# Exploring Ice Nucleation Particle concentrations in a Boreal Environment: limits of machine-learning-assisted variable screening

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 (CFDCs), PINC and PINCii (Portable Ice Nucleation Chambers I and II), were deployed with high-frequency sampling to measure INP concentrations. We applied machine-learning techniques to explore predictors of INP concentrations using more than 500 high-resolution atmospheric, aerosol, and ecosystem variables measured continuously at the Station for Measuring Ecosystem-Atmosphere Relations (SMEAR) II, of which 84 were retained after quality screening for the analysis. 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 aerosol 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. A core cautionary finding of this study is that, even with more than 500 co-located, high-resolution variables at one of the world's most heavily instrumented atmospheric stations, strong, deterministic links between INP concentrations and monitored parameters remain elusive. These results underscore the need for site-specific parameterizations to capture INP variability in the complex boreal environments; local biogenic and chemical proxies, such as fluorescent particle concentrations and nitrate aerosol mass, emerge as the most promising predictors for the spring and summer period.

## 1 Introduction

Clouds, especially at high-latitudes, exist in a sensitive balance. Most clouds, including those that bring precipitation to northern latitudes, exist 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). For ice crystals to form via heterogeneous nucleation, temperatures below 0°C are required; at such temperatures, ice is the thermodynamically stable phase, yet supercooled liquid water persists because a high kinetic barrier inhibits spontaneous freezing. Even with this inherently thermodynamically metastable co-existence of ice and supercooled liquid water—ice is the stable phase below 0°C, but a high kinetic barrier to nucleation sustains liquid water in the absence of efficient INPs—, 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). At high latitudes, particularly in the Arctic, mixed-phase clouds play out-sized roles in regulating climate, where amplifying feedbacks—such as the ice-albedo feedback and the water vapour–temperature amplification characteristic of polar warming—have particularly pronounced effects (Shupe and Intrieri, 2004; Morrison et al., 2012). Importantly, mixed-phase clouds primarily form with the help of small particles in the atmospheric aerosol that provide the initial seeds for ice and liquid droplet formation (ice can also enter a cloud via sedimentation from above, e.g., the seeder–feeder mechanism). Those seeds, ice nucleating particles (INPs) and cloud condensation nuclei (CCN) are fundamental to cloud formation, precipitation efficiency, and radiative properties. 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 (whether INPs are present at all) and abundance (how many are present per unit volume of air) (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 with a 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., the Integrated Carbon Observation System (ICOS; Heiskanen et al., 2022), the Aerosol, Clouds and Trace Gases Research Infrastructure (ACTRIS; Pandolfi et al., 2018), and the Swedish Infrastructure for Ecosystem Science (SITES; Swedish Infrastructure for Ecosystem Science 2021) 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 extending the campaign (Schneider et al., 2021) have been published. Key findings include the characterisation of condensation/immersion mode INP concentrations and their likely origin from long-range transport (Paramonov et al., 2020), the identification of seasonal trends linked to biogenic emissions (Schneider et al., 2021), instrument intercomparison results (Brasseur et al., 2022), the vertical distribution of INPs over the boreal forest (Brasseur et al., 2024), and the role of fluorescent biological aerosol particles as INP tracers (Vogel et al., 2024). 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 as importance-ranking and pattern-exploration tools to investigate the emergence of non-obvious (or intuitive) connections within the available high-frequency data. When many variables are screened simultaneously, the probability of finding spurious correlations increases; the results presented here should therefore be interpreted as hypothesis-generating rather than hypothesis-confirming, and the identified associations require independent validation.

Previous studies have linked INP concentrations to a range of aerosol properties, including total particle number concentration above 0.5 $\mu$m (DeMott et al., 2010; Tobo et al., 2013), fluorescent biological aerosol particle concentrations (Tobo et al., 2013), mineral dust loading (DeMott et al., 2015), and black carbon (DeMott, 1990). At the same site, Paramonov et al. (2020) found that no single parameter predicted INP concentrations over the full campaign period, although short-timescale correlations with black carbon, supermicron biological particles, and sub-0.1 $\mu$m particles were observed. These precedents motivate the present study's open-ended screening approach while providing the physical context for interpreting the resulting variable rankings.

The results of our study are educational but cautionary. Strong links between INP concentrations and fundamental chemical signatures of the atmospheric aerosol do exist, but are likely open to over-interpretation. Especially 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.

### 2.2 Complementary Data and Machine Learning

Although INP concentration 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 INP concentration.

Detailed instrument specifications for the HyICE-2018 campaign instruments are provided in Brasseur et al. (2022); metadata for the full SMEAR II monitoring suite are available via the SmartSMEAR portal (Junninen et al., 2009).

The machine learning analysis employed random forest and decision tree models to derive feature importance rankings, and principal component analysis (PCA) together with K-means clustering for dimensionality reduction and pattern exploration. These algorithms were applied as exploratory importance-ranking tools; no train/test data splitting or cross-validation was performed, as the objective was hypothesis generation rather than predictive modelling. Prior to analysis, all variables were harmonised to a common 20-minute time base; variables measured at coarser resolution were assigned to the nearest 20-minute interval, and those at finer resolution were averaged. The analyses were implemented in R: random forests with the `randomForest` package, decision trees with `rpart`, and PCA and K-means with the base functions `prcomp` and `kmeans`. Random-seed sensitivity is quantified by running the random forest under 50 independent seeds; the seed-resolved importance scores are archived with the analysis code (see Code availability).

In addition to the INP measurements, for the purpose 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).

It is important to note that these machine learning methods were applied here as importance-ranking and exploratory tools, not to build or evaluate a predictive model; no train/test data splits were applied.

### 2.3 INP concentration measurements

The following instrument specifications are provided to establish the conditions under which INP concentrations were measured (particularly the lamina temperature and humidity, which define the INP activation threshold) and to facilitate comparison with other CFDC-based INP studies.

Specifically, we test whether new particle formation (NPF) events generate particles that grow into the INP-relevant size range ($>0.5$ $\mu$m); whether primary biological aerosol (proxied by fluorescent particle counts from the WIBS) is a dominant INP source in this environment; and whether aerosol chemical markers—including black carbon, nitrate, and organic aerosol mass—can serve as practical INP proxies.

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 have been directly compared 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 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 $T_l = -31°\text{C}$ and a relative humidity with respect to water $\text{RH}_w = 105\%$. 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 $\mu$m. 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 $T_l = -32°\text{C}$ and $\text{RH}_w = 105\%$ and a 2.5 $\mu$m cyclone impactor.

Both PINC and PINCii operate with the same measurement cycle: a 5-minute background (particle-free) period followed by a 15-minute ambient sampling period, cycling continuously. Each INP concentration value is derived from the ambient-period particle count minus the interpolated background, yielding one data point every 20 minutes. At a sample flow rate of 1 L min$^{-1}$, the 15 L sample volume per data point is sufficient for statistically meaningful INP counts even at the low concentrations observed during winter ($<$1 L$^{-1}$). Instrumental constraints and availability made it impossible for PINC and PINCii to operate simultaneously during HyICE-2018. However, their operation in succession means that for the bulk of the campaign there exists high-frequency data coverage, which appears self-consistent. An instrument comparison summary is available within the campaign Measurement Report (Brasseur et al., 2022), where the inconsistencies with another parallel-plate CFDC chamber, SPectrometer for Ice Nuclei (SPIN), make plain its absence here. The comparison showed that PINC and PINCii agree within a factor of two during the overlapping measurement period, with both instruments capturing the same general trends in INP concentrations; minor differences in absolute concentration are attributable to the 1 K difference in operating temperature ($T_l = -31°\text{C}$ for PINC, $-32°\text{C}$ for PINCii) and instrument-specific detection efficiencies.

## 3 Results and Discussion

### 3.1 INP and other monitored variables

The time series of INP concentrations measured by PINC and PINCii each with a 15-minute ambient sampling window (one data point every 20 minutes) (Brasseur et al., 2022) is presented in Fig. 1(a). INP concentrations were measured at $T_l = -31°C$ (PINC) and $T_l = -32°C$ (PINCii). Each point represents a 15-minute ambient sampling period (with 5-minute background periods before and after, yielding one data point every 20 minutes) 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 of Fig. 1(a), the PINCii spring and early summer time series appears somewhat more spread on the logarithmic concentration axis than the PINC winter series, consistent with seasonal changes in aerosol sources and meteorology; we do not report distribution fits or geometric moments here because PINC and PINCii did not operate simultaneously, so we restrict the comparison to this qualitative description.

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 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). New particle formation events are relevant here because freshly nucleated particles can grow into the INP-relevant size range during sustained growth episodes, potentially contributing to the INP-active aerosol population.

### 3.2 Machine Learning

The seasonal progression visible in Fig. 1—the transition in temperature and snow depth, the spring/summer increase in fluorescent biological particles, the shift in aerosol chemical composition (increasing organic fraction, panel e), and the apparent co-variability of INP with these tracers—motivates 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 machine-supported analysis techniques, including random forest models

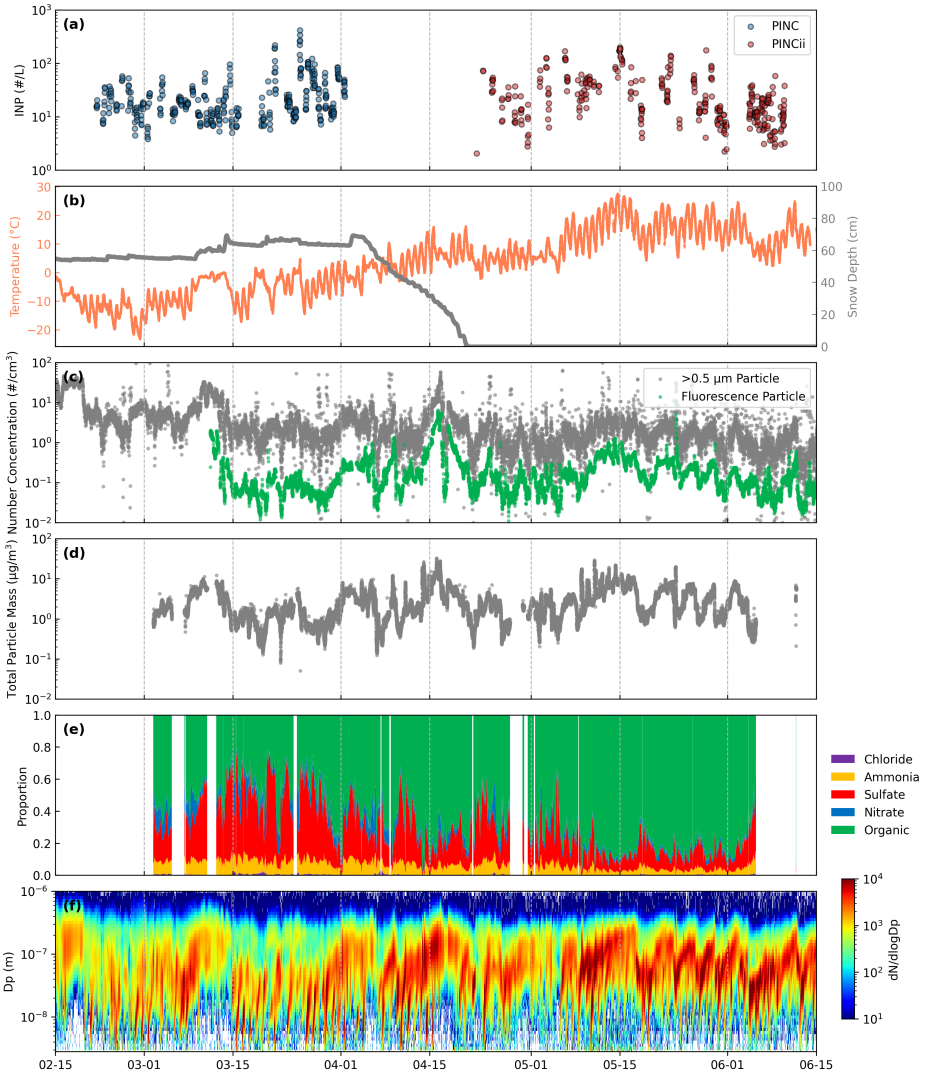


**Figure 1.** Time series of atmospheric, aerosol, and INP properties measured during the HyICE-2018 campaign at the SMEAR II station in Hyytiälä, Finland (19 February–10 June 2018). (a) INP concentrations measured by the Portable Ice Nucleation Chamber (PINC; circles) and its updated version, PINCii (circles, differentiated by colour~~squares~~), shown on a logarithmic scale. INP concentrations were measured at $T_l = -31°$C (PINC; blue) and $T_l = -32°$C (PINCii; red). (b) Air temperature (coral, left axis) and snow depth (grey, right axis), illustrating the transition from winter through spring to summer. (c) Number Concentration (#/cm$^3$) of particles with diameters >0.5 $\mu$m (grey) and fluorescent biological aerosol particles (>0.5 $\mu$m, green); WIBS fluorescence data begin on 11 March 2018, when the instrument was first deployed. (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 (colours follow AMS convention): 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.

(Fig. 2), along with pairwise correlation, decision tree, principal component, and K-means clustering analyses. While it was difficult to extract quantitative results in all cases, to quantify the random-seed sensitivity of the random-forest feature importance we ran the analysis with 50 independent random seeds (1 . . . 50). Figure 2 reports the per-variable median importance with 5–95$^{th}$ percentile whiskers across seeds; the top-ranked variables discussed below remain in the top group across all seeds. ~~each separate treatment resulted in qualitatively similar results, which are illustrated in Fig. 2.~~ A formal multi-method comparison (e.g., side-by-side feature rankings from random forest, decision tree, and PCA) is not shown but was verified by the authors during analysis; these approaches consistently highlighted similar top-ranking variables.

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 INP concentration. For example, biological particles that yield a fluorescence signal serve as a tracer for primary biological aerosol, which may include INP-active species, and are among the most studied INP types in forest environments (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. Several other highly ranked variables (e.g., acetone, methanol concentrations) are not examined individually because they largely co-vary with the selected predictors (particularly nitrate) and thus provide redundant information for the purposes of this exploratory analysis. A systematic investigation of all high-ranking but unexpected variables is an important avenue for future work.

In Fig. 3 six variables are selected to illustrate their pairwise correlation with INP. The six variables were selected on two complementary grounds: (i) high importance rank in the random forest analysis (fluorescent particles, black carbon, nitrate, particle mass); and (ii) established physical or empirical connection to INP in prior literature or at this specific site ($>$0.5 $\mu$m concentration per DeMott et al. 2010; Tobo et al. 2013; organic aerosol mass per the biogenic seasonality documented by Schneider et al. 2021). 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. Indeed, nitrate and acetone concentrations correlate positively with total aerosol mass during the PINCii spring/summer period ($r \approx 0.5$–$0.6$), suggesting that their association with INP may partly reflect the total aerosol burden rather than a specific chemical mechanism. 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). Notably, Paramonov et al. (2020) also reported a positive correlation between BC and INP concentrations at short timescales during the HyICE-2018 campaign at the same site (their Fig. 5), independently supporting this finding. Possible mechanisms include enhancement of BC's ice-nucleating ability through atmospheric aging, oxidation, and coating with organic material (DeMott, 1990; Mahrt et al., 2020b, a). 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.

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 (fluorescent biological aerosol particles from the WIBS; total particle mass from the AMS; $>$0.5 $\mu$m number concentration from the APS). In addition to fluorescent biological aerosol particles (WIBS), particle number concentrations for particles larger than 0.5 $\mu$m (APS) are plotted. Particle concentrations of this size appear in several,

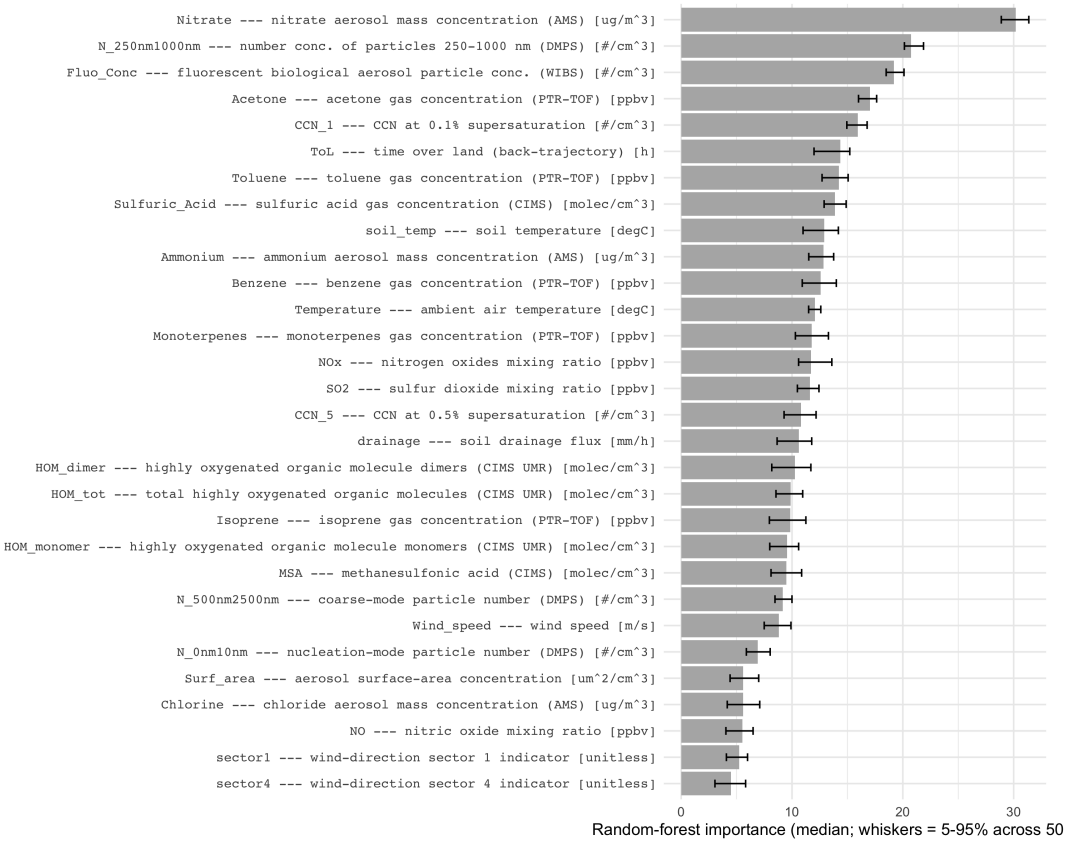


**Figure 2.** Feature importance analysis derived from a Random Forest model trained to predict ice-nucleating particle (INP) concentrations during the HyICE-2018 campaign at SMEAR II. Variables are shown in decreasing order of random-forest importance. The six highest-ranked variables in this panel are nitrate aerosol mass concentration (AMS), 250–1000 nm particle number concentration (DMPS), fluorescent biological aerosol particle concentration (WIBS), acetone gas-phase concentration (PTR-TOF), CCN concentration at 0.1% supersaturation, and the back-trajectory-derived time over land (ToL). Fluorescent particle concentration and nitrate aerosol mass are subsequently examined in pairwise correlation in Fig. 3 alongside particle mass (AMS), $>0.5$ $\mu$m number concentration (APS), organic mass (AMS), and black carbon mass (aethalometer)—variables drawn from a separate aerosol-only feature set and therefore not part of the random-forest input shown here (selection criteria in Sect. 2.2).~~The six variables examined in pairwise correlation in Fig. 3—fluorescent particle number concentration (WIBS), particle mass (1 nm–10 $\mu$m aerodynamic diameter, AMS), number concentration of particles with diameters $>0.5$ $\mu$m (APS), organic aerosol mass concentration (AMS), nitrate aerosol mass concentration (AMS), and black carbon mass concentration (aethalometer)—appear among the top-ranked variables in this panel (selection criteria in Sect. 2.2).~~ The bars show the relative importance of each variable, expressed in arbitrary units. Higher values indicate 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). Only the top-ranked variables (by median random-forest importance across 50 seeds) are shown for legibility; the full ranked list and per-seed importance scores are archived with the analysis code (see Code availability). Short names on the y-axis are SMEAR-II identifiers paired with their spelled-out long names and units; see `https://smear.avaa.csc.fi` for the full SMEAR variable registry.

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)) (all three from the AMS except BC, which is from the aethalometer), 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. Data for PINC and PINCii are shown separately because the two instruments operated in different seasons (winter vs. spring/summer) with distinct ambient aerosol and INP characteristics; combining them would conflate seasonal differences and obscure the key finding: the observed correlations are specific to the PINCii spring/summer subset and do not hold generally across the full campaign (the PINC winter data show no meaningful correlation with any of the examined predictors). The Pearson correlation coefficients were calculated from the raw (untransformed) data; the log–log axes are used for visualisation only to accommodate the wide dynamic range of the measurements. Using 20-minute data instead of hourly means yields qualitatively similar results, suggesting that the dominant co-varying processes operate on timescales exceeding 20 minutes (diurnal or synoptic scale) rather than at sub-hourly variability. This subset-specific character of the correlations has broader implications for campaign-based INP correlations reported in the literature: INP–aerosol correlations derived from short, season-specific field campaigns are conditioned on the ambient regime of the measurement period and should be interpreted with caution when generalised to other seasons, sites, or ecosystems.

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 identify statistical associations between predictor variables and INP concentrations but cannot establish causal relationships. 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 contain INPs. The correlations likely reflect co-varying air-mass properties rather than direct INP composition: INPs are a vanishingly small fraction of the total aerosol ($\sim$1 in $10^6$), and without single-particle analysis of ice crystal residuals, the identity of the actual ice-nucleating species remains undetermined (Paramonov et al., 2020). 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. Log-normal concentration distributions arise naturally from the multiplicative dilution and mixing of aerosol during atmospheric transport (Ott, 1990), consistent with INPs at this site originating from diverse, distant sources. This interpretation aligns with the findings of Paramonov et al. (2020), who hypothesised long-range transport as the dominant INP source at SMEAR II based on the same PINC dataset. Lagrangian back trajectories or dispersion models (e.g., HYbrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT; Stein et al. 2015) and the FLEXible PARTicle dispersion model (FLEXPART; Pisso et al. 2019)~~HYSPLIT, FLEXPART~~) can contextualise air-mass history and are often used together with chemical tracers or receptor modelling to assess long-range transport, but they do not by themselves uniquely verify INP provenance or the dilution interpretation above:

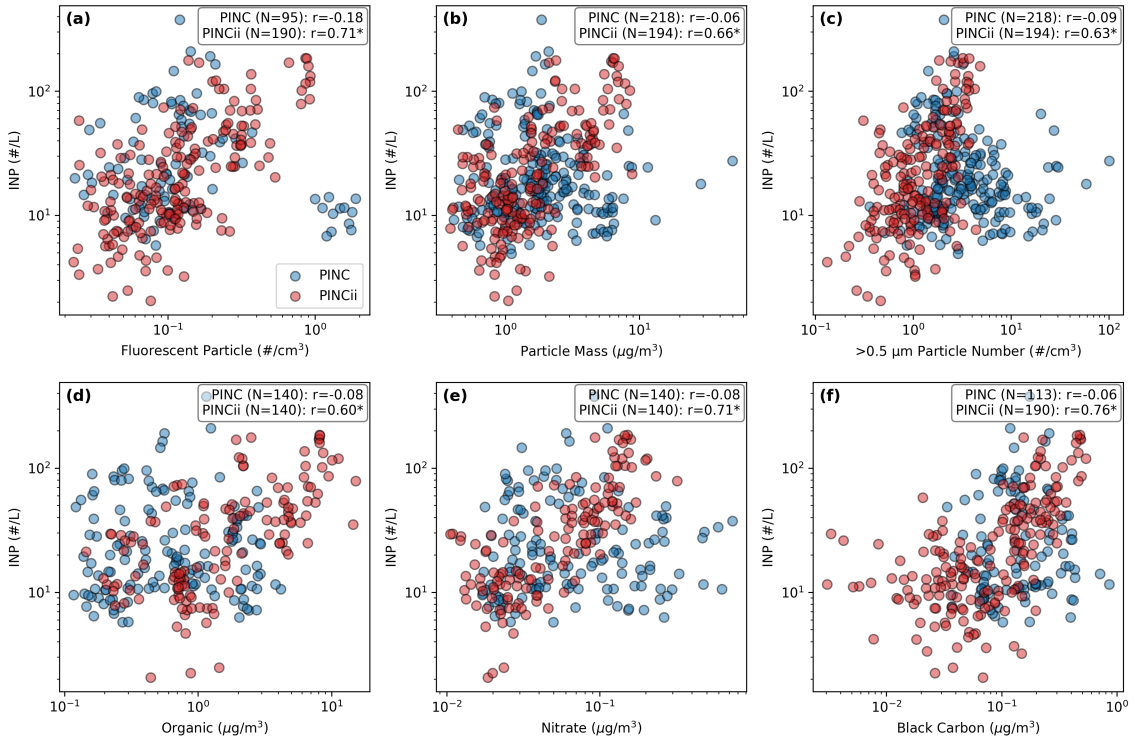


**Figure 3.** Log–log scatter plots showing correlations between ice-nucleating particle (INP) concentrations measured by PINC (blue) and PINCii (red) during HyICE-2018 and six selected aerosol parameters measured at SMEAR II, Hyytiälä, Finland: (a) fluorescent particle number concentration (WIBS), (b) particle mass (1 nm–10 $\mu$m aerodynamic diameter, AMS), (c) number concentration of particles with diameters >0.5 $\mu$m (APS), (d) organic aerosol mass concentration (AMS), (e) nitrate aerosol mass concentration (AMS), and (f) black carbon mass concentration (aethalometer). Asterisks (*) denote statistically significant correlations ($p < 0.05$). Pearson coefficients were calculated from the raw (untransformed) data; the log–log axes are for visualisation only. Hourly means, computed for each full hour when both variables were above zero, were used to align with the coarser temporal resolution of some complementary datasets; while using 20-minute data yielded qualitatively similar results (see text). $N$ values in each panel indicate the number of PINC and PINCii data points, respectively. 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. PINC data (blue) correspond to winter measurements (19 February–2 April); PINCii data (red) to spring/summer (22 April–10 June).

trajectory errors, mixing along the path, and the lack of INP-specific tagging in kinematic histories mean that such analyses complement—rather than uniquely test—inferences from local correlations and distribution shape. We highlight combined trajectory–tracer or trajectory–receptor studies as a priority for future work. 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 $\chi^2$ statistics (using raw bin counts; see caption), the formal tests reject the simple unimodal log-normal null at conventional significance for both PINCii ($\chi^2 = 81.98$, $p = 0.001$) and PINC ($\chi^2 = 157.19$, $p < 0.001$). Nevertheless, the histograms are

broadly similar in shape and scale, so log-normality remains a useful approximate description of the bulk distribution. A mod-
265 est elevated tail in the PINCii histogram motivates the bimodal decomposition in Fig. 4(b): for PINCii, $\chi^2$ decreases to $63.54$ ($p = 0.028$), whereas for PINC the bimodal fit does not improve overall agreement ($\chi^2 = 242.72$, $p < 0.001$). We therefore do not interpret the PINCii tail as evidence for a statistically robust separate population at the 5% level, and we caution that the $\chi^2$ values indicate residual structure beyond ideal unimodal log-normal behaviour for both instruments.

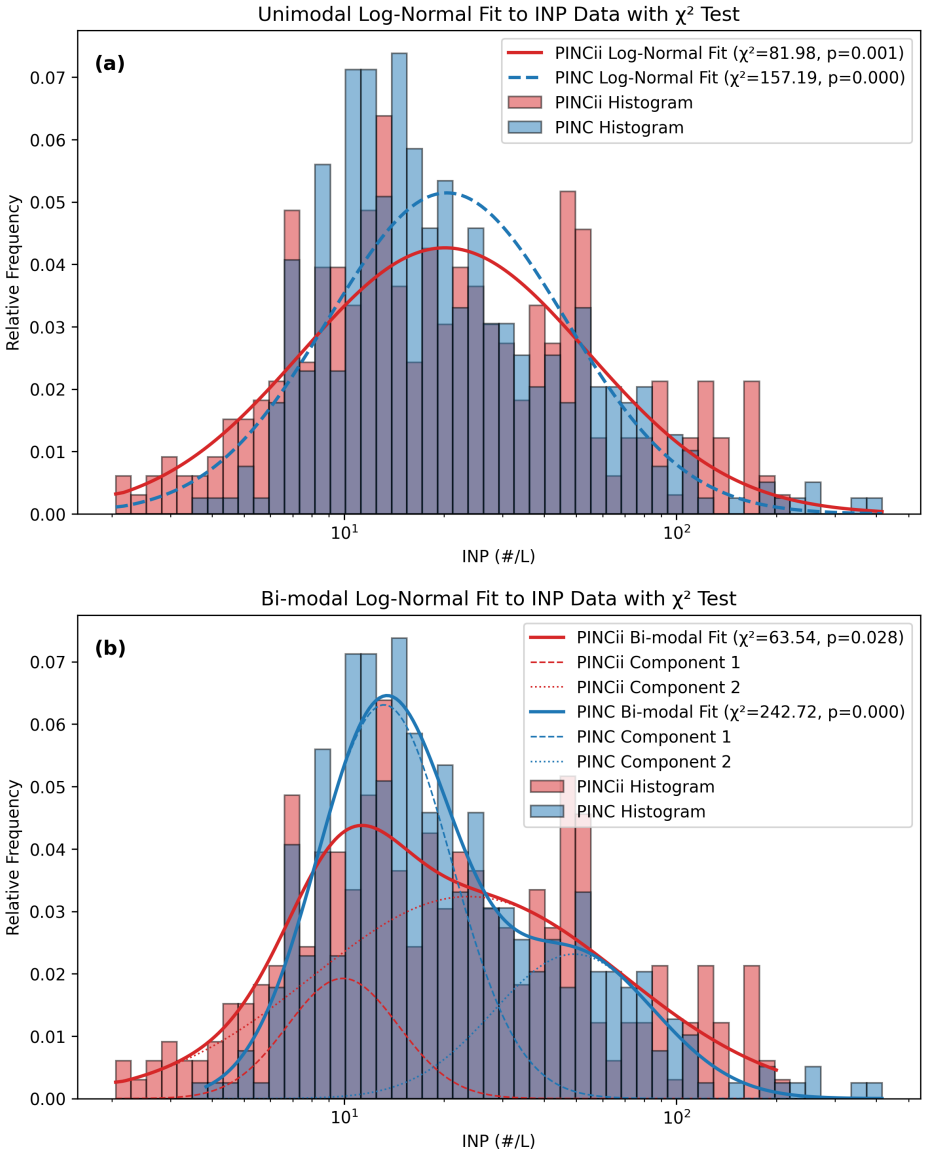


**Figure 4.** Normalized relative frequency distributions for INP concentrations measured by PINC and PINCii during the HyICE-2018 campaign, along with fitted log-normal probability density functions. (a) Unimodal log-normal fits to each instrument's INP histogram, with fit parameters evaluated using maximum likelihood estimation and goodness-of-fit assessed via $\chi^2$ statistics. (b) Bimodal log-normal fits decomposed into two component modes, illustrating potential multi-population structure in the INP size/activation spectrum. The $\chi^2$ goodness-of-fit test uses raw bin counts as observed frequencies, with expected counts derived by integrating the fitted probability density over each bin and scaling by total sample size $N$; degrees of freedom are adjusted for the number of estimated parameters ($\mathrm{ddof} = 2$ for unimodal; $\mathrm{ddof} = 5$ for bimodal fits). The unimodal fits yield $\chi^2 = 81.98$ ($p = 0.001$) for PINCii and $\chi^2 = 157.19$ ($p < 0.001$) for PINC. The bimodal fits give $\chi^2 = 63.54$ ($p = 0.028$) for PINCii and $\chi^2 = 242.72$ ($p < 0.001$) for PINC; $\chi^2$ decreases for PINCii relative to the unimodal case, but the test still rejects the bimodal fit at the 5% level ($p = 0.028$), whereas the PINC bimodal model is not supported as an improvement over its unimodal counterpart.

### 3.3 Parameterization

If the dominant INP signal originates from long-range transport, local parameterisations based on co-located measurements will have limited predictive power—a prediction directly borne out by the PINC winter results below. The modest success of the PINCii spring/summer parameterisations may then reflect a local biogenic INP contribution that augments the transported background during the growing season.

The limited correlations identified in Sect. 3.2 suggest that existing parameterisations—which rely on variables such as $n_{\mathrm{AP}>0.5\,\mu\mathrm{m}}$ that do not strongly correlate with INP in this dataset—are unlikely to reproduce the observed INP variability. Testing them nonetheless serves two purposes: (i) it quantifies the parameterisation failure against a comprehensive boreal dataset, and (ii) it motivates the exploration of alternative local proxies in Sect. 3.3.2.

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

Although episodic long-range transport of high-latitude dust to Finland cannot be excluded a priori, we cross-referenced the PINC measurement window (19 February–2 April 2018) against the systematic Finnish Meteorological Institute catalog of long-range dust transport events to Finland from 1980 to 2022 (86 documented events: 59 Saharan, 22 Aral-Caspian, 5 Middle East; Meinander et al., 2023). No catalog event overlaps our campaign window. The nearest documented Saharan dust depositions over Finland are dated to 8–10 March 1991 (northern Finland) and 21–23 February 2021 (southern Finland); neither falls within our measurement period. Although the FMI catalog does not explicitly enumerate Icelandic or Greenlandic glacial-outwash dust events, no major high-latitude glacial-dust outbreak affecting Northern Europe in spring 2018 has been reported in the peer-reviewed literature. We therefore find no documented evidence that long-range dust transport materially contributed to the winter INP signal observed by PINC, supporting our choice to compare measurements against the dust-independent DeMott et al. (2010) formulation rather than the dust-aware DeMott et al. (2015) revision.

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 $\mu$m ($n_{AP>0.5\mu m}$), 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, where no local

mineral dust sources exist, although episodic long-range transport of high-latitude glacially sourced dust (e.g., Tobo et al. 2019; Sanchez-Marroquin et al. 2020) cannot be excluded, we refer back to the 2010 formulation, which is:

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

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 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. Note that both Eqs. (1) and (2) use the total aerosol number concentration $n_{\mathrm{AP}>0.5\,\mu\mathrm{m}}$ as the predictor. Tobo et al. (2013) also provide a separate parameterisation based on fluorescent biological aerosol particle (FBAP) concentrations; that variant is not applied here. 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.

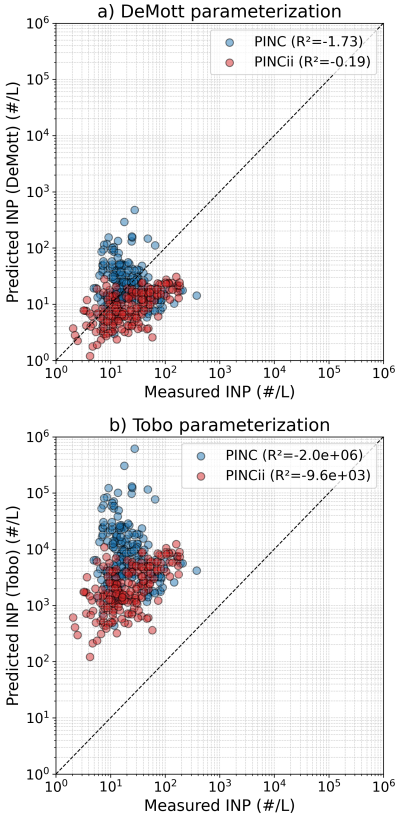


**Figure 5.** Predicted INP concentrations from (a) DeMott et al. (2010) and (b) Tobo et al. (2013) parameterizations plotted versus the INP concentrations measured using the PINC(blue) and PINCii(red) CFDCs. The 1:1 line is plotted and $R^2$ values corresponding to the coefficient(s) of determination 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.

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). Specifically, the DeMott et al. (2010) underprediction is consistent with the intercomparison findings of Brasseur et al. (2022), whereas the poor performance of Tobo et al. (2013) over the full campaign contrasts with its relatively better performance during their four targeted intercomparison days.. The better performance noted by Brasseur et al. (2022) (their Fig. 8 time series, and the consolidated 28 March inter-instrument spectrum summary in their Fig. 11) reflects a focused comparison over four targeted intercomparison days that are not representative of the full seasonal range covered here; Brasseur et al. (2022) themselves caution that their comparison

"might not be representative of the entire HyICE-2018 campaign." Direct comparison between CFDC-based and bulk immersion freezing assay results is not straightforward because CFDCs primarily activate condensation and deposition freezing on a timescale of seconds, whereas bulk immersion assays (such as INSEKT, used to derive the Schneider et al. 2021 and Brasseur et al. 2024 parameterisations) probe a different nucleation pathway over longer timescales.

### 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 $n_{\mathrm{INP}}$ 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 exponent $j$ reflects the sensitivity of INP concentration to changes in the predictor: values near unity indicate approximately linear relationships (>0.5 $\mu$m number: $j = 1.13$; BC mass: $j = 1.03$), values well below unity suggest weak sensitivity (organic mass: $j = 0.56$), and near-zero or negative values indicate absence of a meaningful relationship (all PINC predictors). The pre-factor $i$ sets the absolute scale and is influenced by the ambient INP concentration range during the respective measurement period. The consistently low or negative adjusted $R^2$ for PINC confirms the absence of predictive skill during the winter period, regardless of the predictor chosen.

The persistent overprediction by Tobo et al. (2013) for PINCii reflects two site-specific factors: (i) at Hyytiälä, frequent NPF events grow secondary organic aerosol into the >0.5 $\mu$m size range (Dal Maso et al., 2005; Kulmala et al., 2013), inflating $n_{\mathrm{AP}>0.5\,\mu\mathrm{m}}$ without a commensurate increase in biological INPs, whereas the parameterisation was calibrated at a North American temperate forest where large particles are predominantly primary biological aerosol; and (ii) the sub-Arctic boreal spring supports lower primary biological aerosol emission than the temperate calibration environment, further reducing the fraction of large particles that are ice-active at these temperatures.

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, yield near-constant predicted INP concentrations because the fitted exponents ($j$ in Eq. 3) are near zero for all PINC predictors (Table 1), meaning the predicted values collapse to approximately the pre-factor $i$ regardless of predictor variability. 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.

**Table 1.** Regression coefficients and adjusted $R^2$ values calculated for different predictors based on the simplified power law relationship (Eq. 3). The presented predictors were selected based on their high importance ranking in the random forest analysis and established physical or empirical connections to INP in prior literature (see Sect. 3.2) (§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 |
| $>0.5$ $\mu$m number concentration | PINC | 35.75 | $-0.21$ | 0.01 |
| $>0.5$ $\mu$m 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 |

## 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 show statistically detectable but weak associations with other measured variables, insufficient for reliable prediction, and no single parameter emerges that is strongly linked to INP.

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 reflect INP concentrations driven by processes external to the measurement site—most plausibly long-range transport from diverse, distant sources, as indicated by the log-normal concentration distribution (Fig. 4) and the absence of local predictor correlations. 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

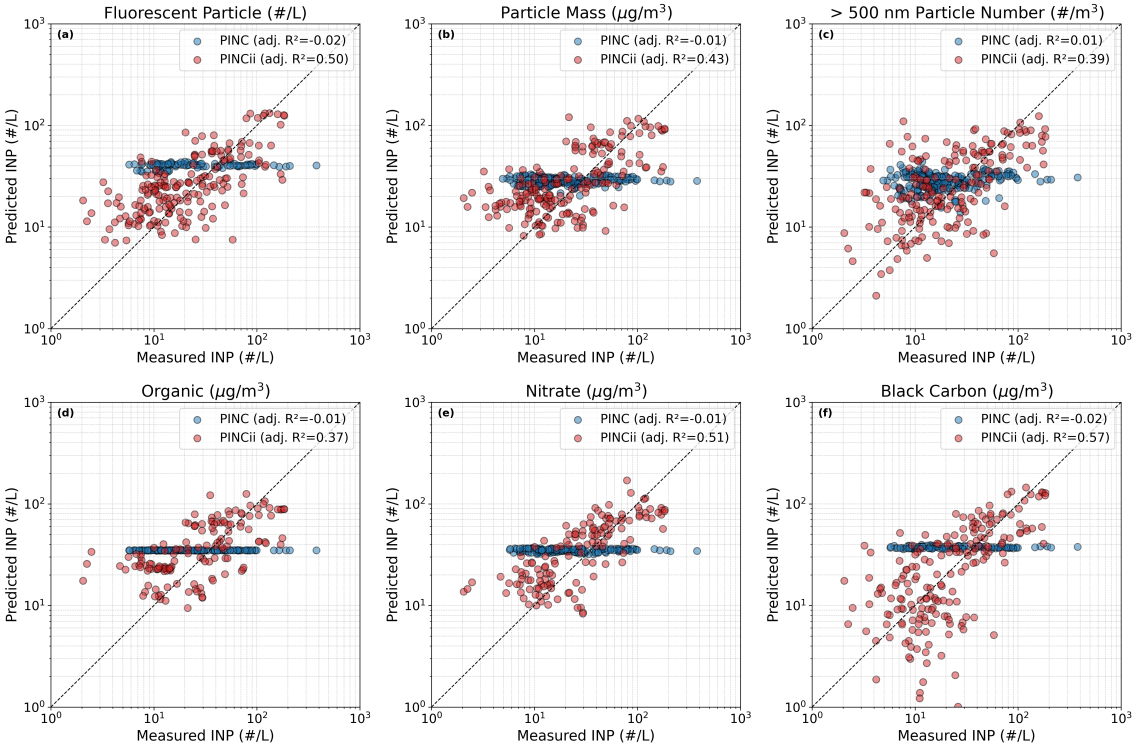


**Figure 6.** Comparison between measured and parameterized ice-nucleating particle (INP) concentrations for six empirical fits using (a) fluorescent particle number concentration, (b) particle mass (1 nm–10 $\mu$m), (c) number concentration of particles with diameters $>$0.5 $\mu$m, (d) organic aerosol 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 during spring and early summer.

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.

The weak correlations observed in this study, combined with the log-normal INP concentration distributions consistent with random dilution of trace species, suggest that INP at this boreal site are dominated by long-range transport from diverse, distant sources. In such a regime, no local measurement suite—however comprehensive—can be expected to yield strong, causal predictor–INP relationships, because the INP identity and source vary stochastically with air-mass origin. This interpretation is independently supported by Paramonov et al. (2020), who reached a similar conclusion using the same PINC dataset at the same site. We therefore propose that future campaigns should complement high-frequency INP measurements

with source-apportionment tools (e.g., back-trajectory analysis, receptor modelling) to evaluate consistency with the long-range transport hypothesis and to constrain the geographic origins and aerosol types that plausibly contribute to the INP population—recognising that trajectory-based tools alone cannot uniquely verify INP provenance without tracer or receptor constraints. The question "why can't we predict INP from local measurements?" is itself a scientifically valuable finding: it constrains the problem space and guides the field toward approaches that account for air-mass history rather than relying solely on local co-located tracers.

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. Rather than simply calling for more of the same local measurements, we suggest that future long-term studies at heavily-equipped complementary stations—for example, the ACTRIS cloud in-situ (CCice) effort and the instrument co-location it enables—should pair high-frequency INP measurements with source-apportionment tools (back-trajectory analysis, receptor modelling, and, where available, single-particle ice-residual analysis), so that the air-mass history behind each measurement is explicitly accounted for. Such an effort could significantly enhance future mobile measurements, where investigators typically have limited scope and resources to choose the "important" complementary measurements.

*Code availability.* The full R analysis code used in this study—comprising the multi-seed random-forest sensitivity sweep, the dust-episode screen, the figure-generation script for Fig. 2, and the parent analysis notebook—is archived at https://doi.org/10.5281/zenodo.20367476 together with the analysis-ready data. The R version and full list of package versions used to generate the results are captured in the archived session-info file; the analysis was performed in R 4.6.0 with the `randomForest`, `rpart`, `ggplot2`, `dplyr`, and `readr` packages, among others.

*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). Derived analysis-ready datasets (the harmonised 20-minute feature matrix used as input to the random-forest analysis, the multi-seed random-forest importance table, and the dust-episode screen output) are deposited together with the analysis scripts at https://doi.org/10.5281/zenodo.20367476 (see Code availability).~~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.

*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.

## 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., Vogel, F., Wu, Y., Abbatt, J. P. D., Atanasova, N. S., Bamford, D. H., Bertozzi, B., Boyer, M., Brus, D., Daily, M. I., Fösig, R., Gute, E., Harrison, A. D., Hietala, P., Höhler, K., Kanji, Z. A., Keskinen, J., Lacher, L., Lampimäki, M., Levula, J., Manninen, A., Nadolny, J., Peltola, M., Porter, G. C. E., Poutanen, P., Proske, U., Schorr, T., Silas Umo, N., Stenszky, J., Virtanen, A., Moisseev, D., Kulmala, M., Murray, B. J., Petäjä, T., Möhler, O., 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, https://doi.org/10.5194/acp-22-5117-2022, 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.

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.

Heiskanen, J., Brümmer, C., Buchmann, N., Calfapietra, C., Chen, H., Gielen, B., Gkritzalis, T., Hammer, S., Hartman, S., Herbst, M., et al.: The integrated carbon observation system in Europe, Bulletin of the American Meteorological Society, 103, E855–E872, 2022.

Junninen, H., Lauri, A., Keronen, P., Aalto, P., Hiltunen, V., Hari, P., and Kulmala, M.: Smart-SMEAR: on-line data exploration and visualization tool for SMEAR stations, Boreal Environment Research, 14, 447, 2009.

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.

Meinander, O., Kouznetsov, R., Uppstu, A., Sofiev, M., Kaakinen, A., Salminen, J., Rontu, L., Welti, A., Francis, D., Piedehierro, A. A., Heikkilä, P., Heikkinen, E., and Laaksonen, A.: African dust transport and deposition modelling verified through a citizen science campaign in Finland, Scientific Reports, 13, 21 379, https://doi.org/10.1038/s41598-023-46321-7, 2023.

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.

Pandolfi, M., Alados-Arboledas, L., Alastuey, A., Andrade, M., Angelov, C., Artiñano, B., Backman, J., Baltensperger, U., Bonasoni, P., Bukowiecki, N., et al.: A European aerosol phenomenology–6: scattering properties of atmospheric aerosol particles from 28 ACTRIS sites, Atmospheric Chemistry and Physics, 18, 7877–7911, 2018.

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.

Pisso, I., Sollum, E., Grythe, H., Kristiansen, N. I., Cassiani, M., Eckhardt, S., Arnold, D., Morton, D., Thompson, R. L., Groot Zwaaftink, C. D., Evangeliou, N., Sodemann, H., Haimberger, L., Henne, S., Brunner, D., Burkhart, J. F., Fouilloux, A., Brioude, J., Philipp, A., Seibert, P., and Stohl, A.: The Lagrangian particle dispersion model FLEXPART version 10.4, Geoscientific Model Development, 12, 4955–4997, https://doi.org/10.5194/gmd-12-4955-2019, 2019.

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.

Sanchez-Marroquin, A., Arnalds, O., Baustian-Dorsi, K., Browse, J., Dagsson-Waldhauserova, P., Harrison, A., Maters, E., Pringle, K., Vergara-Temprado, J., Burke, I., et al.: Iceland is an episodic source of atmospheric ice-nucleating particles relevant for mixed-phase clouds, Science advances, 6, eaba8137, 2020.

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/.

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

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.

Swedish Infrastructure for Ecosystem Science: SITES: Swedish Infrastructure for Ecosystem Science, https://www.fieldsites.se/, 2021.

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.

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.