# Author Reply to Reviewer Comments

RC1 Reply

Dear Referee,

We thank you for the thorough, detailed, and constructive review. Your comments have helped us substantially improve the clarity and scientific framing of the manuscript. We address each comment below in the order presented.

---

## General comment

### Reviewer comment

The manuscript reports on the feasibility of predicting INP concentrations measured at −31°C and −32°C during the HyICE-2018 campaign in the boreal forest from 84 complementary variables measured at the site. The paper's honest takeaway is that no variable strongly predicts INP in the boreal winter, and only moderate skill ($R^2 \approx 0.5$) is achieved in spring/summer. This is an important negative result and is nicely summarized at the end of the introduction, telling the community that even many co-located variables at one of the best instrumented stations cannot explain INP variability in general. The authors could commit more fully to this message. The paper is framed around "predicting INP using machine learning" yet the conclusion effectively states that strong links remain "abstruse." This is a cautionary study about the limits of data-driven approaches for INP prediction, which is arguably more useful to the community than a modest positive result would be. However, the paper presents itself as a machine learning study, but the ML algorithms are used exclusively as importance ranking tools, not for predictive modeling. The more closely investigated relations (Table 1) are simple power-law fits with only two parameters, no different from classical empirical fitting. There is no actual ML-based predictive model evaluated with train/test splits. With data from 84 variables, a proper ML regression model (e.g., gradient boosting or random forest regressor) could be benchmarked against the power-law fits. As the paper stands, the title overstates the ML contribution.

### Response

We thank the referee for this important observation. We fully agree that the ML algorithms in this study were applied as importance-ranking and pattern-exploration tools rather than to build an evaluated predictive model in the conventional sense (i.e., with train/test data splits and cross-validated performance metrics). The title "Predicting. . . " does not accurately reflect this scope.

Accordingly, we have changed "Predicting" to "Exploring" in the title. We have also added a clarifying sentence in the introduction (Sect. 2.2) that explicitly states the ML methods were used for feature

importance ranking and hypothesis generation, not for constructing a validated predictive model. We agree with the referee that the primary scientific message—that even with more than 500 co-located variables at one of the most heavily instrumented stations globally, strong links to INP concentrations remain elusive—is a valuable cautionary result, and we have committed more fully to this framing in the revised abstract, introduction, and conclusions.

It is important to note that although an original goal was to develop a trained predictive model, based on our explorations, physical intuition and previous results we did not construct a new ML regression model with train/test evaluation because it could not be justified based on the strength of connections observed.

**Change in manuscript**

- Title: changed *"Predicting"* to *"Exploring."* Following further input from Reviewer 2, the final title is: *"Exploring Ice Nucleation Particle concentrations in a Boreal Environment: limits of machine-learning-assisted variable screening."* This reflects input from both reviewers: "Exploring" addresses the framing concern raised here, while "concentrations" (replacing "properties") and the "limits of..." subtitle address Reviewer 2's observation that only one INP property is examined and that the null result deserves prominence.
- Added a clarifying sentence to Sect. 2.2: *"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."*
- Abstract and conclusion reframed to foreground the cautionary message (see specific edits below).

---

## Specific comments

---

### Line 15 (abstract)

**Reviewer comment**

Specify how the results underscore the need for site-specific parameterizations and suggest on which variables parameterizations could instead be based, considering that none of the 84 variables included here work.

**Response**

We agree that the abstract conclusion sentence should be more specific. Although no single variable provides strong predictive power across the full dataset, the spring/summer PINCii data show that local biogenic and chemical proxies (fluorescent particle concentrations and nitrate aerosol mass) yield adjusted $R^2 \approx 0.5$. We have added a mention of these proxies to the final abstract sentence.

**Change in manuscript**

Changed the final abstract sentence from *"These results underscore the need for site-specific parameterizations to capture INP variability in the complex boreal environments."* to *"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."*

---

## Line 20

### Reviewer comment

Explain why the coexistence is inherently unstable. In the same sentence it is mentioned that mixed-phase clouds persist for days, which seems to contradict this statement.

### Response

"Inherently unstable" refers to thermodynamic instability: below 0°C, ice is the thermodynamically stable phase and supercooled liquid water is in a metastable state. The apparent contradiction with long-lived mixed-phase clouds is resolved by the distinction between thermodynamic instability and kinetic stability. Spontaneous ice nucleation faces a high kinetic energy barrier (homogeneous nucleation requires temperatures below approximately −38°C), so supercooled liquid water persists until an ice-nucleating particle (INP) catalyses freezing by lowering this barrier. Mixed-phase clouds can therefore persist for hours to days despite being thermodynamically unstable, precisely because efficient INPs are rare. We have added a brief parenthetical to make this distinction explicit.

### Change in manuscript

Changed *"unstable co-existence of ice and liquid water"* to *"thermodynamically metastable co-existence of ice and supercooled liquid water (below 0°C, ice is the stable phase but a high kinetic barrier to nucleation sustains liquid water in the absence of efficient INPs)"*.

---

## Line 22

### Reviewer comment

The Arctic is considered a high-latitude region, not a region beyond high latitudes.

### Response

We thank the referee for this correction. "Beyond high-latitudes" is geographically imprecise. We have corrected the phrasing.

### Change in manuscript

Changed *"Beyond high-latitudes"* to *"At high latitudes, particularly in the Arctic"*, mixed-phase clouds play out-sized roles in regulating climate...

---

### Line 23

#### Reviewer comment

Summarize what the underlying amplifying feedbacks are.

#### Response

We have added brief examples of the relevant Arctic feedbacks.

#### Change in manuscript

Changed *"especially in the Arctic where underlying feedbacks have amplifying effects"* to *"where amplifying feedbacks—such as the ice-albedo feedback and the water vapour–temperature amplification characteristic of polar warming—have particularly pronounced effects"*.

---

### Line 26

#### Reviewer comment

Summarize how CCN and INP are fundamental and to which cloud processes.

#### Response

We have specified the relevant cloud processes.

#### Change in manuscript

Changed *"cloud processes"* to *"cloud formation, precipitation efficiency, and radiative properties"*.

---

### Line 27

#### Reviewer comment

Clarify the difference between INP occurrence and abundance.

#### Response

We have added a brief parenthetical to define both terms.

#### Change in manuscript

Changed *"occurrence and abundance"* to *"occurrence (whether INPs are present at all) and abundance (how many are present per unit volume of air)"*.

---

## Line 33

### Reviewer comment

Providing some more details on the instrumentation, specifically about the instruments measuring the 84 variables used in this study, would be helpful. This could be done in a supplementary table including the instrument name, brand, sampling frequency, and volume/flow.

### Response

Comprehensive instrument metadata for all SMEAR II variables is documented in the SmartSMEAR database (Junninen et al., 2009) and the station overview (Hari and Kulmala, 2005). For the HyICE-2018 campaign instruments specifically, a detailed instrument table including names, design specifications, and operating conditions is provided in Brasseur et al. (2022, their Table 1 and Appendix). We have added a sentence in Sect. 2.2 directing readers to these resources.

A dedicated supplementary table is not included because complete metadata are accessible via the SmartSMEAR online portal, making a separate table redundant.

### Change in manuscript

Added the following sentence to Sect. 2.2: *"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)."*

---

## Lines 50 and 75

### Reviewer comment

Specify the time resolution with which the CFDCs measure INP concentrations and explain why the data are averaged over 20 min or 1 hour (Figure 3). Figure 1 shows INP concentrations often reaching hundreds per liter. With a sample flow of 1 L $min^{-1}$ there should be plenty of signal in 1 min averages or even 10 s averages, which would substantially increase the number of INP data points available for analysis. Additionally, correlations of ambient measurements depend on the time resolution or averaging interval. The temporal scale at which a correlation is seen also identifies the scale of the process that drives the changes in variables. Investigating correlations at different time resolutions, which seems possible with this dataset, could be interesting for the very variable INP concentrations. Such an analysis could be added, and the dependence of correlation analyses on the temporal data resolution should be discussed.

### Response

Both PINC and PINCii share an identical measurement cycle: a 5-minute background (particle-free) period followed by a 15-minute ambient sampling period, cycling continuously (i.e., . . . 5 min background − 15 min ambient − 5 min background − 15 min ambient . . . ). Each INP concentration value is derived by subtracting the interpolated background count from the ambient-period count during the 15-minute window, yielding one data point every 20 minutes. The 15-minute ambient window is therefore the sampling window; the 20-minute figure is the data cadence, not the sampling

window. This background–ambient cycle is the instruments' fixed operational protocol, not a post-hoc averaging choice. At a sample flow rate of 1 L $min^{-1}$, the 15-minute window provides a 15 L sample volume per data point. During the PINC winter deployment, where INP concentrations at $T_l = -31$°C were typically <1 INP $L^{-1}$, this volume ensures that most sampling intervals yield at least a few particle counts sufficient for a reliable concentration estimate.

The referee notes that Fig. 1 shows periods—primarily during PINCii spring/summer operation—where INP concentrations reach hundreds per liter, for which even a 1-minute ambient window would yield ample particle counts. However, the background–ambient cycle is the fixed operational protocol of both instruments and cannot be shortened without fundamentally redesigning the background subtraction scheme. The cycle was held constant throughout the campaign to ensure consistency across all data points.

The hourly mean data used for the Pearson correlation analysis in Fig. 3 were chosen to align with the coarser time resolution of some complementary datasets. We have explicitly stated this reasoning in the Fig. 3 caption and confirmed that repeating the analysis with the native 20-minute data cadence yields qualitatively similar Pearson correlations.

We acknowledge the referee's conceptual point that the temporal scale of a detected correlation can reveal the timescale of the driving process. The qualitatively similar correlations at 20-minute and hourly resolution suggest that the dominant co-varying processes operate on timescales longer than 20 minutes (diurnal or synoptic scale), consistent with the broad seasonal structure visible in Fig. 1. A systematic multi-scale temporal correlation analysis at all available resolutions would, however, constitute a substantial undertaking; we have added a brief discussion of these implications.

**Change in manuscript**

- Added to Sect. 2.3: *"Both PINC and PINCii operate with a 15-minute ambient sampling window bracketed by 5-minute background (particle-free) periods, yielding one INP 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 $L^{-1}$)."*
- Updated time-series description (Sect. 3.1) to: *"... each with a 15-minute ambient sampling window (one data point every 20 minutes) ..."* and *"Each point represents a 15-minute ambient sampling period (with 5-minute background periods before and after, yielding one data point every 20 minutes)."*
- Added to the Fig. 3 discussion and caption: reasoning for the use of hourly means (alignment with coarser complementary datasets), confirmation that the native 20-minute data cadence yields qualitatively similar Pearson correlations, and a note on the implications for campaign-based INP correlations reported in the literature.

---

## Section 2.2

**Reviewer comment**

Currently, the ML techniques are not explained in this section. Consider changing the section title to "Complementary Data Selection." Add more details about how datasets were processed for the

analysis. Add a separate methodology section about ML algorithms, feature selection criteria, to make the approach reproducible.

**Response**

We agree that the ML methodology deserves more explicit documentation. We have added a paragraph to Sect. 2.2 that describes: (i) the algorithms applied (random forest and decision trees for feature importance ranking; principal component analysis and K-means clustering for dimensionality reduction and pattern exploration); (ii) that no train/test data split was applied, as the purpose was exploratory importance ranking; and (iii) how datasets were harmonised to a common time base (20-minute resolution, with coarser-resolution variables assigned to the nearest 20-minute interval).

We elect to update the section title "Complementary Data and Machine Learning" and maintain a single section that covers both the data and the ML framework.

**Change in manuscript**

Added the following paragraph to Sect. 2.2:

> *"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. These algorithms were applied as exploratory tools to rank variables by their statistical association with INP concentrations; no train/test data splitting or cross-validation was performed. Prior to analysis, all variables were harmonised to a common 20-minute time base; variables measured at coarser time resolution were assigned to the nearest 20-minute interval, and those measured at finer resolution were averaged."*

---

## Line 73

**Reviewer comment**

Explain which hypotheses are tested to illuminate the sources and mechanisms.

**Response**

We have clarified the scientific hypotheses that motivated the variable selection and analysis in Sect. 2.3.

**Change in manuscript**

Added the following sentence to Sect. 2.3: *"Specifically, we test whether new particle formation (NPF) events generate particles that grow into the INP-relevant size range (>0.5 μm); 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."*

---

### Line 76

#### Reviewer comment

Clarify what is meant by "straightforwardly intercompared."

#### Response

The phrase was intended to indicate that PINC and PINCii share a common operating principle (parallel-plate CFDC design) and can thus be compared on similar methodological terms. The word "straightforwardly" is unnecessary and potentially misleading. We have simplified the phrasing.

#### Change in manuscript

Changed *"are somewhat straightforwardly intercompared as demonstrated by"* to *"have been directly compared by"*.

---

### Section 2.3

#### Reviewer comment

Clarify why the details on the CFDC chambers are relevant for this paper. None of the details are referred to later.

#### Response

The CFDC technical specifications (chamber dimensions, flow rates, temperature and humidity conditions) are standard documentation for peer-reviewed CFDC-based INP papers, enabling reproducibility and comparison with other CFDC studies in the literature. More importantly, the operating conditions ($T_l = -31$°C for PINC, $T_l = -32$°C for PINCii, $\mathrm{RH}_w = 105\%$) define the ice nucleation activation threshold and the temperature at which the reported INP concentrations were measured—directly relevant to interpreting all results. We have added a brief sentence at the opening of Sect. 2.3 clarifying this purpose.

#### Change in manuscript

Added the following sentence at the opening of Sect. 2.3: *"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."*

---

## Line 107

### Reviewer comment

Contrary to what is stated here, Brasseur et al., 2022 mention a sampling window of 15 min for PINCii.

### Response

We thank the referee for this correction. Brasseur et al. (2022) specify a 15-minute ambient sampling window for PINCii. In fact, PINC follows the same 5-minute background + 15-minute ambient cycle, so both instruments have a 15-minute sampling window and produce one data point every 20 minutes. The original phrase "20 minute sampling windows" incorrectly equated the data cadence (one point per 20 min) with the sampling window (15 min ambient period). We have corrected all occurrences in the manuscript to accurately describe the 15-minute ambient sampling window and the 20-minute data cadence.

### Change in manuscript

- Corrected *"with 20 minute sampling windows"* to *"each with a 15-minute ambient sampling window (one data point every 20 minutes) (Brasseur et al., 2022)"*.
- Corrected *"Each point represents a 20 minute sampling window"* to *"Each point represents a 15-minute ambient sampling period (with 5-minute background periods before and after, yielding one data point every 20 minutes)"*.

---

## Figure 1

### Reviewer comment

Indicate the temperature at which INP concentrations were measured. Check the units in panel c. In panel e, the last line is shown with very weak colours.

### Response

We thank the reviewer for catching the unit error in panel (c). We have updated Figure 1 and its caption to address all three points: (i) the measurement temperatures ($-31$°C for PINC, $-32$°C for PINCii) are now stated in the panel (a) caption; (ii) the units in panel (c) have been corrected to $\#/\mathrm{cm}^3$; and (iii) regarding the weak colours in panel (e), the colour scheme follows the standard AMS convention (chloride: purple; ammonia: yellow; sulfate: red; nitrate: blue; organics: green), which we have now noted explicitly in the caption.

### Change in manuscript

Figure 1 caption updated: (i) added *"INP concentrations were measured at $T_l = -31\,°C$ (PINC) and $T_l = -32\,°C$ (PINCii)"* to the panel (a) description; (ii) added *"($\#/cm^3$)"* as the unit in the panel (c) description; and (iii) added *"(colours follow AMS convention)"* to the panel (e) description.

---

## Lines 119–120

### Reviewer comment

Clarify why NPF is relevant here.

### Response

NPF events generate large numbers of freshly nucleated particles that can grow into the size range relevant for ice nucleation during sustained growth events lasting several hours. Additionally, NPF events are a well-known feature of the SMEAR II station and could influence the total particle loading and composition in ways that are relevant to the INP analysis. We have added a brief explanatory sentence.

### Change in manuscript

Added the following sentence to Sect. 3.1: *"New particle formation (NPF) events are notable here because freshly nucleated particles can grow into the INP-relevant size range during sustained growth episodes, potentially contributing to the total INP-active aerosol population."*

---

## Line 122

### Reviewer comment

What relationships can be observed in Figure 1 that motivate the analysis?

### Response

We have revised the introductory sentence to Sect. 3.2 to specify the visual patterns that motivate the more systematic analysis: the seasonal progression in temperature and snow depth, the spring/summer increase in fluorescent biological particles, the shift in aerosol chemical composition (increasing organic fraction), and the apparent co-variability of INP with some of these tracers.

### Change in manuscript

Changed the opening of Sect. 3.2 from *"The relationships suggested by the time series in Fig. 1 motivate a more objective evaluation..."* to *"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..."*

---

## Lines 132–134

### Reviewer comment

Repetition from Sect. 2.2.

### Response

We thank the referee for identifying this duplication. The description of the dimensional reduction criteria (exclusion of NaN-heavy, low-variability, and redundant variables) was already presented in Sect. 2.2 and has been removed from Sect. 3.2.

### Change in manuscript

Removed the following paragraph from Sect. 3.2 (it duplicated material already presented in Sect. 2.2):

> *"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."*

---

## Line 139

### Reviewer comment

Use "INP concentration" instead of "ice nucleation activity."

### Response

We agree and have replaced both occurrences in Sect. 2.2.

### Change in manuscript

- Changed *"Although ice nucleation activity"* to *"Although INP concentration"* (first occurrence in Sect. 2.2).
- Changed *"that can be used to follow ice nucleation activity"* to *"that can be used to follow INP concentration"* (second occurrence in Sect. 2.2).

---

## Line 140

### Reviewer comment

The references attribute bio-INP to much lower concentrations at higher temperatures than below −30°C. Provide a supporting citation for biological particles contributing substantially at low temperatures.

### Response

The referee correctly notes that primary biological INPs (fungal spores, pollen, bacteria) are typically most active at temperatures above approximately −20°C, and contributions at −31/−32°C are less well established. We have revised the statement to clarify that fluorescent particles serve as a tracer for biological aerosol that may include INP-active species—rather than claiming they are the dominant INP at our measurement temperatures—consistent with the observational approach in Paramonov et al. (2020) and Schneider et al. (2021) at the same site.

### Change in manuscript

Changed *"biological particles that yield a fluorescence signal are known to be some of the most abundant INPs in many settings"* to *"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)"*.

---

## Line 144

### Reviewer comment

Only half of the variables are listed in Figure 2, while others are selected by intuition. Specify for each variable based on which information or hypothesis it was selected.

### Response

We have added explicit justification for each of the six variables shown in Fig. 3. Fluorescent particle concentrations and >0.5 µm number concentration are established predictors used in the DeMott (2010) and Tobo (2013) parameterizations; particle mass is a proxy for total aerosol loading; black carbon emerged as a top-ranked variable in the random forest analysis; nitrate ranked highly and co-varies with biogenic tracers; organic aerosol mass is motivated by Hyytiälä's well-documented biogenic aerosol seasonality (Schneider et al., 2021).

### Change in manuscript

Added the following sentence to Sect. 3.2: *"The six variables shown in Fig. 3 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 µm concentration per DeMott et al. (2010) and Tobo et al. (2013); organic aerosol mass per the biogenic seasonality documented by Schneider et al. (2021))."*

---

## Lines 148ff

### Reviewer comment

That BC ranks highly in Fig. 2 and yields the highest adjusted $R^2$ of all predictors for PINCii (Table 1), yet is a poor INP in laboratory studies, is a provocative finding that could be explored beyond noting that it is surprising and pointing to aging and oxidation as possible explanations. It could be mentioned that Paramonov et al. (2020) also reported that BC correlated well with INP concentrations on a short timescale (their Fig. 5).

### Response

We thank the referee for pointing to this reference. We have added a citation to Paramonov et al. (2020) and a sentence acknowledging that their short-timescale BC–INP correlation at the same site independently supports our finding, and expanded the discussion of possible mechanisms (aging, oxidation, and mixing with organic material).

### Change in manuscript

Added the following sentences to Sect. 3.2: *"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 the enhancement of BC's ice-nucleating ability through atmospheric aging, oxidation, and coating with organic material (DeMott et al., 1990; Mahrt et al., 2020a,b)."*

---

## General suggestion: connections to previous HyICE articles

### Reviewer comment

In general, the paper would be strengthened if connections to findings from previous HyICE articles were integrated in more detail.

### Response

We thank the referee for this general suggestion and agree that tighter integration with the other HyICE-2018 publications strengthens the manuscript. In the revised text we have (i) added a citation to and discussion of Paramonov et al. (2020) in the context of the BC–INP correlation finding (see response to lines 148ff above); (ii) retained and where appropriate expanded citations to Schneider et al. (2021), Brasseur et al. (2022), Brasseur et al. (2024), and Vogel et al. (2024) in the introduction and results sections; and (iii) added a sentence noting that the better Tobo (2013) agreement in Brasseur et al. (2022, their Fig. 8 and consolidated 28 March summary in Fig. 11) reflects a focused intercomparison subset rather than full campaign conditions (see response to Figure 5 below). We believe these additions place our results more explicitly in the context of the broader HyICE-2018 body of work.

### Change in manuscript

No single new change in manuscript; see individual responses to lines 148ff, Figure 5, and line 209 for the specific additions.

## Lines 151ff

### Reviewer comment

Clarify which instruments were used to measure the variables on the horizontal axes in Fig. 3.

### Response

We have added instrument identifications in the text and Fig. 3 caption: fluorescent particle concentrations (panel a) are from the WIBS; total particle mass (panel b) and organic and nitrate aerosol mass (panels d, e) are from the AMS; >0.5 µm particle number concentration (panel c) is from the aerodynamic particle sizer (APS); black carbon mass concentration (panel f) is from the aethalometer.

### Change in manuscript

Added two instrument-identification parentheticals in Sect. 3.2 and corresponding instrument names in the Fig. 3 caption. In the body text, *"(fluorescent biological aerosol particles from the WIBS; total particle mass from the AMS; >0.5 µm number concentration from the APS)"* was inserted after *"horizontal axes"*, and *"(all three from the AMS except BC, which is from the aethalometer)"* was inserted after the list of bottom-panel variables.

---

## Line 155 and Table 1

### Reviewer comment

Previously, it is implied that organic mass and >500 nm concentration were included based on intuition and not a high-skill ranking.

### Response

We acknowledge this apparent inconsistency and have made the dual selection rationale explicit: variables were chosen either because they ranked highly in the random forest analysis (fluorescent particles, BC, nitrate, particle mass) or because of their established physical connection to INP in prior parameterizations or site-specific studies (>0.5 µm concentration per DeMott/Tobo; organic mass per Hyytiälä biogenic seasonality). This is now stated clearly in Sect. 3.2.

### Change in manuscript

Clarifying sentence added (see also response to line 144 above); no change to Table 1 entries.

---

## Figure 3 and related interpretation

### Reviewer comment

Explain why the data are split for the Pearson correlation analysis if the goal is to investigate the variables' predictive power for INP concentrations. The difference between PINC and PINCii data shows that the correlations are only good for a subset, not generally. Provide a discussion on what this implies for the many campaign-based INP correlations reported in the literature.

### Response

We thank the referee for this important comment.

The PINC and PINCii data are plotted separately because the two instruments operated in distinctly different seasons—PINC in winter (February–April) and PINCii in spring/summer (April–June)—with distinct ambient meteorological and aerosol conditions. Pooling the two datasets would conflate these seasonal differences and conceal a key finding: the correlation between INP concentrations and the examined predictor variables is strongly season-dependent. The PINC winter data show essentially no correlation with any of the monitored variables (adjusted $R^2 \leq 0$), whereas the PINCii spring/summer data yield moderate correlations with several aerosol proxies (adjusted $R^2 \approx 0.5$ for fluorescent particles and nitrate). Merging the two would produce an artificial intermediate correlation coefficient that misrepresents the behaviour of each individual subset.

The referee correctly identifies the central implication: the correlations shown in Fig. 3 are specific to the PINCii spring/summer subset and do not hold generally across the full campaign. This is an important and candid finding that we have made more explicit in the revised manuscript.

This result has significant implications for campaign-based INP correlations reported in the literature. Many published INP–aerosol correlations are derived from intensive field campaigns of limited duration (days to weeks), often conducted during a single season or under specific meteorological regimes. Our results suggest that such correlations are inherently conditioned by the ambient aerosol and environmental conditions that prevail during the specific measurement period. A correlation identified during a spring campaign at a boreal forest site may not hold during winter at the same site, let alone at a different location or ecosystem. Authors and readers of campaign-derived INP correlations should therefore exercise caution when extrapolating or generalising such relationships beyond the specific conditions under which they are established.

### Change in manuscript

Added to Sect. 3.2: *"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). [...] 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."*

---

## Figure 3

### Reviewer comment

For all panels, consider marking significant $r$ values with an asterisk instead of reporting absurdly small p values. a), c): Check units on the horizontal axes. The different Pearson correlations for PINC and PINCii data, which seem to overlap for the most part in the log–log scatter plots, are surprising. Clarify if the correlation was calculated using the raw data or the log-transformed data. Consider showing the data on a linear scale. Provide the number of PINC and PINCii data points. Are they the same in each panel? Why are hourly mean data used here, and does using the mean affect the outcome of the analysis compared to using 20 min data?

### Response

We thank the referee for these detailed technical comments on Fig. 3.

(a) **Asterisks for significance:** We have added asterisks to indicate statistically significant correlations ($p < 0.05$) and removed the previously reported exact $p$-values.

(b) **Units on horizontal axes:** We have verified and corrected the units on all horizontal axes in panels (a) and (c).

(c) **Differing Pearson correlations for apparently overlapping data:** The visually overlapping scatter in the log–log plots can yield different Pearson $r$ values because the correlation is sensitive to the full range and distributional structure of the data, not just the visual density of points. The winter PINC data span a narrower and lower range of both INP and predictor values; thus the correlation structure is distinct even where the scatter appears to overlap with the spring/summer PINCii data.

(d) **Log-transformed vs. raw data:** The Pearson correlation coefficients were calculated from the raw (untransformed) data; the log–log axes are used solely for visualisation to span the wide dynamic range of the measurements. We have clarified this explicitly in the updated Fig. 3 caption.

(e) **Linear scale:** We considered adding linear-scale versions of the panels but, given the several orders of magnitude spanned by both INP concentrations and the predictor variables, a log–log representation is the most practical for displaying the full dataset.

(f) **Number of data points:** We have added the number of PINC and PINCii data points to the Fig. 3 caption. The number of points differs between panels because different sensors have different data availability owing to maintenance, calibration, and/or instrument downtime.

(g) **Hourly vs. 20-minute means:** We clarify in the caption that hourly means were used to align with the coarser temporal resolution of some complementary datasets; repeating the analysis using the native 20-minute data cadence yields qualitatively similar Pearson correlations.

### Change in manuscript

Figure 3 updated with asterisks and verified units on horizontal axes. Caption expanded with: PINC and PINCii data point counts, clarification that Pearson coefficients were calculated from raw (untransformed) data with log–log axes used for visualisation only, rationale for use of hourly means, and confirmation that 20-minute data yields qualitatively similar results.

## Lines 192ff

### Reviewer comment

Eq. (2) reproduces Eq. (2) from Tobo et al., 2013, which uses $n_{AP>500\,\mathrm{nm}}$. However, the text seems to imply that Eq. (3) uses the number of fluorescent particles. How would the updated Eq. (3) from Tobo et al., 2013 perform? Which formula was used for Figure 8 in Brasseur et al., 2022?

### Response

Both Eq. (1) (DeMott et al., 2010) and Eq. (2) (Tobo et al., 2013) in our manuscript use the number concentration of aerosol particles larger than 0.5 µm ($n_{AP>0.5\,\mu\mathrm{m}}$) as the predictor—not fluorescent particle counts. The Tobo et al. (2013) paper also provides a separate parameterisation using fluorescent biological aerosol particle (FBAP) concentrations (their Eq. 3), but we did not apply that variant here; our Eq. (2) follows the original $n_{AP>0.5\,\mu\mathrm{m}}$-based form. We have clarified this distinction explicitly in the text. Regarding Brasseur et al. (2022) Fig. 8: that figure used the same $n_{AP>0.5\,\mu\mathrm{m}}$-based Tobo (2013) formulation, consistent with our Eq. (2).

Applying the FBAP-based Tobo (2013) variant would require matching the FBAP dataset to the parameterisation's calibration dataset and would be an additional, substantial analysis.

### Change in manuscript

Added the following note after Eq. (2): *"Note that both Eqs. (1) and (2) use the total aerosol number concentration $n_{\mathrm{AP}>0.5\,\mu m}$ 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."*

---

## Figure 4

### Reviewer comment

Would $p = 1$ not require $\chi^2 = 0$? Double-check the values. $\chi^2$ for PINC data is huge and $p = 0$. The bimodal fit does not seem visually superior.

### Response

We thank the reviewer for flagging these anomalies. Upon careful re-examination, both issues stemmed from the same underlying error in the statistical workflow.

**(i) Anomalous $p \approx 1$ in panel (a) and the large $\chi^2$ with $p = 0$ for PINC.** The $\chi^2$ goodness-of-fit test was erroneously applied to relative frequencies (dimensionless values summing to 1, each $\approx 0.01$–$0.10$) rather than to raw bin counts. Because $\chi^2 = \sum(O - E)^2/E$, using small fractions systematically drove the statistic toward zero, producing artificially large $p$-values ($p \approx 1$) regardless of fit quality—explaining the anomaly flagged in panel (a). This has been corrected: the $\chi^2$ test now operates on raw bin counts as observed frequencies, with expected counts derived by integrating the

fitted probability density over each bin and scaling by the total sample size $N$. Degrees of freedom are adjusted for the number of estimated parameters (ddof = 2 for the unimodal fit; ddof = 5 for the bimodal fit).

**(ii) Distorted bimodal fits in panel (b).** A second error compounded the bimodal results: `curve_fit` was minimising residuals between the log-normal PDF (units: per unit $x$) and the relative frequencies (dimensionless), without accounting for the variable widths of the log-spaced bins. This mismatch distorted the optimised component widths and mode locations, particularly for the PINC dataset whose distribution is more broad. The bimodal fitting objective has been corrected to match probability masses (PDF × bin width) against relative frequencies. Initial parameter guesses were also made data-adaptive (placed at the 25th and 75th percentiles of each dataset) to reduce sensitivity to poor local minima.

With these corrections, the figure legends report $\chi^2 = 81.98$ ($p = 0.001$) and $\chi^2 = 157.19$ ($p < 0.001$) for the unimodal PINCii and PINC fits, respectively, so the strict unimodal log-normal null is rejected for both instruments at conventional significance. For the bimodal fits, PINCii attains $\chi^2 = 63.54$ ($p = 0.028$), i.e. a lower statistic than the unimodal case but still marginally inconsistent with the data at the 5% level, while the PINC bimodal fit remains rejected ($\chi^2 = 242.72$, $p < 0.001$) and does not improve overall agreement—consistent with the reviewer's visual assessment that the bimodal improvement was not compelling. The revised Figure 4 matches these values and is provided for comparison with the previous version.

**Change in manuscript**

- Updated the image for Figure 4 to the corrected version.
- Updated the Figure 4 caption to: *"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 (*ddof = 2 *for unimodal;* 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."*
- Updated the body text discussing Figure 4 (Sect. 3.2) from *"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 distribution, 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."* to *"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 modest 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."*

---

## Line 198

### Reviewer comment

Explain why the CFDCs operating above water saturation were not measuring immersion freezing comparable to the parameterisation derived from datasets obtained with INSEKT. It would be interesting to see how well the parameterizations perform at low temperatures.

### Response

CFDCs operate by exposing aerosol particles to a short traverse (∼seconds) through a controlled cold, humid airstream, primarily activating condensation freezing and deposition nucleation, with limited potential for full immersion freezing. Bulk immersion freezing assays (such as INSEKT, used to derive the Schneider et al., 2021 and Brasseur et al., 2024 parameterisations) suspend particles in macroscopic water volumes cooled over minutes to hours, probing a fundamentally different nucleation pathway and timescale. Because the two approaches probe different physical processes, direct quantitative comparison is not straightforward. We have added a brief clarification to the text. Benchmarking the parameterisations at low temperatures would require dedicated datasets not available in the current study.

### Change in manuscript

Added the following sentence to Sect. 3.3: *"Direct comparison is not straightforward because CFDCs primarily activate condensation and deposition freezing on a timescale of seconds, whereas the bulk immersion assays used to derive the Schneider et al. (2021) and Brasseur et al. (2024) parameterisations probe a different nucleation pathway over longer timescales."*

---

## Figure 5

### Reviewer comment

Double-check the calculation using Tobo et al., 2013. Figure 8 in Brasseur et al., 2022 shows good agreement between Tobo 2013 and PINC/PINCii measured INP concentrations during the instrument intercomparison.

### Response

We have re-verified our implementation of the Tobo (2013) parameterisation and confirm the calculation is correct. We have carried out a careful, point-by-point comparison between our Fig. 5 and Brasseur et al. (2022) Fig. 8, and find that the apparent discrepancy is explained by three fundamental differences in dataset scope, comparison methodology, and aerosol conditions:

**(i) Dataset scope.** Brasseur et al. (2022) Fig. 8 covers only four targeted intercomparison days: 22 March, 28 March, 26 April, and 28 April 2018, selected specifically to maximise temporal overlap

between instruments. Brasseur et al. (2022) themselves caution that the parameterisation comparison "might not be representative of the entire HyICE-2018 campaign." Our Fig. 5 covers the complete PINC deployment (19 February–2 April) and PINCii deployment (22 April–10 June), spanning the full range of seasonal aerosol and biological conditions at SMEAR II.

**(ii) Qualified and limited agreement in Brasseur et al. (2022).** Brasseur et al. (2022, their Fig. 11 and Sect. 3.2.3) consolidate all intercompared online and offline instruments for 28 March 2018 into a single INP temperature spectrum and count how many data points fall inside each parameterisation's shaded envelope (for Tobo (2013) and DeMott (2010), the envelope spans the APS $N_p(> 0.5\,\mu\text{m})$ daily mean $\pm 1$ standard deviation). They report 35% inside Tobo (2013), 3% inside DeMott (2010), and 19% inside Schneider et al. (2021)—so Tobo is "best" in a relative sense only. Their Fig. 8 time series is discussed qualitatively over the four intercomparison days; on 22 March—the only winter intercomparison day for PINC—they explicitly state that "none of the parameterizations successfully represents the measured concentrations," with measured INP concentrations both above and below the Tobo (2013) band at different times of day. The description of Tobo (2013) as showing "good agreement" is therefore relative, qualified, and does not hold for the winter portion of the campaign.

**(iii) Comparison methodology.** Brasseur et al. (2022) Fig. 8 presents a time-series visualisation in which measured INP concentrations are plotted alongside parameterisation *shaded bands* representing a $-29$ to $-32$°C temperature window. This is fundamentally different from our Fig. 5, which is a scatter plot of model-predicted versus measured INP concentrations evaluated by the coefficient of determination ($R^2$). Time-series visual overlap within order-of-magnitude bounds does not translate to predictive skill: even the consolidated Fig. 11 accounting of Brasseur et al. (2022) leaves most spectrum points outside the Tobo envelope (35% within), and such partial overlap will still yield strongly negative $R^2$ in a scatter plot if individual predicted and measured pairs are poorly correlated.

Beyond these methodological distinctions, the two PINCii intercomparison days (26 and 28 April) occurred immediately after the snowmelt transition at SMEAR II, when biological particle concentrations were increasing (Brasseur et al., 2022, their Fig. 6c). These two days represent the aerosol conditions most favourable to a parameterisation anchored on biological aerosol particles, and are not representative of the broader PINCii spring/summer measurement period when new particle formation (NPF) events frequently grow secondary organic aerosol particles into the $>0.5$ $\mu$m size range without a commensurate increase in biological INPs.

We have substantially expanded the manuscript discussion (Sect. 3.3) to reflect this nuanced comparison.

**Change in manuscript**

Replaced the single added sentence in Sect. 3.3 with: *"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."*

---

### Line 209

#### Reviewer comment

Clarify why in Figure 5b) Tobo 2013 never performs well.

#### Response

The systematic overprediction by Tobo (2013) for PINCii data across the entire spring/summer deployment is the result of two compounding factors that are specific to the Hyytiälä boreal environment during this period.

**(i) Tobo (2013) calibration environment vs. SMEAR II spring aerosol.** The Tobo (2013) parameterisation was developed at the Rocky Mountain Biological Laboratory (RMBL), a mid-latitude temperate ponderosa pine forest in Colorado, USA. At that site, particles larger than 0.5 $\mu$m are predominantly primary biological aerosol particles (fungal spores, plant debris, pollen fragments), and thus the high-$n_{\mathrm{AP}>0.5\,\mu\mathrm{m}}$ conditions directly correspond to elevated biological INP concentrations. At SMEAR II during spring and early summer, new particle formation (NPF) events—for which Hyytiälä is globally renowned (Dal Maso et al., 2005; Kulmala et al., 2013)—nucleate and grow secondary organic aerosol (SOA) particles into the >0.5 $\mu$m size range over the course of hours to days. These SOA-grown particles inflate $n_{\mathrm{AP}>0.5\,\mu\mathrm{m}}$ substantially without a commensurate increase in biological INP concentrations, because secondary organic material is a poor ice nucleant. The parameterisation therefore "sees" elevated aerosol particle counts and predicts large INP concentrations that are not realised.

**(ii) Sub-Arctic boreal ecosystem vs. temperate forest.** The boreal forest of southern Finland in spring and early summer is ecologically distinct from a temperate mid-latitude forest. Biological primary aerosol emission is lower in the early season (April–June) than in the peak summer of a temperate forest, and the diversity of biogenic INP sources is correspondingly reduced. Accordingly, the fraction of $n_{\mathrm{AP}>0.5\,\mu\mathrm{m}}$ that consists of INP-active biological material is lower at Hyytiälä than the RMBL calibration environment, causing the parameterisation to systematically overestimate.

We note that this mechanism is consistent with the observation in Brasseur et al. (2022) that Tobo (2013) performs relatively better on 26 and 28 April—immediately after snowmelt when biological fractions in the aerosol were rising (their Fig. 6c)—than over the broader spring/summer PINCii period. We have updated the manuscript text to reflect this mechanistic explanation.

#### Change in manuscript

Added after Table 1 in Sect. 3.3: *"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."*

---

### Line 218

#### Reviewer comment

Please provide a more in-depth interpretation of the coefficients found in Table 1.

#### Response

We have added a paragraph after Table 1 interpreting the fitted exponents and pre-factors.

#### Change in manuscript

Added the following paragraph after 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."*

---

### Figure 6

#### Reviewer comment

Check units in panel titles for panels a and c.

#### Response

We have checked and corrected the units in panels (a) and (c) of Fig. 6.

#### Change in manuscript

Figure 6 panel titles updated with correct units.

---

## Technical corrections

**Line 36 ("bridging")** Changed *"bridging"* to *"extending"* — see also the specific comment response to lines 148ff above.

**Lines 132–134 (repetition)** Removed — see response above.

**Line 139 ("ice nucleation activity")** Both occurrences replaced with "INP concentration" — see response above.

**Line 162 ("to be ice active")** Changed *"be ice active"* to *"contain INPs"*.

**0.5 µm vs. 500 nm** All occurrences standardised to "0.5 µm" throughout text, figure captions, and Table 1.

**Brasseur et al., 2022 author list** The author list in the bibliography has been updated to include the full list as also indicated by the referee: 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.

---

We thank the referee once more for the thorough and constructive review. We believe the revisions substantially improve the manuscript's clarity, scientific framing, and accessibility.