https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

AEROSOL Research Discussions

Predicting Ice Nucleation Particle properties in a Boreal Environment using machine learning

Yusheng Wu[1], Zoé Brasseur[1,2], Dimtri Castarède[3], Paavo Heikkilä[4], Jorma Keskinen[4], Ottmar Möhler[5], Markku Kulmala[1], Tuukka Petäjä[1], Erik S. Thomson[3], and Jonathan Duplissy[1,6]

[1]Institute for Atmospheric and Earth System Research/Physics, Faculty of Science, University of Helsinki, Helsinki, Finland
[2]Svalbard Integrated Arctic Earth Observing System (SIOS), SIOS Knowledge Centre, 9170 Longyearbyen, Norway
[3]Department of Chemistry and Molecular Biology, Atmospheric Science, University of Gothenburg, Gothenburg, Sweden
[4]Aerosol physics laboratory, Physics Unit, Faculty of engineering and natural sciences, Tampere University, Tampere, Finland
[5]Institute of Meteorology and Climate Research Atmospheric Aerosol Research, Karlsruhe Institute of Technology, Karlsruhe, Germany
[6]Helsinki Institute of Physics, University of Helsinki, Helsinki, Finland

Correspondence: Erik S. Thomson (erik.thomson@gu.se) and Jonathan Duplissy (jonathan.duplissy@helsinki.fi)

Abstract.

Mixed-phase clouds, which are dominant in mid- and high-latitude regions, strongly influence Earth's radiative balance and precipitation processes. Their formation depends critically on the presence of ice-nucleating particles (INPs), which are rare relative to cloud condensation nuclei. The HyICE-2018 measurement campaign took place at the SMEAR II station in the high-latitude boreal forest of Hyytiälä, Finland, between February and June 2018. Two continuous-flow diffusion chambers Portable Ice Nucleation Chamber I and II (PINC and PINCii) with high-frequency sampling were deployed to measure INP concentrations. We applied machine-learning techniques to explore predictors of INP variability using more than 500 high-resolution atmospheric, aerosol, and ecosystem variables measured continuously at Station for Measuring Ecosystem-Atmosphere Relations (SMEAR) II. We identify distinct differences between winter and spring/summer measurements. The winter measurements conducted with PINC appear to be nearly independent of any monitored variable. In contrast, the spring/-summer measurements conducted with PINCii appear to be more closely linked to and responsive to ambient aerosol properties. Furthermore, we find that classical parameterizations based on particle concentration overestimate observed INP concentrations in the boreal environment. However, similar empirical fits based on local proxies, such as a marker of biogenic aerosol or nitrate, yield improved agreement during spring and summer, while no improvement occurs during winter. These results underscore the need for site-specific parameterizations to capture INP variability in the complex boreal environments.

1 Introduction

Clouds, especially at high-latitudes, exist in a sensitive balance. Most clouds, including those that bring precipitation to northern latitudes, form between 0°C and -40°C and are "mixed-phase", meaning that both liquid water and ice exist within their bounds (Mülmenstädt et al., 2015). Obviously, for ice to exist, the temperature must be less than 0°C, meaning that at these temperatures liquid water would prefer to be ice. Even with this inherently unstable co-existence of ice and liquid water,

https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

mixed-phase clouds are widespread (Mülmenstädt et al., 2015) and persist for many hours and even days, which is important for Earth's energy budget and precipitation processes (Shupe et al., 2013). Beyond high-latitudes, mixed-phase clouds play out-sized roles in regulating climate, especially in the Arctic where underlying feedbacks have amplifying effects (Shupe and Intrieri, 2004; Morrison et al., 2012). Importantly, mixed-phase clouds only appear with the help of small particles in the atmospheric aerosol that provide the initial seeds for ice and liquid droplet formation. Those seeds, ice nucleating particles (INPs) and cloud condensation nuclei (CCN) are fundamental to cloud processes. Of these, the consensus is that INP are significantly more rare than CCN, and there remains great uncertainty when it comes to predicting INP occurrence and abundance (DeMott et al., 2010).

In 2018 an intensive measurement campaign (HyICE-2018; Brasseur et al., 2022) was undertaken at the Hyytiälä, Finland, Station for Measuring Ecosystem-Atmosphere Relations (SMEAR) II to focus on measuring INPs in the boreal environment. A strong motivation for co-locating the HyICE-2018 INP measurements at the SMEAR II station was due to the fact that SMEAR II is a heavily instrumented station for monitoring a plethora of meteorological, ecological, hydrological, etc., variables (Hari and Kulmala, 2005). It is in fact one of the most significantly instrumented such stations globally, as illustrated by its inclusion in many measurement networks, e.g., ICOS, ACTRIS, SITES, etc. Previously, several results from both the intensive HyICE-2018 campaign (Paramonov et al., 2020; Schneider et al., 2021; Brasseur et al., 2022, 2024; Vogel et al., 2024), and long-term studies bridging the campaign (Schneider et al., 2021) have been published. To complement those studies, herein we attempt to exploit the wide scope of parameters that are measured at SMEAR II with high time resolution. In addition to classical correlation studies, commensurate with the already published HyICE studies, we attempt to use several machine learning algorithms to investigate the emergence of non-obvious (or intiuitive) connections within the available high-frequency data.

The results of our study are educational but cautionary. Strong links between INP data and fundamental chemical signatures of the atmospheric aerosol do exist, but are likely open to over-interpretation. Even with several hundred variables measured with high frequency time resolution, correlation does not necessarily illuminate causation. With regard to INPs, which are a small fraction of all atmospheric particulate, the indication is that improved mechanistic understanding remains abstruse.

2 Methods and Data

2.1 Study site and period

The HyICE-2018 campaign took place at the SMEAR II measurement station in Hyytiälä, Finland (Hari and Kulmala, 2005) located within a sub-Arctic boreal environment at 61°51'N, 24°17'E and 181 m above sea level. The station set up and details during HyICE-2018 are fully described in Brasseur et al. (2022). For the purposes of this study, the time period of interest is from mid-February 2018 (February 19) to mid-June 2018 (June 10), during which time two continuous flow diffusion chambers (CFDCs; PINC - Portable Ice Nucleation Chamber, and PINCii - second-generation PINC) sharing design characteristics were operated to sample ambient aerosol and measure INPs with high time resolution.

https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

AEROSOL Research Discussions

2.2 Complementary Data and Machine Learning

A significant motivation to co-locate the HyICE-2018 campaign at the SMEAR II infrastructure was to take advantage of the available ecosystem and atmospheric monitoring that is continuously undertaken at the station. Sampling atmospheric particles is an exercise in capturing rare events, a task that is amplified when measuring INPs. Although ice nucleation activity is highly temperature dependent, for much of the heterogeneous freezing temperature spectra (between $\approx -38°$and $0°$C) INPs represent a small fraction (as low as one in a million) of all particles (DeMott et al., 2010). That means that there is significant scientific interest in identifying tracers or other indicators of freezing activity that can be used to follow ice nucleation activity.

In addition to the INP measurements, for the purposes of this investigation, 509 individually monitored variables that are continuously recorded at SMEAR II were interrogated (data are available online at smear.avaa.csc.fi). Those measurements, recorded with high time resolutions, are mostly atmospheric in character and can be broadly categorized as: meteorological, radiological, soil, characteristics of aerosols and gases, and associated data products of the aforementioned. Naturally, the resulting multidimensional data includes many potential redundancies and/or irrelevancies, in terms of illuminating connections to INPs. Therefore, data filtering techniques were implemented to reduce the dimensionality and redundancy within the data. As a first pass data features were not considered if they contained the following: I. Excessive NaN values, such as are often generated when concentrations are below instrument sensitivities. II. Data with little or no variability, i.e., constant values. III. Data which was duplicated, like the same parameter sampled at different heights but without showing systematic differences.

The study also employs the Wideband Integrated Bioaerosol Sensor (WIBS), a real-time, single-particle instrument designed for atmospheric bioaerosol detection. The WIBS uses dual ultraviolet excitation wavelengths and subsequent fluorescence emission measurement to infer the presence of biological material in individual aerosol particles. Light scattering from a 635 nm diode laser determines particle size ($> 0.5\mu$m; Tang et al., 2022).

2.3 INP concentration measurements

The foundation for this study is trying to link measured INPs with other environmental parameters, in an effort to illuminate the sources and mechanisms that may drive atmospheric ice formation above boreal environments.

Of the operational sampling units from the HyICE-2018 campaign, two CFDCs have time resolutions that make their data practical to compare with other parameters sampled at high frequency, and are somewhat straightforwardly intercompared as demonstrated by Brasseur et al. (2022). PINC and PINCii are parallel-plate CFDCs designed to measure INPs, present in a sample air flow that is sandwiched between clean sheath air. Ice nucleation conditions within the sample air flow are modeled for CFDCs based on measured wall temperatures and the known saturation vapor pressure for the ice covered walls (Garimella et al., 2016, 2017; Castarède et al., 2023).

PINC, the first-generation instrument, has been widely used in field campaigns for over a decade (Chou et al., 2011; Paramonov et al., 2020). Originally built for airborne deployments, it features a compact design with a main chamber (568 mm in length) and a 230 mm evaporation section. It is effective for ambient INP measurements, but has limitations in terms of cooling power, wall temperature control, and laminar flow stability. Its operation typically involves fixed temperature and humidity

https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

conditions that target mixed-phase cloud scenarios. During HyICE-2018 PINC was operated from 19 February to 2 April at a fixed sample lamina temperature $T_l = -31°C$ and a relative humidity with respect to water $RH_w = 105\%$. PINC utilized a 1 L min^{-1} dried ($RH \leq 30\%$) sample flow with an inline cyclone impactor used to eliminate sampling particles larger than 2.5 μm. PINC measurements from HyICE-2018 are presented in detail in Paramonov et al. (2020).

PINCii was originally designed and built based on upgrading PINC (Castarède et al., 2023), utilizing knowledge gained from more than a decade of experimental use (Stetzer et al., 2008; Chou et al., 2011; Kanji et al., 2013, 2019), and thus significantly enhancing the PINC design and capabilities. PINCii is fully described in Castarède et al. (2023). It has a much larger chamber, approximately twice the length(s) in both the main (1000 mm) and the evaporation (440 mm) sections, which provides extended residence time for aerosol particles and ice nucleation and crystal growth. These longer sections, and thus increased growth time, improve the counting statistics and the instrument resolution. PINCii incorporates a pre-cooled sheath air system, as well as an enhanced wall temperature control and monitoring, with dense thermocouple arrays and a sophisticated cooling system that allows cooling to ≈ -67 °C. PINCii also introduces improved methods for analyzing thermodynamic conditions by accounting for wall inhomogeneities and ice layer thickness, increasing the accuracy of RH and temperature estimations. During HyICE-2018, PINCii also utilized a dried sample aerosol flow of 1 L min^{-1} (measuring from 22 April to 10 June) with aerosol lamina temperature $T_l = -32°C$ and $RH_w = 105\%$ and a 2.5 μm cyclone impactor.

Instrumental constraints and availability made it impossible for PINC and PINCii to operate simultaneously during HyICE-2018. However, their operation in succession means that for the bulk of the campaign there exists high-frequency data coverage, which appears self-consistent. An instrument comparison summary is available within the campaign Measurement Report (Brasseur et al., 2022), where the inconsistencies with another parallel-plate CFDC chamber, SPectrometer for Ice Nuclei (SPIN), make plain its absence here.

3 Results and Discussion

3.1 INP and other monitored variables

The time series of INP concentrations measured by PINC and PINCii with 20 minute sampling windows is presented in Fig. 1(a). Each point represents a 20 minute sampling window with a subtracted background interpolated from particle free measurements before and after the sampling periods. Between instruments and across the entire campaign period no significant differences in the absolute concentrations and spread of measurements are noted. However, by visual inspection the PINCii measurements in spring and early summer appear to contain slightly greater variability.

Also depicted in Fig. 1 are the time series of several other key monitored variables. Some, like temperature (Fig. 1 (b)) and snow depth (Fig. 1 (b), gray trace) reflect the seasonal change from winter to spring and summer, while others are quite commonly associated with INP abundance (e.g., particle number concentrations, Fig. 1 (c)). Given that previously published results from daily and/or multi-day samples have demonstrated seasonal variability for Hyytiälä INP and linked observed trends with biogenic emissions (Schneider et al., 2021; Proske et al., 2025), the time evolution of aerosol mass fractions are

https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

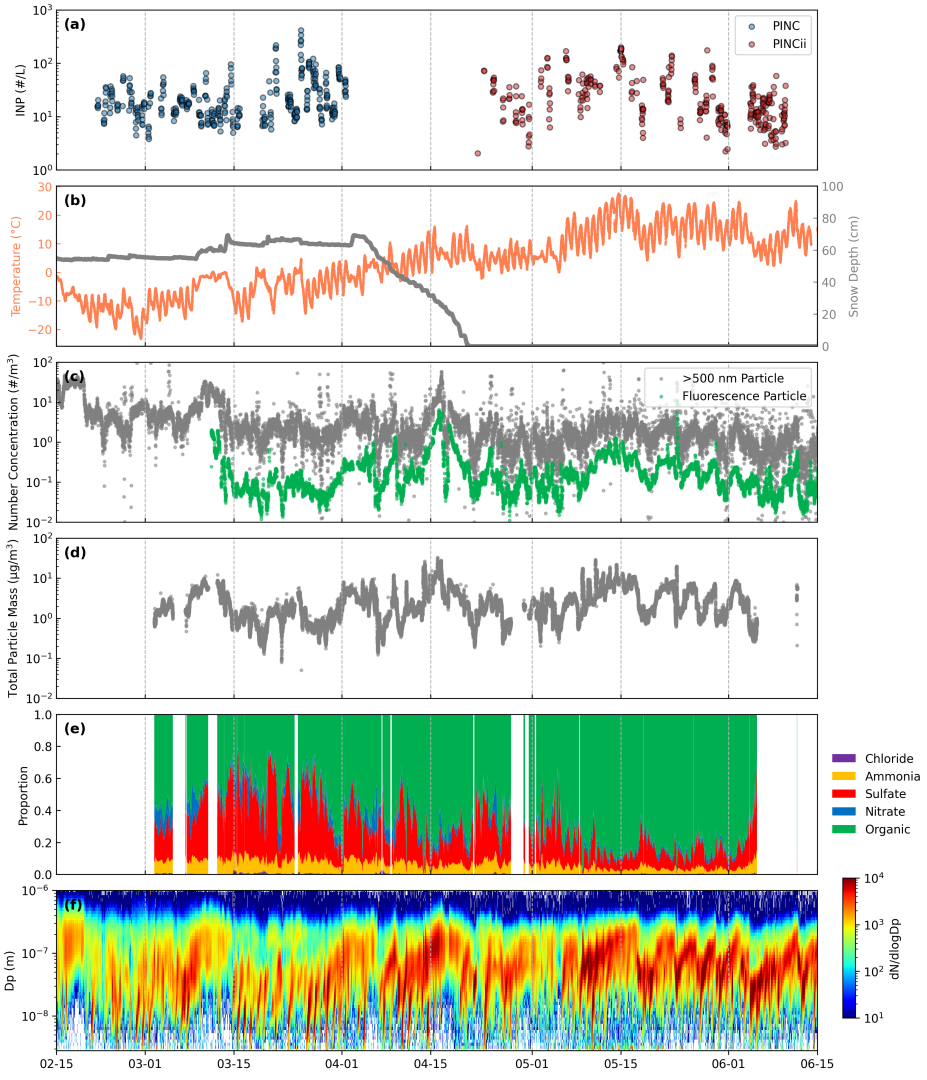

Figure 1. Time series of atmospheric, aerosol, and INP properties measured during the HyICE-2018 campaign at the SMEAR II station in Hyytiälä, Finland (19 February–10 June 2018). (a) INP concentrations measured by the Portable Ice Nucleation Chamber (PINC; circles) and its updated version, PINCii (squares), shown on a logarithmic scale. (b) Air temperature (coral, left axis) and snow depth (grey, right axis), illustrating the transition from winter through spring to summer. (c) Number concentrations of particles with diameters > 500 nm (grey) and fluorescent biological aerosol particles (> 500 nm, green). (d) Total particle mass concentration for particles with d $< 1\mu$m measured with an Aerosol Mass Spectrometer (AMS). (e) Mass fractions of AMS-measured chemical components: chloride (purple), ammonia (yellow), sulfate (red), nitrate (blue), and organics (green). (f) Particle number size distributions (PNSD) measured with a Differential Mobility Particle Sizer (DMPS), with color indicating dN/dlogDp. Together, the panels capture seasonal changes, aerosol chemical and physical characteristics, and the variability in INP abundance across the intensive campaign period.

https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

also plotted (Fig. 1 (e)). The most notable change in that time series is the increasing organic aerosol fraction with the change in season from winter to summer.

Finally, the time series of sub-micron particles (Fig. 1 (f)), clearly exhibits the characteristic "banana" curves, indicative of new particle formation, that Hyytiälä is well-known for (Dal Maso et al., 2005; Kulmala et al., 2013).

3.2 Machine Learning

The relationships suggested by the time series in Fig. 1 motivate a more objective evaluation of how INP variability relates to the broader set of concurrently measured parameters. Given the large number of variables and their inherent covariance, simple visual or pairwise analyses are insufficient to robustly rank their relevance. Therefore we apply statistical and machine-assisted approaches to explore reducing the feature space and to identify consistent associations with INP concentrations.

The variables presented in Fig. 1 are a small subset of the total 509 recorded variables that were investigated within this study. Of those 509 sampled variables the previously outlined first-pass dimensional reduction left 84 variables. The remaining 84 parameters were further interrogated using several classic and newer machine-supported analysis techniques. These included pairwise correlation, decision tree and random forest learning algorithms (Fig. 2), and principle component and K-means clustering analyses. While it was difficult to extract quantitative results in all cases, each separate treatment resulted in qualitatively similar results, as illustrated in Fig. 2.

Variables were excluded from the original set of 509 if they contained excessive numbers of NaN values, exhibited very low variability (i.e., were nearly constant), or were effectively redundant (for example, the same parameter, such as temperature, measured at different heights without meaningful differences). Particle size distribution measurements were consolidated into number concentrations over selected size ranges. In addition, several features were found to be strongly correlated (based on Pearson correlation), indicating redundant information; one example is the close correspondence between highly oxygenated organic molecule (HOM) monomers and organic nitrate.

The feature importance rankings from random forest models (e.g., Fig. 2) consistently assign high importance to variables that are known to be strongly correlated with ice nucleation activity. For example, biological particles that yield a fluorescence signal are known to be some of the most abundant INPs in many settings (Murray et al., 2012; Morris et al., 2014; Proske et al., 2025). The connection(s) between other highly ranked quantities that emerge and INP are sometimes less clear, but suggest closer investigation.

In Fig. 3 six variables are selected to illustrate their pairwise correlation with INP. The individual variables were selected based upon their highly ranked outcomes and physical intuition connecting the variables, measurements, and ice nucleation. For example, in Fig. 3 fluorescence and nitrate are more closely examined, but acetone and methanol, which largely co-vary with nitrate, are not. Ice nucleation has a documented dependence on particle size that emerges in many parameterizations (DeMott et al., 2015; Tobo et al., 2019), and combined with mass accounts for volume. Several indicators of black carbon (BC) are surprisingly highly ranked, given many INP studies suggest weak ice activity for BC (Mahrt et al., 2020a; Testa et al., 2024). Finally, because of Hyytialla's well documented legacy in organic aerosol measurements and new particle formation (Kulmala et al., 2013), organics are added as a variable of interest.

https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

AEROSOL Research Discussions

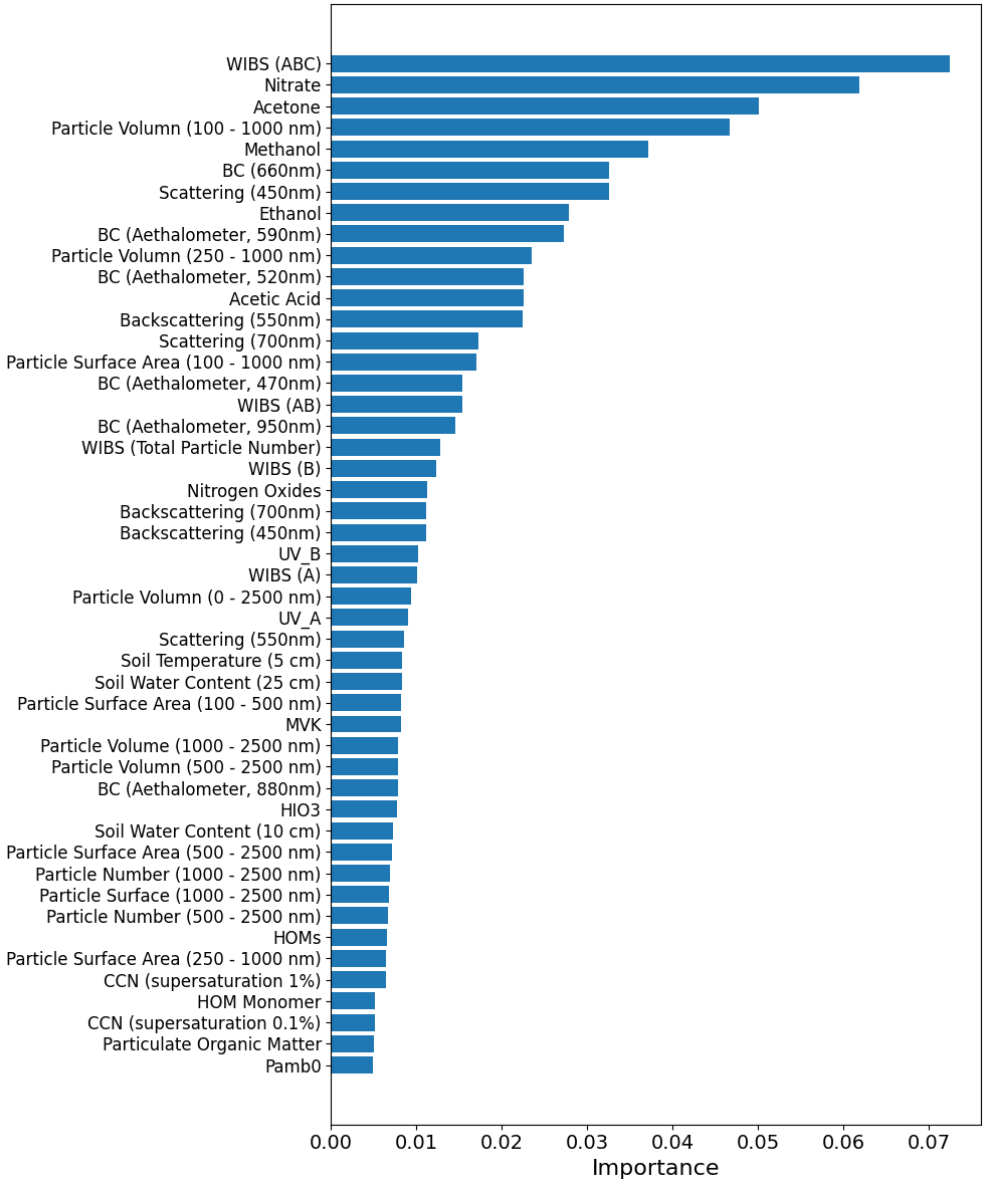

Figure 2. Feature importance analysis derived from a Random Forest model trained to predict ice-nucleating particle (INP) concentrations during the HyICE-2018 campaign at SMEAR II. A Random Forest is an ensemble of many simple decision-based models (decision trees) that each make a prediction using different subsets of the data. By averaging over all trees, the method identifies which input variables most strongly influence the model's ability to predict INP variability. The bars show the relative importance of each variable, expressed in arbitrary units. Higher values indicate that a variable contributed more often to successful decisions within the model. Variables with larger bars therefore have a stronger statistical association with INP concentrations, although this does not necessarily imply a causal relationship. For WIBS, ABC denotes particles that fluoresce above the detection threshold in three WIBS fluorescence channels; further details are given in (Savage et al., 2017).

https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

AEROSOL Research Discussions

In Fig. 3 measured INP is plotted on the vertical axis. In the upper panels variables with previously established linkages to INP are plotted on the horizontal axes. In addition to flourescent particles, particle number concentrations for particles larger than 0.5 μm are plotted. Particle concentrations of this size appear in several, widely used, INP parameterization schemes (Tobo et al., 2013; DeMott et al., 2015). Particle mass concentration, again largely a proxy for the number of large particles present in the aerosol, is plotted in Fig. 3(b). In the bottom panels several highly-ranked variables that constrain aerosol chemistry are plotted. For all of these cases, the organic mass concentration (Fig. 3(d)), nitrate mass concentration (Fig. 3(e)), and black carbon mass concentration (Fig. 3(f)), the direct correlation is as good or better than what is demonstrated in the upper panels for variables with previously established links to ice activity.

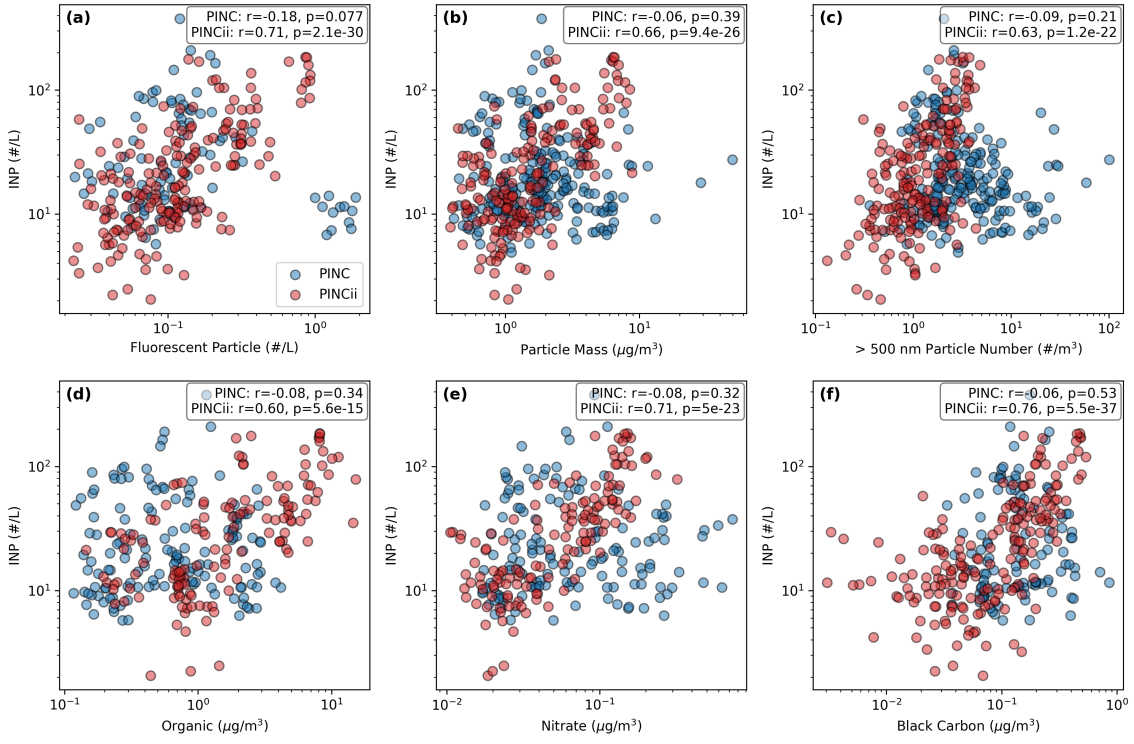

Figure 3. Log–log scatter plots showing correlations between ice-nucleating particle (INP) concentrations measured by PINC (blue) and PINCii (red) during HyICE-2018 and six selected aerosol parameters measured at SMEAR II, Hyytiälä, Finland: (a) fluorescent particle number concentration, (b) particle mass (1 nm–10 μm aerodynamic diameter), (c) number concentration of particles with diameters > 500 nm, (d) organic aerosol mass concentration, (e) nitrate aerosol mass concentration, and (f) black carbon mass concentration. Each point is the hourly mean, computed for each full hour when both variables were above zero. Pearson correlation coefficients (r) and corresponding p-values are shown for each instrument. Data are plotted separately for PINC and PINCii to highlight similarities and differences in observed relationships, and to illustrate potential links between INP abundance and aerosol chemical or physical properties.

While the individual learning algorithms that were tested were consistent and succeeded in bringing to light several variables that also exhibit strong pairwise correlation with INP, they do in and of themselves not succeed in shedding light on causation

https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

or sources of INP. This is perhaps unsurprising, as taken wholistically the highly ranked variables simply suggest that more dirty, mixed aerosol with higher particulate concentrations is more likely to be ice active. Similar observations have been made previously in several diverse ecosystems, from the subtropical marine boundary layer (Welti et al., 2018), to ocean basins (Welti et al., 2020), and globally distributed land-based samples ranging from Arctic to equatorial latitudes (Schrod et al., 2020). In those studies it has been noted that INP concentrations measured remotely from strong sources, often exhibit log-normal frequency distributions (Fig. 4(a)). This observation is directly analogous to random mixing (dilution) of trace pollutant species, (Ott, 1990) and suggests that the dominant INP signal originates from long-range transport of well-mixed aerosol. In Fig. 4 the relative frequency distributions of INP concentrations measured with PINC(blue) and PINCii(red) are plotted with unimodal (a) and bimodal (b) fitted curves. From the fitting it is clear that to a large degree the observed INP concentrations are well represented by simple log-normal distributions. Moreover, the PINC and PINCii distributions are highly similar. The exception, manifest as a spike in the tail of the PINCii distirbution, may be related to a particular source or series of events that were present during the PINCii sampling, but absent during the deep winter season. However, the divergence is not significant enough to robustly identify any real signal.

3.3 Parameterization

The qualitative connections that emerge between INP and aerosol characteristics (Fig. 3) from the machine learning algorithms (Fig. 2), signal an opportunity to explore the parameterization space. We expect existing parameterizations to account for the observed correlations related to aerosol particle abundance and biogenic origin (top panels Fig. 3). In contrast, the connections to aerosol chemical composition (bottom panels Fig. 3) are less obvious and, in fact, despite considerable explorations of black carbon, it is typically observed to have limited ice activity (Thomson et al., 2018; Adams et al., 2020; Santos et al., 2024), especially when it occurs as fresh soot. That said, there is some evidence that transport, oxidation, and aging enhance soot's ice activity (DeMott, 1990; Mahrt et al., 2020a, b).

3.3.1 Previous parameterizations

Several empirical parameterizations have been developed to predict INP concentrations as a function of temperature and aerosol properties. Among the most widely used are the formulations by DeMott et al. (2010) and Tobo et al. (2013), which express INP number concentration (n_{INP}) as a function of cloud temperature T in degrees Kelvin and aerosol number concentration [scm^{-3}] for particles larger than 0.5 μm ($n_{AP>0.5\mu m}$), and were derived specifically from CFDC datasets. DeMott et al. (2015) revised the 2010 parameterization for predicting INP concentrations to explicitly include mineral dust as a primary source. However, since mineral dust likely represents a negligible component in our boreal study environment, that is far from dust sources, we refer back to the 2010 formulation, which is:

$$n_{\mathrm{INP}} = a \times (273.16 - T)^{b} \times (n_{\mathrm{AP}>0.5\mu m})^{(c(273.16-T)+d)}, \quad (1)$$

where $a = 0.0000594$, $b = 3.33$, $c = 0.0264$, and $d = 0.0033$. The subsequent Tobo et al. (2013) parameterization follows a similar power law formulation but was updated using primary biological aerosol particle measurements. It is also investigated

https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

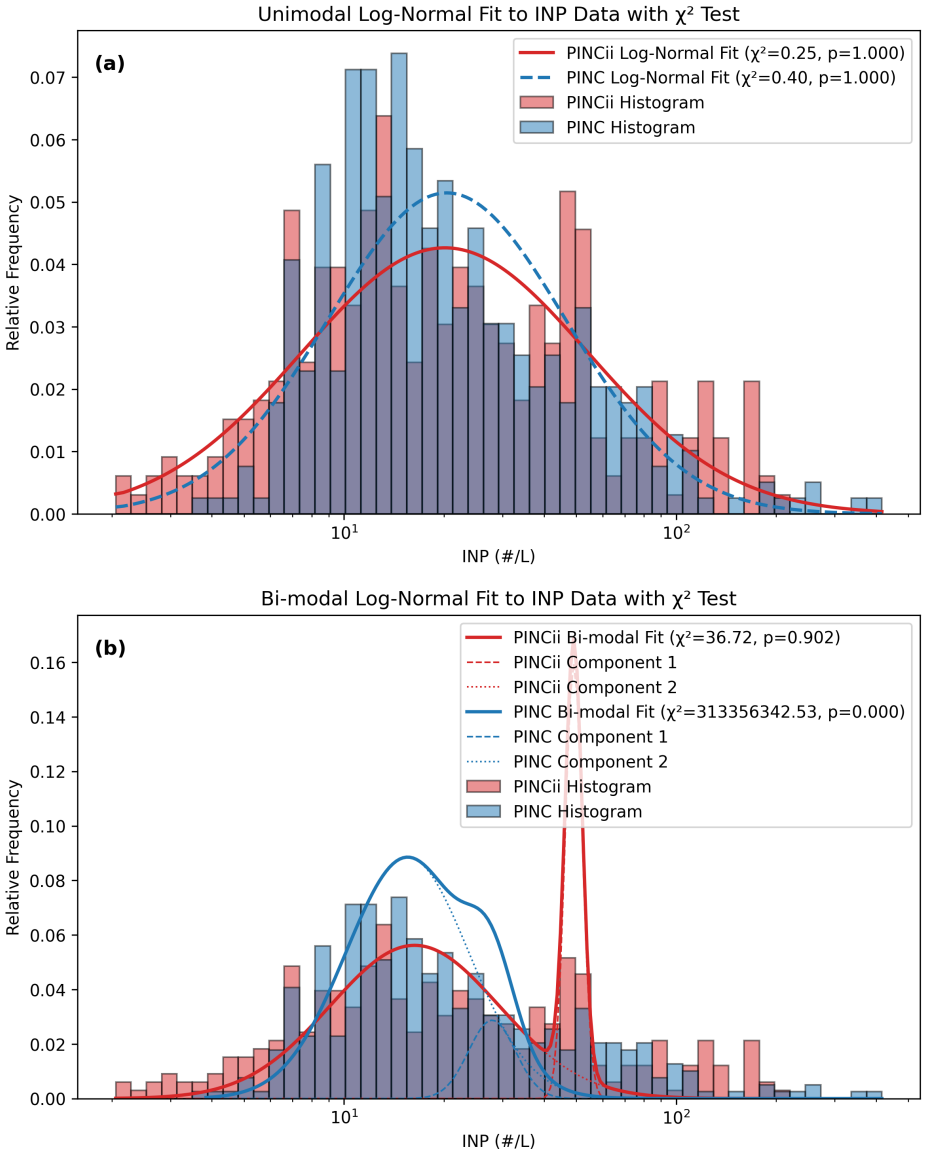

Figure 4. Normalized relative frequency distributions for INP concentrations measured by PINC and PINCii during the HyICE-2018 campaign, along with fitted log-normal probability density functions. (a) Unimodal log-normal fits to each instrument's INP histogram, with fit parameters evaluated using maximum likelihood estimation and goodness-of-fit assessed via χ^2 statistics. (b) Bimodal log-normal fits decomposed into two component modes, illustrating potential multi-population structure in the INP size/activation spectrum. The analysis highlights that while unimodal fits capture the central tendency, bi-modal representations better resolve the observed distribution tails for both instruments.

https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

because the work of Schneider et al. (2021), Vogel et al. (2024), and Proske et al. (2025), suggest a strong link between INP and primary biological activity at SMEAR-II. The Tobo et al. (2013) parameterization is,

$$n_{\mathrm{INP}} = (n_{\mathrm{AP}>0.5\mu m})^{(\alpha(273.16-T)+\beta)} \exp(\gamma(273.16-T)+\delta), \tag{2}$$

where $\alpha = -0.074$, $\beta = 3.8$, $\gamma = 0.414$, and $\delta = -9.671$ are suggested as coefficients. Further INP parameterizations have been developed from SMEAR-II-based immersion freezing measurements (Schneider et al., 2021; Brasseur et al., 2024). However, the immersion freezing focus differs from the operating principle of CFDCs and is largely in a different, warmer temperature regime. Moreover, the Schneider et al. (2021) study sought to capture longer-term INP trends. Thus, although the sampling of those studies was also located at and above SMEAR-II, it is not applied in our analysis of this high-frequency data.

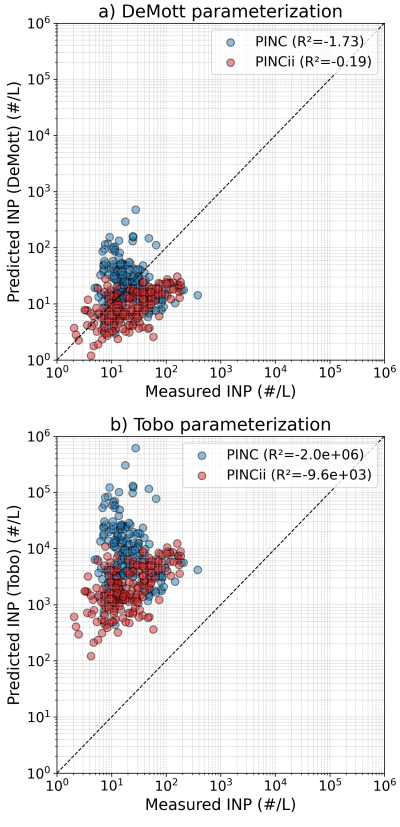

Figure 5. Predicted INP concentrations from (a) DeMott et al. (2010) and (b) Tobo et al. (2013) parameterizations plotted versus the INP concentrations measured using the PINC(blue) and PINCii(red) CFDCs. The 1:1 line is plotted and R^2 values corresponding to the coefficient(s) of determination for each fit are given within the legend. Negative R^2 values indicate that the parameterizations provide no more predictive strength than the arithmetic mean value.

https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

In Fig. 5 the INP predicted from these models is compared to the PINC (blue) and PINCii (red) measurements and the coefficient of determination (R^2) is used to quantify how well the models explain the observed variability. In all cases the R^2 has nonintuitive negative values, which can occur when fitting non-linear functions and indicates that the model predictions are less accurate than simply applying the observational mean as a predictor (i.e., a horizontal line at the mean value). Thus, both model formulations perform poorly when applied to the HyICE-2018, PINC and PINCii data, indicating that these existing parameterizations do not capture the observed boreal-specific INP variability (Fig. 5). This observation both agrees and disagrees with shorter-term observations made in the context of a focused intercomparison between instruments during the HyICE-2018 campaign, for which the Tobo et al. (2013) parameterization performed well, even as the DeMott et al. (2010) formulation consistently underpredicted measurements.

3.3.2 New parameterizations for boreal conditions

Although previously described parameterizations do not adequately capture the observed variability, they suggest the exploration of simple empirical power-law fitting between n_{INP} and aerosol and chemical tracers that have been highly ranked by the machine learning algorithms. We seek to determine whether such an approach can elucidate potential local INP proxies. We utilize a highly simplified, generic power law relationship,

$$n_{\mathrm{INP}} = i \times X^j, \tag{3}$$

where X is used to represent variables such as those in Fig. 3 and i and j are fitting parameters. The coefficients that result from fitting (3) and adjusted R^2 values are summarized in Table 1.

The best-performing relationships are obtained for the PINCii dataset, with nitrate mass and WIBS fluorescence both yielding moderate predictive skill (Fig. 6). This suggests that INP abundance in the boreal boundary layer is more closely linked to chemically complex and biologically active aerosols than to bulk particle number alone. However, the PINC measurements, collected during the snow-covered (Fig. 1) winter season, have almost no response to the predictors. While the total variability of the PINC data largely spans the PINCii data, the muted response to all predictors, mimics the Fig. 5 results, and no model seems to improve prediction beyond the mean PINC value.

4 Conclusions

Here we have attempted to deepen our understanding of the sources, abundance, and variability of INP in the boreal environment. A major objective of the HyICE-2018 campaign was to utilize high-frequency INP measurements, co-located with the over 500 time-resolved monitoring measurements at SMEAR-II, to illuminate INP characteristics in more detail. The results are mixed. Using CFDCs, we capture higher frequency INP variability, with as short as 20 minute time increments. However, the observed INP concentrations can only be qualitatively linked to other measured variables, and no single parameter emerges that is strongly linked to INP.

https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

AEROSOL Research Discussions

Table 1. Regression coefficients and adjusted R^2 values calculated for different predictors based on the simplified power law relationship (Fig. 3). The presented predictors were selected based on their consistently high qualitative scoring using the various machine learning techniques (§3.2).

Predictor (X)	**Instrument**	i	j	**Adjusted** R^2
Fluorescent particle	PINC	36.22	−0.05	−0.02
Fluorescent particle	PINCii	140.51	0.79	0.50
Particle mass	PINC	30.24	−0.10	−0.01
Particle mass	PINCii	18.50	0.86	0.43
> 500 nm number concentration	PINC	35.75	−0.21	0.01
> 500 nm number concentration	PINCii	20.88	1.13	0.39
Organic aerosol mass	PINC	34.86	0.00	−0.01
Organic aerosol mass	PINCii	27.66	0.56	0.37
Nitrate aerosol mass	PINC	36.22	−0.05	−0.02
Nitrate aerosol mass	PINCii	140.51	0.79	0.50
Black carbon mass	PINC	39.92	0.03	−0.02
Black carbon mass	PINCii	274.50	1.03	0.57

One qualitatively strong connection, is that aerosol bulk chemical composition parameters (e.g., nitrate, acetone) are consistently highly ranked predictor variables. The origin of the connection is unclear, and it may simply be a connection between total aerosol burden and INP abundance. Other, previously established connections, for example, with large particles and with biogenic particles, also exhibit highly ranked importance. However, even these features do not have the strength to suggest cause and effect.

There are distinct differences between measurements in the winter and spring/summer seasons. The winter measurements made with PINC appear to be nearly independent of any monitored variable, and rather appear to simply capture the intrinsic natural variability of INP. This potentially suggests that in southern Finland, winter INP largely originate from long range transport, and are reflective of the mixing and dilution of INP from many sources. The spring/summer measurements conducted with PINCii appear to be more linked to, and respond to, the ambient aerosol properties. This is consistent with previous work with longer-term averages (Schneider et al., 2021) that showed that in 2018 at SMEAR-II the INP concentrations increased as spring arrived and the ecosystem awoke. That said more work and longer-term high frequency measurements would be needed to close the loop and to examine whether there is consistency between seasonal trends and high resolution measurements.

Given the rarefied nature and high spatial and temporal variability of INP, one underlying conclusion of this study is that, even with vast amounts of complementary data, drawing strong conclusions that can illuminate causality will likely remain illusive. We proffer that more and longer-term studies should be undertaken at heavily-equipped complementary measurement stations. Perhaps, for example, the ACTRIS cloud in-situ (CCice) effort and instrument co-location can help the community to

https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

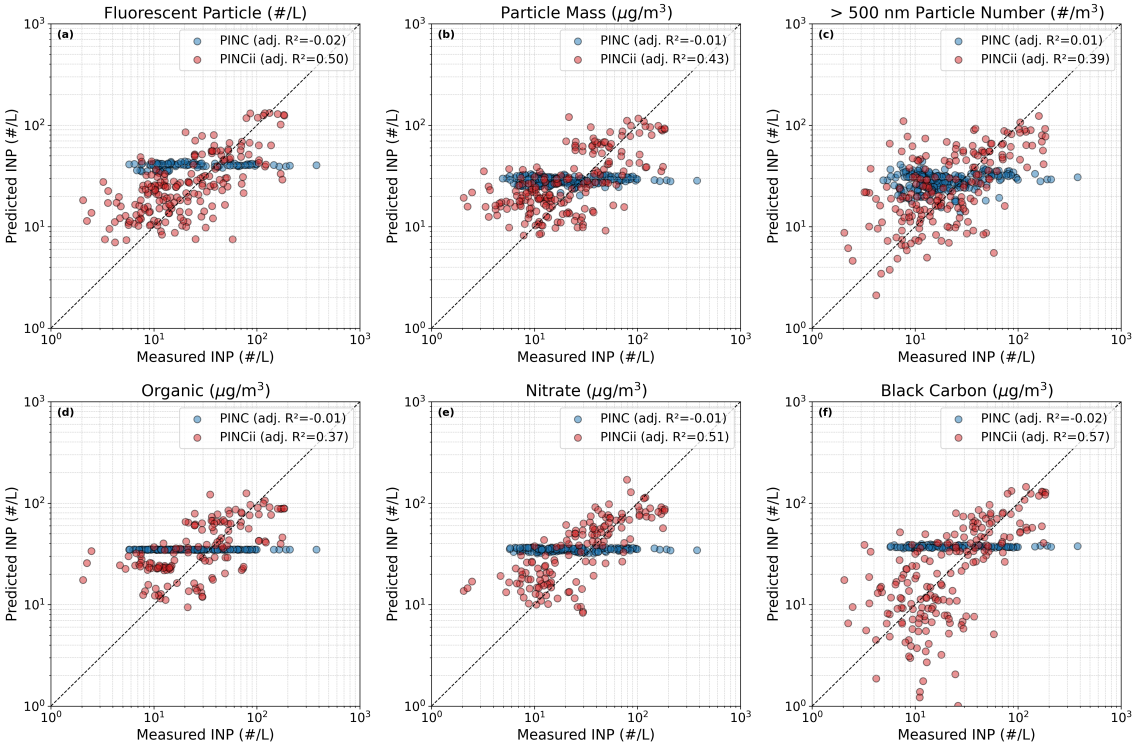

Figure 6. Comparison between measured and parameterized ice-nucleating particle (INP) concentrations for six empirical fits using (a) fluorescent particle number concentration, (b) particle mass (1 nm–10 μm), (c) number concentration of particles with diameters > 500 nm, (d) organic aerosol mass, (e) nitrate aerosol mass, and (f) black carbon mass. Each panel shows PINC (blue) and PINCii (red) data from the HyICE-2018 campaign. The 1:1 dashed lines indicate agreement between the measured INP and the simplified power law parameterizations for both data sets. The adjusted coefficients of determination (adjusted R^2) are given in Table 1. Moderate skill was obtained for the WIBS- and NO_3-based parameterizations (adjusted $R^2 \approx 0.5$ for PINCii), suggesting that biologically and chemically enriched aerosol components contribute to INP variability in the boreal atmosphere.

identify key predictor parameters. Such an effort could significantly enhance future mobile measurements, where investigators typically have limited scope and resources to choose the "important" complementary measurements.

Data availability. The aerosol, trace gas and meteorological data are available at the SmartSMEAR data repository (https://avaa.tdata.fi/web/smart). Contact with the original data contributors can be requested from atmdata@helsinki.fi. The INP data presented in this study are available at https://doi.org/10.5281/zenodo.5141574 (Brasseur et al., 2021). Other data are available upon request from the corresponding authors. The data discussed and presented from PINC are also included in (Paramonov et al., 2020) and are available at https://doi.org/10.3929/ethz-b-000397022.

https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

AEROSOL Research Discussions

Author contributions. OM, MK, TP, JD, and JK initiated and planned the HyICE-2018 campaign. ZB and JD largely coordinated and oversaw the campaign with the support of the permanent SMEAR II staff. DC, ZB, YW, JD, and EST constructed, troubleshot, and deployed the PINCii during the campaign. ZB, YW, and DC conducted measurements using the PINCii, while PH conducted measurements using the WIBS. EST and JD also participated as campaign supervisors. YW analyzed the data and prepared the figures presented here. EST, JD, and YW wrote the manuscript. All authors read and reviewed the manuscript and contributed to its improvement before and during the review process.

Competing interests. At least one of the (co-)authors is a member of the editorial board of Aerosol Research. The peer-review process was handled by an independent editor, and the authors declare no other competing interests.

Acknowledgements. The authors thank the technical staff of the Hyytiälä Forestry Field Station for their help throughout the HyICE-2018 campaign.

This project received funding from the European Union's Horizon 2020 research and innovation program under grand agreement nos. 654109 and 739530 and TransNational Access via ACTRIS-2 HyICE-2018 TNA project. The work of the University of Helsinki was supported by the Academy of Finland Centre of Excellence in Atmospheric Science (grant no. 307331) and NANOBIOMASS (307537), ACTRIS-Finland (328616), ACTRISCF (329274) and Arctic Community Resilience to Boreal Environmental change: Assessing risks from fire and disease (ACRoBEAR, 334792) Belmont Forum project. In addition, the work of the University of Helsinki was financially supported by the European Commission through ACTRIS2 (654109) and ACTRIS-IMP (871115) and ACTRIS2 TransNational Access and through integrative and Comprehensive Understanding on Polar Environments (iCUPE, 689443), ERA-NET-Cofund and by the University of Helsinki (ACTRIS-HY). EST has been supported by the Swedish Research Councils, VR (2013-05153, 2020-03497) and FORMAS (2017-00564), and the Swedish Strategic Research Area MERGE.

https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

AEROSOL Research Discussions

References

Adams, M. P., Tarn, M. D., Sanchez-Marroquin, A., Porter, G. C., O'Sullivan, D., Harrison, A. D., Cui, Z., Vergara-Temprado, J., Carotenuto, F., Holden, M. A., et al.: A major combustion aerosol event had a negligible impact on the atmospheric ice-nucleating particle population, Journal of Geophysical Research: Atmospheres, 125, e2020JD032 938, 2020.

Brasseur, Z., Castarède, D., Thomson, E. S., Adams, M. P., Drossaart van Dusseldorp, S., Heikkilä, P., Korhonen, K., Lampilahti, J., Paramonov, M., Schneider, J., and Duplissy, J.: Datasets to: Measurement Report: Introduction to the HyICE-2018 Campaign for Measurements of Ice-Nucleating Particles and Instrument Inter-Comparison in the Hyytiälä Boreal Forest, Zenodo, 2021.

Brasseur, Z., Castarède, D., Thomson, E. S., Adams, M. P., Drossaart van Dusseldorp, S., Heikkilä, P., Korhonen, K., Lampilahti, J., Paramonov, M., Schneider, J., and Duplissy, J.: Measurement Report: Introduction to the HyICE-2018 Campaign for Measurements of Ice-Nucleating Particles and Instrument Inter-Comparison in the Hyytiälä Boreal Forest, Atmospheric chemistry and physics, 22, 5117–5145, 2022.

Brasseur, Z., Schneider, J., Lampilahti, J., Vakkari, V., Sinclair, V. A., Williamson, C. J., Xavier, C., Moisseev, D., Hartmann, M., and Poutanen, P.: Vertical Distribution of Ice Nucleating Particles over the Boreal Forest of Hyytiälä, Finland, Atmospheric Chemistry and Physics, 24, 11 305–11 332, 2024.

Castarède, D., Brasseur, Z., Wu, Y., Kanji, Z. A., Hartmann, M., Ahonen, L., Bilde, M., Kulmala, M., Petäjä, T., Pettersson, J. B. C., Sierau, B., Stetzer, O., Stratmann, F., Svenningsson, B., Swietlicki, E., Thu Nguyen, Q., Duplissy, J., and Thomson, E. S.: Development and Characterization of the Portable Ice Nucleation Chamber 2 (PINCii), Atmospheric Measurement Techniques, 16, 3881–3899, https://doi.org/10.5194/amt-16-3881-2023, 2023.

Chou, C., Stetzer, O., Weingartner, E., Juranyi, 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/DOI 10.5194/acp-11-4725-2011, 2011.

Dal Maso, M., Kulmala, M., Riipinen, I., Wagner, R., Hussein, T., Aalto, P. P., and Lehtinen, K. E.: Formation and growth of fresh atmospheric aerosols: eight years of aerosol size distribution data from SMEAR II, Hyytiala, Finland, Boreal environment research, 10, 323, 2005.

DeMott, P. J.: An exploratory study of ice nucleation by soot aerosols, Journal of Applied Meteorology and Climatology, 29, 1072–1079, 1990.

DeMott, P. J., Prenni, A. J., Liu, X., Kreidenweis, S. M., Petters, M. D., Twohy, C. H., Richardson, M. S., Eidhammer, T., and Rogers, D. C.: 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.

DeMott, P. J., Prenni, A. J., McMeeking, G. R., Sullivan, R. C., Petters, M. D., Tobo, Y., Niemand, M., Möhler, O., Snider, J. R., Wang, Z., and Kreidenweis, S. M.: Integrating laboratory and field data to quantify the immersion freezing ice nucleation activity of mineral dust particles, Atmospheric Chemistry and Physics, 15, 393–409, https://doi.org/10.5194/acp-15-393-2015, 2015.

Garimella, S., Kristensen, T. B., Ignatius, K., Welti, A., Voigtländer, J., Kulkarni, G. R., Sagan, F., Kok, G. L., Dorsey, J., Nichman, L., Rothenberg, D. A., Rösch, M., Kirchgäßner, A. C. R., Ladkin, R., Wex, H., Wilson, T. W., Ladino, L. A., Abbatt, J. P. D., Stetzer, O., Lohmann, U., Stratmann, F., and Cziczo, D. J.: The SPectrometer for Ice Nuclei (SPIN): an instrument to investigate ice nucleation, Atmospheric Measurement Techniques, 9, 2781–2795, https://doi.org/10.5194/amt-9-2781-2016, 2016.

https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

Garimella, S., Rothenberg, D. A., Wolf, M. J., David, R. O., Kanji, Z. A., Wang, C., Rösch, M., and Cziczo, D. J.: Uncertainty in counting ice nucleating particles with continuous flow diffusion chambers, Atmospheric Chemistry and Physics, 17, 10 855–10 864, https://doi.org/10.5194/acp-17-10855-2017, 2017.

Hari, P. and Kulmala, M.: Station for measuring Ecosystem-Atmosphere relations (SMEAR II), Boreal Environment Research, 10, 315–322, 2005.

Kanji, Z. A., Welti, A., Chou, C., Stetzer, O., and Lohmann, U.: Laboratory studies of immersion and deposition mode ice nucleation of ozone aged mineral dust particles, Atmospheric Chemistry and Physics, 13, 9097–9118, https://doi.org/10.5194/acp-13-9097-2013, 2013.

Kanji, Z. A., Sullivan, R. C., Niemand, M., DeMott, P. J., Prenni, A. J., Chou, C., Saathoff, H., and Möhler, O.: Heterogeneous ice nucleation properties of natural desert dust particles coated with a surrogate of secondary organic aerosol, Atmospheric Chemistry and Physics, 19, 5091–5110, https://doi.org/https://doi.org/10.5194/acp-19-5091-2019, 2019.

Kulmala, M., Kontkanen, J., Junninen, H., Lehtipalo, K.and Manninen, H. E., T., N., Petaja, T., Sipila, M., Schobesberger, S., R. P., Franchin, A., Jokinen, T., Jarvinen, E., Aijala, M., Kangasluoma, J., Hakala, J., Aalto, P. P., Paasonen, P., Mikkila, J., Vanhanen, J., Aalto, J., Hakola, H., Makkonen, U., T., R., Mauldin, R. L., Duplissy, J., Vehkamaki, H., Back, J., Kortelainen, A., Riipinen, I., Kurten, T., Johnston, M. V., Smith, J. N., Ehn, M., Mentel, T. F., Lehtinen, K. E. J., Laaksonen, A., V.-M., K., and Worsnop, D. R.: Direct Observations of Atmospheric Aerosol Nucleation, Science, 339, 943–946, https://doi.org/10.1126/science.1227385, 2013.

Mahrt, F., Alpert, P. A., Dou, J., Grönquist, P., Arroyo, P. C., Ammann, M., Lohmann, U., and Kanji, Z. A.: Aging induced changes in ice nucleation activity of combustion aerosol as determined by near edge X-ray absorption fine structure (NEXAFS) spectroscopy, Environmental Science: Processes & Impacts, 22, 895–907, 2020a.

Mahrt, F., Kilchhofer, K., Marcolli, C., Grönquist, P., David, R. O., Rösch, M., Lohmann, U., and Kanji, Z. A.: The impact of cloud processing on the ice nucleation abilities of soot particles at cirrus temperatures, Journal of Geophysical Research: Atmospheres, 125, e2019JD030 922, 2020b.

Morris, C. E., Conen, F., Alex Huffman, J., Phillips, V., Pöschl, U., and Sands, D. C.: Bioprecipitation: a feedback cycle linking Earth history, ecosystem dynamics and land use through biological ice nucleators in the atmosphere, Global change biology, 20, 341–351, 2014.

Morrison, H., De Boer, G., Feingold, G., Harrington, J., Shupe, M. D., and Sulia, K.: Resilience of persistent Arctic mixed-phase clouds, Nature Geoscience, 5, 11–17, 2012.

Murray, B., O'sullivan, D., Atkinson, J., and Webb, M.: Ice nucleation by particles immersed in supercooled cloud droplets, Chemical Society Reviews, 41, 6519–6554, 2012.

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.

Ott, W. R.: A physical explanation of the lognormality of pollutant concentrations, Journal of the Air & Waste Management Association, 40, 1378–1383, 1990.

Paramonov, M., Drossaart van Dusseldorp, S., Gute, E., Abbatt, J. P. D., Heikkilä, P., Keskinen, J., Chen, X., Luoma, K., Heikkinen, L., Hao, L., Petäjä, T., and Kanji, Z. A.: Condensation/immersion mode ice-nucleating particles in a boreal environment, Atmospheric Chemistry and Physics, 20, 6687–6706, https://doi.org/10.5194/acp-20-6687-2020, 2020.

Proske, U., Adams, M. P., Porter, G. C. E., Holden, M. A., Bäck, J., and Murray, B. J.: Measurement report: The ice-nucleating activity of lichen sampled in a northern European boreal forest, Atmospheric Chemistry and Physics, 25, 979–995, https://doi.org/10.5194/acp-25-979-2025, 2025.

https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

Santos, L., Salo, K., Kong, X., Hartmann, M., Sjöblom, J., and Thomson, E.: Marine fuel regulations and engine emissions: Impacts on physicochemical properties, cloud activity and emission factors, Journal of Geophysical Research: Atmospheres, 129, e2023JD040 389, 2024.

Savage, N. J., Krentz, C. E., Könemann, T., Han, T. T., Mainelis, G., Pöhlker, C., and Huffman, J. A.: Systematic characterization and fluorescence threshold strategies for the wideband integrated bioaerosol sensor (WIBS) using size-resolved biological and interfering particles, Atmospheric Measurement Techniques, 10, 4279–4302, https://doi.org/10.5194/amt-10-4279-2017, 2017.

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., Moisseev, 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, Atmospheric Chemistry and Physics, 21, 3899–3918, https://doi.org/10.5194/acp-21-3899-2021, 2021.

Schrod, J., Thomson, E. S., Weber, D., Kossmann, J., Pöhlker, C., Saturno, J., Ditas, F., Artaxo, P., Clouard, V., Saurel, J.-M., et al.: Long-term INP measurements from four stations across the globe, Atmospheric Chemistry and Physics Discussions, 2020, 1–37, 2020.

Shupe, M. D. and Intrieri, J. M.: Cloud Radiative Forcing of the Arctic Surface: The Influence of Cloud Properties, Surface Albedo, and Solar Zenith Angle, Journal of climate, 17, 616–628, 2004.

Shupe, M. D., Persson, P. O. G., Brooks, I. M., Tjernström, M., Sedlar, J., Mauritsen, T., Sjogren, S., and Leck, C.: Cloud and Boundary Layer Interactions over the Arctic Sea Ice in Late Summer, Atmospheric Chemistry and Physics, 13, 9379–9399, https://acp.copernicus.org/articles/13/9379/2013/, 2013.

smear.avaa.csc.fi: SMEAR: Station for Measuring Ecosystem–Atmosphere Relations, https://smear.avaa.csc.fi/.

Stetzer, O., Baschek, B., Lueoeond, F., and Lohmann, U.: The Zurich Ice Nucleation Chamber (ZINC) - A new instrument to investigate atmospheric ice formation, Aerosol Science and Technology, 42, 64–74, https://doi.org/DOI 10.1080/02786820701787944, 2008.

Tang, D., Wei, T., Yuan, J., Xia, H., and Dou, X.: Observation of bioaerosol transport using wideband integrated bioaerosol sensor and coherent Doppler lidar, Atmospheric Measurement Techniques, 15, 2819–2838, https://doi.org/10.5194/amt-15-2819-2022, 2022.

Testa, B., Durdina, L., Alpert, P. A., Mahrt, F., Dreimol, C. H., Edebeli, J., Spirig, C., Decker, Z. C., Anet, J., and Kanji, Z. A.: Soot aerosols from commercial aviation engines are poor ice-nucleating particles at cirrus cloud temperatures, Atmospheric Chemistry and Physics, 24, 4537–4567, 2024.

Thomson, E., Weber, D., Bingemer, H., Tuomi, J., Ebert, M., and Pettersson, J.: Intensification of ice nucleation observed in ocean ship emissions, Scientific Reports, 8, 1111, 2018.

Tobo, Y., Prenni, A. J., DeMott, P. J., Huffman, J. A., McCluskey, C. S., Tian, G., Pöhlker, C., Pöschl, U., and Kreidenweis, S. M.: Biological aerosol particles as a key determinant of ice nuclei populations in a forest ecosystem, Journal of Geophysical Research: Atmospheres, 118, 10,100–10,110, https://doi.org/10.1002/jgrd.50801, 2013.

Tobo, Y., Adachi, K., DeMott, P. J., Hill, T. C. J., Hamilton, D. S., Mahowald, N. M., Nagatsuka, N., Ohata, S., Uetake, J., Kondo, Y., and Koike, M.: Glacially sourced dust as a potentially significant source of ice nucleating particles, Nature Geoscience, 12, 253–258, https://doi.org/10.1038/s41561-019-0314-x, 2019.

Vogel, F., Adams, M. P., Lacher, L., Foster, P. B., Porter, G. C., Bertozzi, B., Höhler, K., Schneider, J., Schorr, T., and Umo, N. S.: Ice-Nucleating Particles Active below-24° C in a Finnish Boreal Forest and Their Relationship to Bioaerosols, Atmospheric Chemistry and Physics, 24, 11 737–11 757, 2024.

Welti, A., Müller, K., Fleming, Z. L., and Stratmann, F.: Concentration and variability of ice nuclei in the subtropical maritime boundary layer, Atmospheric Chemistry and Physics, 18, 5307–5320, 2018.

https://doi.org/10.5194/ar-2026-4
Preprint. Discussion started: 29 January 2026

385 Welti, A., Bigg, E. K., DeMott, P. J., Gong, X., Hartmann, M., Harvey, M., Henning, S., Herenz, P., Hill, T. C., Hornblow, B., et al.: Ship-based measurements of ice nuclei concentrations over the Arctic, Atlantic, Pacific and Southern Ocean, Atmospheric Chemistry and Physics Discussions, 2020, 1–22, 2020.