ti5_diff
comparison to pre-set Fe, Fm	664, 665	warning	run	deviation: $\pm 15\%$	Fe, Fm
run times	670	error	run	see Sec. 4.3.5	increasing_time
icing of inlet	677	error	run	see Sec. 4.3.11	dur_expansion
Pch range	684	error	data point	495 to 1055 mbar	Pch_range
DP range	685	info/warning	data point	-60 to 8 °C	DP_range
Fe range	686, 688	info/warning	data point	0.9 to 5.1 L/min & 0 to 0.02 L/min	Fe_range & Fe_range_zero
Fm range	687	info/warning	data point	0.9 to 5.1 L/min	Fm_range
Ti1 - Ti3 range	689 - 691	warning	data point	-60 to 0 °C	Ti1_range - Ti3_range
Ti4, Ti5 range	692, 693	error	data point	-60 to 0 °C	Ti4_range, Ti5_range
Tw1 - Tw3 range	694 - 696	warning	data point	-60 to 0 °C	Tw1_range - Tw3_range
DP flatline	698	warning	data point	stuck points: 10	DP_flatline
Pch flatline	699	warning	data point	stuck points: 10	Pch_flatline
Ti1 - Ti4 flatline	700 - 703	warning	data point	stuck points: 5	Ti1_flatline - Ti4_flatline
Ti5 flatline	704	error	data point	stuck points: 5	Ti5_flatline
duration of expansion	708	error	run	see Sec. 4.3.9	dur_expansion
comparison to pre-set Pch	709	warning	run	deviation: $\pm 5\%$	pch_end
particles during refill	710	warning	data point	see Sec. 4.3.4	opc_refill
large aerosols during flush	711	warning	run	threshold: 50 % of INP concentration	inp_flush
missing opc data	997	error	run		concentration_missing
gaps in opc data	998	error	run	see Sec. 4.3.12	opc_data
gaps instrument data	999	error	data point		—