

Supplement of

The spatio-temporal variability of ice-nucleating particle concentrations quantified with a new benchmark dataset

Kristoffer Kvist^{1,2}, Eirik Mikal Samuelsen^{1,2,3}, and Rune Grand Graversen¹

¹Department of Physics and Technology, The Arctic University of Norway (UiT), Tromsø, 9019, Norway

²The Norwegian Air Ambulance Foundation, Oslo, 0184, Norway

³Norwegian Meteorological Institute, P.O. Box 6314 Langnes, Tromsø, 9037, Norway

Correspondence: Kristoffer Kvist (kristoffer.kvist@uit.no)

S1 Comparing the normalized accumulated variance of PINE and IS deployments

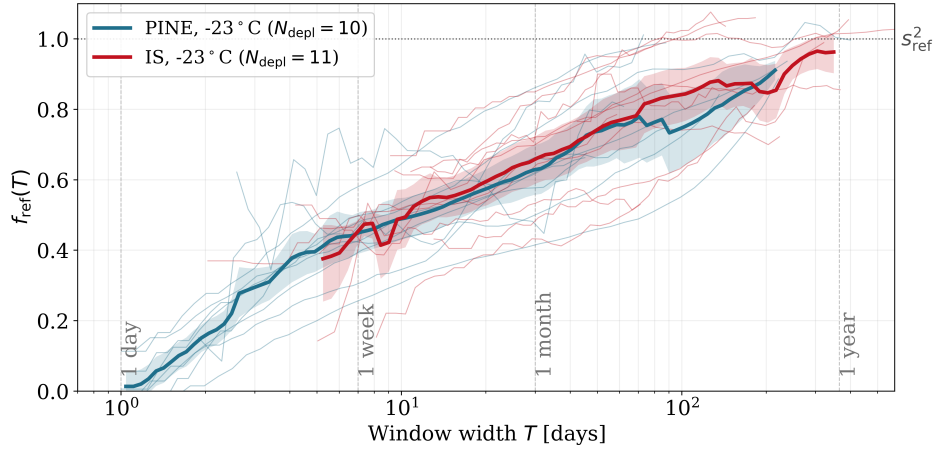


Figure S1. Normalized accumulated variance $f_{\text{ref}}(\Delta t)$ at -23°C of the GloPINE (blue) and IS (red) *deployments* (unique combinations of campaign and platform) of Fig. 6, adjusted to the reference sampling routine with an exposure time of 24 h and a campaign time span of 365 d with $\mathcal{W}(\Delta t)$ (Sect. 3.5). Thin lines show the individual deployments, colored by instrument. Thick lines show the mean over the deployments of each instrument, which gives every deployment equal weight and is drawn where at least three deployments contribute, and the shading gives its 90 % confidence interval from bootstrap resampling of the deployments. The curves are computed as in Fig. 6b. In the legend, N_{depl} is the number of deployments contributing to the respective mean.

S2 Biological and organic fractions of the INP concentrations measured during the KCG campaign

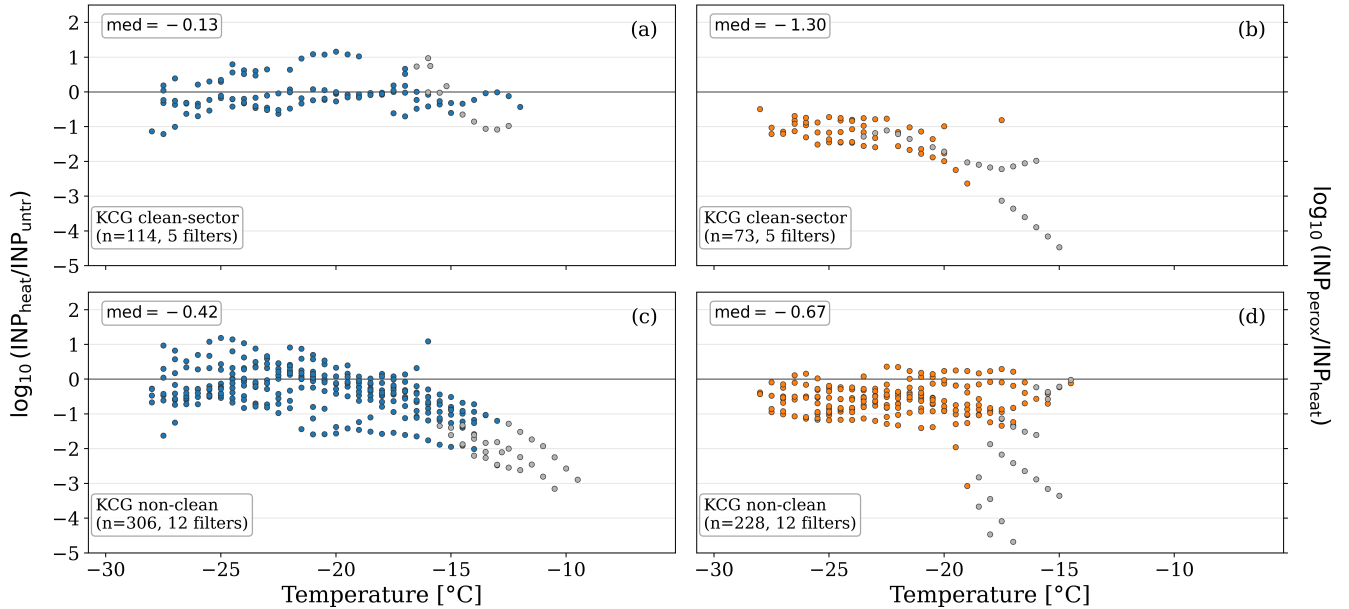


Figure S2. Same as Fig. 7, but for Kennaoook/Cape Grim alone and split by the two sampling modes described in Sect. S3. The upper row (a, b) holds the filters collected in clean-sector mode, which sample air masses from the Southern Ocean only, and the lower row (c, d) those collected in continuous mode. Every matched pair is drawn at its own temperature rather than binned into boxes, because the observations are sparse. Gray points are warm-tail extrapolated, as in Fig. 7.

S3 Environmental and sampling characteristics of the individual campaigns

This section provides details about each campaign, with the goal of guiding users of the dataset in selecting campaigns that reflect their purpose of use. Each entry opens with the environmental setting and relevant sampling details, and closes with a judgment of whether the record is deemed representative of ambient INP conditions over the period it covers. Records shorter than 4 weeks are categorized as not representative regardless of collection methodology, since they include too little synoptic variability. Snow cover is noted for sites that the ERA5 snow-depth field places under snow for a substantial part of the record, which is of interest because a snow-covered surface suppresses local INP sources.

10 S3.1 PINE campaigns

All 20 PINE campaigns are taken from the GloPINE dataset (Herbert et al., 2026) and follow the same protocol. The expansion chamber (Möhler et al., 2021) performs an expansion every 5–15 min regardless of the ambient conditions, and the resulting record is averaged into fixed 6 h intervals and 2 °C temperature bins. The sampling times are therefore determined by the clock

alone. Gaps in the records reflect instrument downtime, not a decision to sample only certain conditions. Below, uptime denotes the fraction of the campaign time span covered by valid measurements.

BEIJING: A low-altitude continental site on the North China Plain near Beijing, surrounded by farmland and low-rise buildings. The aerosol population is influenced by agricultural activity and by intermittent anthropogenic plumes from the city. Air was drawn through a non-heated inlet 12 m above ground. The campaign spans 16 months with little downtime and is the most continuous record in the dataset. The record is representative.

CALISHTO: A mountaintop site at 2314 m on Mount Helmos, Greece. The site is frequently above the boundary layer and exposed to eastern Mediterranean air. Air was drawn through an omnidirectional inlet on the observatory roof. The campaign spans about 7 months at two-thirds uptime. Most 6 h intervals resolve only a single temperature, so the record poorly constrains the shape of the INP spectrum. The record is representative.

CORONA-PINE01, CORONA-PINE04: A lowland site on the research campus of the AIDA facility in Karlsruhe, Germany, in the densely populated upper Rhine valley. In both campaigns, air was drawn through a non-heated inlet 9 m above ground, and two PINE chambers were operated alternately. CORONA-PINE01 lasted less than 3 weeks and CORONA-PINE04 about 5 weeks, both at high uptime and over a narrow temperature range. Neither campaign has an accompanying publication. CORONA-PINE04 is representative of the spring period it covers. CORONA-PINE01 is too short to be representative.

CountIce-pt1, CountIce-pt2: An urban site on the University of Leeds campus, UK, surrounded by low-rise buildings and without strong local dust sources. Part one spans about 4 months at 44 % uptime and part two about 5 months at 51 % uptime. Both records are representative of the periods they cover.

ExINP-ENA: A marine site at the ARM observatory on Graciosa Island, Azores, in the remote eastern North Atlantic. The aerosol is dominated by sea spray within the boundary layer and by long-range transported dust. Air was drawn through an inlet 5.5 m above ground. The campaign spans just under 6 months, from October to March, at 78 % uptime. The record is representative of autumn and winter conditions.

ExINP-GVB: A high-Arctic coastal site at the Gruebadet Atmosphere Laboratory near Ny-Ålesund, Svalbard. Air was drawn through a heated whole-air inlet. The campaign spans only 25 days, and the record is therefore not representative.

ExINP-NSA: A flat, low-lying Arctic coastal site at the Barrow Atmospheric Baseline Observatory, adjacent to the ARM North Slope of Alaska site, on tundra near the Beaufort Sea. Air was drawn through an inlet 12 m above ground. The site is snow-covered for roughly two-thirds of the record, which restricts local surface emission to the summer months. The campaign spans more than 2 years at 64 % uptime and covers all 12 months. It is the longest PINE record in the dataset. The record is representative.

ExINP-SGP: A continental agricultural site at the ARM Southern Great Plains site, Oklahoma, surrounded by farmland. The campaign spans 6 weeks in October and November 2019 at almost complete uptime. A wide temperature range was scanned, which indicates a relatively large population of warm-temperature INPs. The record is representative of the autumn period it covers.

ExINP-ZEP: A high-Arctic mountain-ridge site at the Zeppelin Observatory, Svalbard, at 474 m and usually above the local boundary layer. The surrounding surface is glaciated and snow-covered throughout the record. The campaign spans 6 months

at 62 % uptime. The record is representative of the period it covers.

50 **HyICE2018:** A boreal forest site at SMEAR II, Hyytiälä, Finland. Air was drawn through a heated inlet 6 m above ground. The ground is snow-covered throughout the late-winter period of the campaign. Only a handful of 6 h intervals contain valid measurements, corresponding to 6 % uptime, and all of them lie within 1 °C of −29°C. The record constrains a single temperature only and is not representative.

LIFE-FROSTDEFEND: An agricultural site in a lemon orchard near Aigio, Greece. Air was sampled from a mobile container, so the aerosol carries a strong signal from the local vegetation. The campaign spans just under 4 weeks at 62 % uptime. The sampling period was chosen to coincide with the frost season, but within it the chamber sampled regardless of conditions. The record is too short to be representative.

M-Phase: A research vessel in the Labrador Sea, far from terrestrial sources. The laboratory was mounted on the foredeck; hence, on the sternside of the exhaust to avoid sample contamination. The campaign spans about 5 weeks at 87 % uptime. 60 Most 6 h intervals resolve only a single temperature. Whether periods with exhaust in the wind sector were excluded is not established. The record is representative of the conditions along the cruise track.

PaCE22: A sub-Arctic site at the Pallas Sammallunturi station in northern Finland, above the boreal tree line. The site is snow-covered for about two-thirds of the record. The campaign spans 3 months, from September to December, at 94 % uptime. The record is representative of the autumn and early-winter period it covers.

65 **PICNIC:** A mountaintop site at 1465 m on Puy de Dôme, France, often located in the free troposphere. The measurements were made during an instrument intercomparison (Lacher et al., 2024). The chamber ran continuously, but the campaign lasted only 2 weeks, and the record is therefore not representative.

PUY: The same site as PICNIC, sampled in a later campaign with a different chamber. Air was drawn through a whole-air inlet with passive drying (Canzi et al., 2025). The campaign spans 3 months at 87 % uptime. The record extends no warmer 70 than about −21 °C, which appears to reflect the scanned temperature range. The record is representative of the winter period it covers.

SBO21: A high-alpine summit site at the Sonnblick Observatory, Austria, at 3106 m. The site is above the boundary layer for much of the year and snow-covered for about half the record. Air was drawn through a heated inlet. The campaign spans 15 months at 90 % uptime and covers all 12 months. Of all PINE campaigns, it reaches by far the warmest activation temperatures, 75 above −6 °C, although most 6 h intervals resolve only a single temperature. The proximity of a ski resort, the warm activation temperatures, and INP concentrations that peak in winter together suggest an influence of snow-making. The record is therefore representative of the current local aerosol, but not of natural aerosol conditions.

SPL01: A high-elevation continental site at Storm Peak Laboratory, Colorado, at 3220 m in the Rocky Mountains. The site is snow-covered for roughly three-quarters of the record. The campaign spans 9 months at 68 % uptime. The record is representative. 80

WAO: A coastal site at the Weybourne Atmospheric Observatory on the Norfolk coast, UK. Depending on the wind direction, the site receives either marine air or continental European air. Air was drawn from a mobile laboratory through a heated inlet.

The campaign spans 2 months, from October to December, at 78 % uptime. The record is representative of the autumn period it covers.

85 **S3.2 Long-term land stations measuring INP by filter collection and subsequent ice spectrometry analysis**

SGP: A continental agricultural site on the Southern Great Plains, Oklahoma, part of the ARM long-term program. The site is surrounded by cropland and grassland with strong seasonal biological emission. Day-long filters were collected on a regular 6-day schedule over more than 4 years, covering all 12 months. It is the longest record in the dataset. The record is representative.

90 **EPC:** A coastal marine site at Scripps Pier, La Jolla, California. The site samples onshore flow from the Pacific, but a large urban area lies immediately inland. Day-long filters were collected on a regular 6-day schedule over a full year. The record is representative.

KCG: A remote coastal station at Kennaook/Cape Grim, Tasmania, with the Southern Ocean to the west and Tasmanian farmland to the east. The site therefore samples strongly contrasting air masses depending on the wind direction. The campaign spans 16 months on a nominal 6-day schedule, with filter exposure times ranging from less than an hour to 9 days. Filters were collected in one of two modes. In the *continuous mode*, the pump ran without interruption. In the *clean-sector mode*, the pump was powered only while air masses arrived from the Southern Ocean, so that the filter sampled Southern Ocean air only. The operators chose the mode according to the meteorological conditions and recorded the choice in the collection notes of each filter. Of the 82 filters, 30 were collected in clean-sector mode and 26 in continuous mode; the remaining 26 either span a change of mode or lack a usable note. Because the sampling preferentially targets Southern Ocean air, the record is not representative of the INP conditions at Cape Grim.

OLI: A low-lying Arctic coastal site at Oliktok Point on the North Slope of Alaska, on tundra beside the Beaufort Sea. The site is snow-covered for roughly two-thirds of the record. The campaign spans 9 months of filters with exposure times of 1 to 2 days, collected on an irregular schedule. The record is representative, with the caveat that laboratory blanks rather than field blanks were used, which may inflate the reported concentrations (Creamean et al., 2025).

GUC: A high-elevation mountain site in the Colorado Rockies, snow-covered for roughly three-quarters of the record. A separate tethered-balloon component is described below. The campaign spans 21 months of day-long filters on a nominal 6-day schedule, covering all 12 months. Filters affected by nearby snow-making and cloud-seeding experiments have been removed from the dataset, following Zhou et al. (2026). The record is representative.

110 **MICRE:** A small, remote sub-Antarctic island in the Southern Ocean, thousands of kilometers from any continental dust source. The campaign spans 11 months of alternating 2- and 3-day filters and covers every calendar month. The record is representative.

BNF: A temperate forest site in Bankhead National Forest, Alabama, with strong local biogenic emission. A separate tethered-balloon component is described below. The fixed site contributes day-long filters over about 5 months, but only 16 filters in total. The record is representative of local autumn and winter conditions.

HOU-M1, HOU-S3: Two coastal sites in the Houston area, operated simultaneously through the summer convective season.

HOU-M1 is an urban site influenced by the city and its petrochemical industry. HOU-S3 is a rural site more exposed to onshore flow from the Gulf of Mexico. Both sites collected day-long filters on a 3-day schedule. The records are representative of the convective season, but not of the annual cycle.

- 120 **COMBLE:** A coastal site at Nordmela Harbor on Andøya, Norway, directly exposed to marine air from the Norwegian Sea. The site is snow-covered for about half the record. The campaign spans just under 6 months, with measurements covering 50 % of that time. Sampling targeted cold-air outbreaks: exposure times ranged from 6 to 74 hours and were timed to integrate over individual events. The record is therefore not representative of the site's climatology.

- Schneider-2021:** A boreal forest site at SMEAR II, Hyytiälä, Finland. In the warm season, the aerosol is dominated by local biogenic emission. The ground is snow-covered for about 40 % of the record. The site was monitored continuously over 15 months by consecutive filters with exposure times of roughly 2 days. The record is representative.

Sze-2023: A high-Arctic coastal site at the Villum Research Station in northern Greenland. The site is snow- and ice-covered throughout the record and far from any local dust source. The campaign spans 26 months of filters with a target exposure time of 3.5 days and covers all 12 months. It is one of the cleanest long-term records in the dataset. The record is representative.

- 130 **Tobo-2020:** A megacity site at the Tokyo Skytree, sampling well above street level within a heavily urbanized coastal plain. The campaign spans 12 months of filters with a uniform exposure time of 3 days, collected consecutively. Coverage is therefore effectively continuous, and every filter starts at the same local hour. The record is representative.

- Eckermann-NM:** A coastal Antarctic site at Neumayer Station III, on a permanently ice-covered shelf with essentially no local terrestrial source. The campaign spans 2 years of consecutive fortnightly filters and covers all 12 months. Each filter was analyzed with two instruments and both spectra are retained, so the number of independent samples is half the number of spectra. The record is representative.

- Eckermann-PrElis:** An inland Antarctic site at Princess Elisabeth Station, on permanent ice far from open water. The filters were analyzed with the same two instruments as at Eckermann-NM. The nominal campaign time span of more than a year is misleading: sampling covers two austral summers only, and a single gap accounts for three-quarters of the campaign time span. The exposure times are long and irregular. The record is representative of austral summer conditions, but not of the annual cycle.

- Wex-2019-Alert:** One of four remote Arctic stations reported together by Wex et al. (2019). Alert lies on tundra on northern Ellesmere Island, Canada, far from any local dust source. The site is snow-covered for roughly 40 % of the record. The campaign spans about 1 year of filters with exposure times of typically 1 week and covers a full annual cycle. The record is representative.

Wex-2019-Ny-Alesund: A high-Arctic coastal site at Ny-Ålesund, Svalbard. The site lies in a glaciated grid cell that is snow- and ice-covered throughout, but within reach of the glacial outwash plains that make Svalbard a seasonal dust source. The exposure times are typically week-long, but the record covers only March to September. The record is representative of the light season alone.

- 150 **Wex-2019-Utqiagvik:** The same low-lying Arctic coastal tundra site as ExINP-NSA, beside the Beaufort Sea. The site is snow-covered for roughly two-thirds of the record. The campaign spans about 1 year of filters with exposure times of typically

1 week and covers a full annual cycle. The record is representative.

Wex-2019-Villum: The same high-Arctic station in northern Greenland as Sze-2023, snow- and ice-covered throughout and far from any local dust source. The record consists of 10 week-long filters, deliberately spread over different months across 155 2 years rather than collected continuously. How the sampling week within each month was chosen could not be established. Despite this uncertainty, the record is treated as representative.

S3.3 Shorter land campaigns measuring INP by filter collection and subsequent ice spectrometry analysis

Andøya-2021: A coastal site on Andøya, northern Norway, facing the open Norwegian Sea. The campaign consists of 2 weeks of uniform 40-minute samples in March 2021, collected at a nominal 6-hourly spacing. The campaign period was 160 characterized by frequent cold-air outbreaks reaching the site on northerly winds (Gjelsvik et al., 2025). The record is too short to be representative.

Bergner-2026: Two fjord locations in southern Greenland, Narsarsuaq and Narsaq. Each location had a low-level station and a second station on adjacent high ground. The low-level station at Narsarsuaq stood on a glacial outwash plain, a type of terrain that ranks among the most active high-latitude dust sources. Sampling spans 7 weeks from mid-June to early August 2023, with 165 exposure times of 1 to 3 days and occasional extensions to 2 weeks. Conditions remained calm throughout, no dust emission was observed, and the INP population was dominated by biological and organic material. The record is representative of calm summer conditions at these sites.

Harrison-2022: A windward marine site at Ragged Point, Barbados, sampled from a tower. The site receives Atlantic trade-wind flow carrying long-range transported African dust. The campaign spans 1 month in July and August, the peak Saharan 170 dust season at the site, with irregular exposure times. The record is representative of the dust season, but not of the annual cycle.

Tobo-2016, Tobo-2017: Two short visits to the Zeppelin Observatory near Ny-Ålesund, a high-Arctic ridge site close to the dust-producing glacial outwash plains of Svalbard. The visits were chosen to contrast seasons: 5 days in July 2016 and 6 days in March 2017. Coverage within each visit is continuous, but each contains only a handful of samples. These are case studies 175 rather than time series, and the records are not representative.

ICE-T: A tropical marine site on St. Croix, US Virgin Islands, in the Caribbean trade-wind belt. The record covers 8 days and contains a handful of ground-based samples and one aircraft sample. The record is not representative.

S3.4 Ship-based campaigns

ACE: A circumnavigation of Antarctica, crossing the full range of Southern Ocean conditions from the subtropics to the 180 sea-ice edge. The track is far from continental dust sources for most of its length. The campaign spans 3 months of uniform 8-hour filters, three per day, and is the most regular ship record in the dataset. The record is representative of Southern Ocean conditions.

PS106: A research icebreaker in the Fram Strait and the Arctic Ocean, operating over sea ice and open leads during the melt season. The campaign spans just under 2 months of predominantly 8-hour filters, with measurements covering 96 % of that

185 time. The record is representative of the conditions along the cruise track during the melt season.

MARCUS: Southern Ocean transects aboard an icebreaker between Tasmania and the Antarctic coast, in remote marine air. The campaign spans 4.5 months of roughly day-long filters. Periods in which the wind carried exhaust from the ship's stack to the inlet were removed, so the retained record depends on the wind direction relative to the ship. The record is nevertheless deemed representative of Southern Ocean conditions during austral summer.

190 **CAPRICORN_II:** A research vessel in the Southern Ocean south of Australia, in remote marine air. Filters were mounted on an upper deck about 23 m above sea level, with exposure times alternating between 1 and 2 days. To avoid exhaust contamination, the filter pump was powered only while three conditions were met: the wind arrived over the bow, the ship-relative wind speed was between 10 and 80 kn, and the condensation particle counter read a stable value below 2000 cm^{-3} . Calm conditions are therefore excluded by design, but such conditions are rare in the Southern Ocean. The record is representative
195 of mid-summer Southern Ocean conditions.

Li-2025: A cruise through the Barents, Kara, and Laptev seas, close to the Siberian coast and thus within reach of terrestrial Arctic sources. INP concentrations were measured over 4 weeks with two filters per day. The record is representative of summer conditions in the Arctic Ocean.

MAGIC: A cargo ship between Los Angeles and Honolulu, crossing the remote subtropical North Pacific. The record contains
200 only two day-long filters and is not representative.

SHIPPO: A ship in the western Bering Sea, a high-latitude marine environment. The record contains a single day-long filter and is not representative.

S3.5 Aircraft campaigns

Filter collection from aircraft is selective by construction. Flights are planned to meet a scientific target, and filters are collected
205 over the flight legs that serve that target. Every campaign below collected INP samples during less than 5 % of its nominal campaign time span. None of the aircraft records is representative of ambient conditions.

SOCRATES: Fifteen research flights over the remote Southern Ocean, sampling marine air far from any continental source. Filters were exposed on approximately level flight legs in the marine boundary layer or above cloud top, in conditions selected as favorable for marine stratocumulus.

210 **CAESAR-airborne:** Research flights over Svalbard and the Greenland Sea, sampling marine air flowing off the sea ice during cold-air outbreaks. The record consists of eight flight periods. Filters were exposed on flight legs below cloud, above cloud, or within aerosol layers, and the archive flags the leg type of each sample.

Raif-2024: Research flights over the Norwegian and Barents seas during a cold-air outbreak in March. The sampled Arctic air had crossed the sea ice and in places carried aged haze from lower latitudes. Filters were exposed during nine flight periods.
215 The measured INP concentrations are among the highest recorded in the Arctic.

CACTI: Research flights over the Sierras de Córdoba, Argentina, a semi-arid region of complex terrain with local soil dust sources. The flights were made above a concurrent ground site and targeted orographic convection.

Sanchez-Marroquin-2017: Research flights over southern Iceland, one of the most active high-latitude dust sources, where

glacial outwash plains emit dust episodically under strong winds. The flights were directed into dust plumes to quantify the contribution of Icelandic dust to INP concentrations over the North Atlantic.

Price-2018: Research flights in the Cape Verde region over 2 weeks in August, directly downwind of the Sahara and beneath the Saharan air layer. The flights targeted Saharan dust outflow, and the campaign reports some of the highest INP concentrations in the dataset.

S3.6 Tethered-balloon deployments

Four sites contribute vertically resolved samples collected from tethered balloons (Creamean et al., 2025). Each balloon component forms a separate *deployment* (a unique combination of campaign and platform, Sect. 2). The samples were collected during balloon flights on selected days of intensive operating periods. Measurements cover between well under 1 % and about 7 % of the nominal campaign time span of each record, and each record consists of a handful of separated periods. None of the balloon records is representative of ambient conditions over its nominal campaign time span.

SGP (balloon): The continental agricultural site described above. The record consists of two flight periods within a fortnight, with samples of about 90 minutes.

GUC (balloon): The snow-covered Colorado mountain site described above. The record consists of four flight periods spread over 11 months, with samples of about 2 hours.

BNF (balloon): The temperate forest site described above, flown in spring and early summer when the local biogenic source is active. The record consists of eight flight periods over 3 months, with samples of about an hour.

CRG (balloon): A site in the Baltimore region of Maryland, in a mixed urban and forested landscape. The record consists of eight flight periods over 5 months, with samples of about an hour. A single gap accounts for nearly 90 % of the nominal campaign time span, so the record effectively consists of two separate visits.

S4 Lag-bin placement by peak finding

240 Calculating the autocorrelation requires sorting sample pairs into bins according to their time lag. When sampling is regular, or partly regular, the pairwise lags do not spread smoothly but pile up at integer multiples of the sampling interval. A campaign sampling every 6 days, for example, produces almost all of its pairs at lags of 6, 12, 18, ... days, with nearly empty stretches in between. Using a fixed binning system results in many sparsely or asymmetrically populated bins. To avoid the inaccuracies resulting from a fixed binning scheme, an algorithmic approach is used that centers bins on prominent peaks in the time-lag
245 distribution.

Step 1: Pooling the pairwise lags.

Let $t_1, t_2, \dots, t_K$ be the sampling times of one campaign. Every pair of samples (i, j) with $i < j$ contributes one lag $\tau_{ij} = |t_j - t_i|$. Lags longer than $\tau_{\max} = \frac{3}{4}(\max_i t_i - \min_i t_i)$ are discarded, i.e., lags longer than three-quarters of the time between
250 the first and the last sample. This clipping is used because at lags near $\tau_{\max}$, only samples near the beginning of the campaign can pair with samples near its end, so that structures on shorter timescales that happen to be present there dominate the signal. All remaining steps work with the base-10 logarithm of the lag, i.e., $x = \log_{10} \tau$.

Step 2: Producing a fine histogram of lags.

255 The values x are sorted into a fine histogram with 80 cells per decade. The count in the k th cell is denoted h_k . This histogram is only a device for locating clusters and is much finer than the final bins.

Step 3: Separating spikes from the smooth continuum.

h_k is the sum of two components: sharp spikes where the sampling interval piles lags onto a single value, and a smooth continuum originating from inconsistent sampling intervals. h_k grows smoother at long lags since the time interval spanned by each
260 bin grows in proportion to the lag by design. To quantify how much a spike exceeds the local continuum, the smooth part is estimated and removed. The estimate of the smooth part, $\tilde{h}_k$, is a rolling median of h_k with a window 0.8 decades wide (32 histogram cells to either side). A median is used rather than a mean because a narrow spike barely moves the median of its neighborhood, so a spike does not inflate the very baseline it is measured against. $e_k = \max(h_k - \tilde{h}_k, 0)$ is the excess of cell k
265 above the local continuum.

Step 4: Locating peaks.

The excess is judged *relative* to the continuum,

$$\hat{e}_k = \frac{e_k}{\max(\tilde{h}_k, 1)}. \quad (\text{S1})$$

270 A histogram cell is accepted as a peak when $\hat{e}_k \geq 0.08$ and when it is at least 0.04 decades away from any larger accepted peak.

Step 5: Separating the structured and continuous regimes.

At short lags where the temporal resolution is high, clusters of pairs with equal lags are well resolved, whereas they merge into a continuum at long lags due to the logarithmic spacing. The peak-finding approach described up to this point works well in the structured short-lag regime, but not in the continuum. The structured regime is identified as the longest run of consecutive peaks whose neighboring gaps are all at most 0.12 decades wide, after which the algorithm transitions to uniform log-space binning, placing 15 bin centers per decade, in what is referred to as the continuous regime.

Step 6: Placing edges between centers.

Let $x_1 < x_2 < \dots < x_n$ be the final sorted centers (detected peaks of the structured part and uniformly placed bin centers of the continuous regime combined). The edges are placed halfway between neighboring centers in log space.

Binning GloPINE campaigns.

GloPINE campaigns bypass the peak finder entirely since their 6-hourly sampling is known a priori, so bins are centered on multiples of 0.25 d out to a lag of 5 d, and longer lags are binned uniformly in log space.

Examples.

Figure S3a shows the autocorrelation function of the campaign Schneider-2021 calculated within each bin, as determined by the algorithm described above, and plotted at the bin centers. A similar example is shown for ExINP-ZEP in Fig. S3b. However, because ExINP-ZEP is a GloPINE campaign, the bin centers in the structured part of the time axis are hard-coded to every 6 hours.

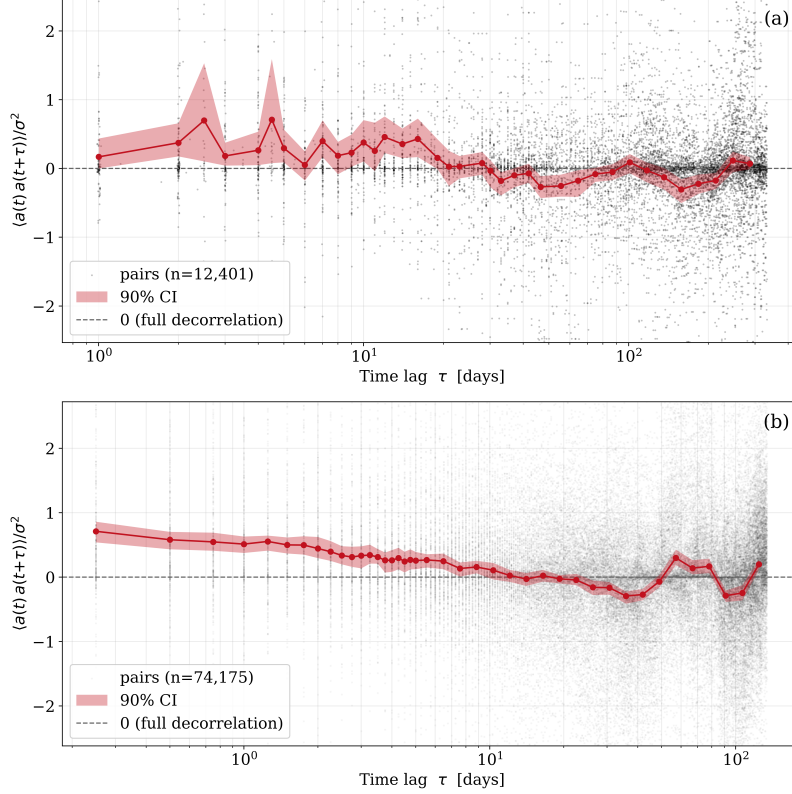


Figure S3. The autocorrelation of two example *deployments* (unique combinations of campaign and platform). Each dot is a single sample pair at its exact time lag τ , showing the variance-normalized product $(y_i - \bar{y})(y_j - \bar{y})/s^2$ of the two samples i and j , where y_i is $\log_{10}(\text{INP}^{\text{proxy}})$ at $T_{\text{ref}} = -23^\circ\text{C}$ of sample i , and $\bar{y}$ and s^2 are the measured mean and variance of the deployment. The red curve is the lag-binned mean on the deployment's own bins and the shaded band its 90 % confidence interval from bootstrap resampling of measurements with replacement (Wilks, 2006). (a) Schneider-2021 (IS): bins are placed by the peak-finding procedure of Sect. S4, with the sampling interval visible as vertical stripes of pairs at integer multiples of 1 day. (b) ExINP-ZEP (PINE): bins follow the fixed 6 h grid for the first 5 days, then transition to being uniformly spaced in log space.

S5 Dependence of the autocovariance of sample anomalies on the exposure time

This section justifies the assumption $C_{L_{\text{ref}}}(\tau) = C_{L_0}(\tau)$ made in Sect. B5; i.e., that the autocovariance of the exposure means
 295 of $a(t)$,

$$C_L(\tau) = \langle \bar{a}_L(t) \bar{a}_L(t + \tau) \rangle, \quad \bar{a}_L(t) = \frac{1}{L} \int_t^{t+L} a(t') dt', \quad (\text{S2})$$

as defined in Eqs. (B21) and (B3), takes the same value for the exposure time L_0 of a record and the reference exposure
 time L_{ref} . Section S5.1 writes $C_L(\tau)$ as a weighted average of C over the lags within $\pm L$ of τ and expands the difference
 $C_L(\tau) - C(\tau)$ in powers of L . Section S5.2 splits the variability of $a(t)$ into a slow and a fast part, which are treated in
 300 Sects. S5.3 and S5.4. Section S5.5 treats periodic variability, which is the one exception where the assumption is not accurate.

S5.1 The covariance of exposure means as a weighted average of the autocovariance

Throughout, $\langle \cdot \rangle$ is the average over the long interval $[0, \Theta]$ defined in Eq. (B1), and $a(t)$ is stationary with $\langle a \rangle = 0$, variance
 σ^2 , and autocovariance $C(\tau)$ as defined in Eq. (B2). As in Eq. (B5), shifting the averaging interval by an amount that is small
 compared with Θ changes the average by a relative amount of that order, which is neglected.

305 Substituting $t' = t + s$ in the exposure mean of Eq. (S2) gives

$$\bar{a}_L(t) = \frac{1}{L} \int_0^L a(t + s) ds, \quad \bar{a}_L(t + \tau) = \frac{1}{L} \int_0^L a(t + \tau + s') ds', \quad (\text{S3})$$

where the second exposure mean is the first with t replaced by $t + \tau$ and the integration variable renamed s' . Thus s is the
 position of an instant within the first exposure, measured from its start, and s' the same for the second exposure. The product
 of the two integrals is the double integral of the product of the integrands:

$$310 \quad \bar{a}_L(t) \bar{a}_L(t + \tau) = \frac{1}{L^2} \int_0^L \int_0^L a(t + s) a(t + \tau + s') ds' ds. \quad (\text{S4})$$

Taking the average $\langle \cdot \rangle$ of Eq. (S4) means integrating over t from 0 to Θ and dividing by Θ . The limits of the integrals over s
 and s' do not depend on t , so the integration over t can be carried out first:

$$C_L(\tau) = \frac{1}{L^2} \int_0^L \int_0^L \left[\frac{1}{\Theta} \int_0^\Theta a(t + s) a(t + \tau + s') dt \right] ds' ds. \quad (\text{S5})$$

In the bracket, the two instants $t + s$ and $t + \tau + s'$ are separated by the lag $\tau + s' - s$, which does not depend on t . Substituting
 315 $t'' = t + s$ and neglecting the shift of the averaging interval by s ,

$$\frac{1}{\Theta} \int_0^\Theta a(t + s) a(t + \tau + s') dt = \frac{1}{\Theta} \int_0^\Theta a(t'') a(t'' + \tau + s' - s) dt'' = C(\tau + s' - s) \quad (\text{S6})$$

by the definition of C in Eq. (B2). Hence

$$C_L(\tau) = \frac{1}{L^2} \int_0^L \int_0^L C(\tau + s' - s) ds' ds. \quad (\text{S7})$$

In words, the covariance of two exposure means is the average of the autocovariance over all pairs of instants, one instant from each exposure.

Substituting $u = s' - s$ for s' in Eq. (S7) gives

$$C_L(\tau) = \frac{1}{L^2} \int_0^L \int_{-s}^{L-s} C(\tau + u) du ds. \quad (\text{S8})$$

The region of integration, $0 \leq s \leq L$ and $-s \leq u \leq L - s$, is a parallelogram in the (s, u) plane that lies between $u = -L$ and $u = L$. Cutting it at $u = 0$, for $-L \leq u \leq 0$ the variable s runs from $-u$ to L , and for $0 \leq u \leq L$ it runs from 0 to $L - u$.

Swapping the order of integration therefore gives

$$C_L(\tau) = \frac{1}{L^2} \left[\int_{-L}^0 \int_{-u}^L C(\tau + u) ds du + \int_0^L \int_0^{L-u} C(\tau + u) ds du \right]. \quad (\text{S9})$$

The integrand does not depend on s , so the inner integrals equal the lengths of the s intervals, $L + u$ and $L - u$:

$$C_L(\tau) = \frac{1}{L^2} \left[\int_{-L}^0 (L + u) C(\tau + u) du + \int_0^L (L - u) C(\tau + u) du \right] = \frac{1}{L^2} \int_{-L}^L (L - |u|) C(\tau + u) du. \quad (\text{S10})$$

With the triangular weight

$$\Lambda_L(u) \equiv \frac{L - |u|}{L^2} \quad \text{for } |u| \leq L, \quad \int_{-L}^L \Lambda_L(u) du = 1, \quad (\text{S11})$$

Eq. (S10) reads

$$C_L(\tau) = \int_{-L}^L \Lambda_L(u) C(\tau + u) du. \quad (\text{S12})$$

Thus $C_L(\tau)$ is a weighted average of C over the lags from $\tau - L$ to $\tau + L$, with the largest weight at τ and weights decreasing linearly to zero at $\tau \pm L$ (Fig. S4a).

Because Λ_L has unit area, the constant $C(\tau)$ equals $\int_{-L}^L \Lambda_L(u) C(\tau) du$, and subtracting this from Eq. (S12) gives

$$C_L(\tau) - C(\tau) = \int_{-L}^L \Lambda_L(u) [C(\tau + u) - C(\tau)] du. \quad (\text{S13})$$

Assume that C is smooth on the interval $[\tau - L, \tau + L]$ covered by the weight, and expand $C(\tau + u)$ in a Taylor series in u about $u = 0$, i.e., about the lag τ :

$$C(\tau + u) - C(\tau) = C'(\tau)u + \frac{C''(\tau)}{2}u^2 + \frac{C'''(\tau)}{6}u^3 + \frac{C''''(\tau)}{24}u^4 + \dots, \quad (\text{S14})$$

340 where the primes denote derivatives with respect to the lag, evaluated at τ . Inserting Eq. (S14) into Eq. (S13), the derivatives do not depend on u and can be taken outside the integrals:

$$C_L(\tau) - C(\tau) = C'(\tau)M_1 + \frac{C''(\tau)}{2}M_2 + \frac{C'''(\tau)}{6}M_3 + \frac{C''''(\tau)}{24}M_4 + \dots, \quad (\text{S15})$$

where

$$M_n \equiv \int_{-L}^L u^n \Lambda_L(u) du \quad (\text{S16})$$

345 are the moments of the triangular weight. Since Λ_L is even, $u^n \Lambda_L(u)$ is odd for odd n , and all odd moments vanish. For even n the integrand is even, and

$$M_n = \frac{2}{L^2} \int_0^L u^n (L - u) du = \frac{2}{L^2} \left[\frac{L^{n+2}}{n+1} - \frac{L^{n+2}}{n+2} \right] = \frac{2L^n}{(n+1)(n+2)}, \quad n \text{ even}, \quad (\text{S17})$$

so that $M_2 = L^2/6$ and $M_4 = L^4/15$. Equation (S15) becomes

$$C_L(\tau) - C(\tau) = \frac{L^2}{12} C''(\tau) + \frac{L^4}{360} C''''(\tau) + \dots \quad (\text{S18})$$

350 Thus, $C_L(\tau)$ differs significantly from $C(\tau)$ only where C has curvature on the scale L near τ .

S5.2 Splitting the variability into a slow and a fast part

Split $a(t)$ into a slow part $a_s(t)$, which holds the variability on timescales long compared with the exposure time L , and a fast part $a_f(t)$, which holds the rest, i.e., the variability on timescales up to about L :

$$a(t) = a_s(t) + a_f(t). \quad (\text{S19})$$

355 The boundary between the two parts is not sharp, and variability on timescales comparable to L belongs to neither part cleanly; this is why the results below hold only approximately.

The slow and the fast variability are assumed to be uncorrelated, so that the two parts contribute additively to the autocovariance, and likewise to the autocovariance of the exposure means:

$$C(\tau) = C_s(\tau) + C_f(\tau), \quad C_L(\tau) = C_{s,L}(\tau) + C_{f,L}(\tau), \quad (\text{S20})$$

360 where C_s and C_f are the autocovariances of a_s and a_f as defined in Eq. (B2), and $C_{s,L}$ and $C_{f,L}$ are the autocovariances of their exposure means as defined in Eq. (S2). Finally, Eq. (S12) was derived for an arbitrary stationary anomaly, so it applies to

a_s and a_f separately:

$$C_{s,L}(\tau) = \int_{-L}^L \Lambda_L(u) C_s(\tau + u) du, \quad C_{f,L}(\tau) = \int_{-L}^L \Lambda_L(u) C_f(\tau + u) du. \quad (\text{S21})$$

Subtracting the first relation of Eq. (S20) from the second,

$$365 \quad C_L(\tau) - C(\tau) = [C_{s,L}(\tau) - C_s(\tau)] + [C_{f,L}(\tau) - C_f(\tau)], \quad (\text{S22})$$

and the two brackets can be examined separately.

S5.3 The slow part

For the slow part, Eq. (S18) applied to C_s gives

$$C_{s,L}(\tau) - C_s(\tau) = \frac{L^2}{12} C_s''(\tau) + \frac{L^4}{360} C_s''''(\tau) + \dots \quad (\text{S23})$$

370 The size of the leading term is estimated by a scaling argument. Let $\hat{C}$ denote the typical magnitude of C_s near the lag τ , and let t_s denote the timescale over which C_s changes by an amount of order $\hat{C}$ there. A function that changes by $\hat{C}$ over an interval of length t_s has a second derivative of order $\hat{C}/t_s^2$, and a fourth derivative of order $\hat{C}/t_s^4$. Inserting this scaling into Eq. (S23) and dividing by $\hat{C}$,

$$\frac{|C_{s,L}(\tau) - C_s(\tau)|}{\hat{C}} \sim \frac{1}{12} \left(\frac{L}{t_s}\right)^2 + \frac{1}{360} \left(\frac{L}{t_s}\right)^4 + \dots \quad (\text{S24})$$

375 The right-hand side is an expansion in $(L/t_s)^2$, and the leading term shows the condition: the exposure average leaves C_s unchanged, to a relative accuracy of $(L/t_s)^2/12$, wherever the timescale t_s on which C_s varies is long compared with the exposure time L . Therefore:

$$C_{s,L}(\tau) \approx C_s(\tau) \quad \text{at every lag } \tau, \quad (\text{S25})$$

and the first bracket of Eq. (S22) vanishes.

380 S5.4 The fast part

For the fast part, the Taylor expansion is of no use, because C_f varies on lag scales shorter than L . Instead, apply the substitution $\tau' = \tau + u$ to Eq. (S21):

$$C_{f,L}(\tau) = \int_{-L}^L \Lambda_L(u) C_f(\tau + u) du = \int_{\tau-L}^{\tau+L} \Lambda_L(\tau' - \tau) C_f(\tau') d\tau', \quad (\text{S26})$$

385 which shows that $C_{f,L}(\tau)$ is given by the overlap of C_f with a triangular window centered at the lag τ and reaching from $\tau - L$ to $\tau + L$ (Fig. S4b).

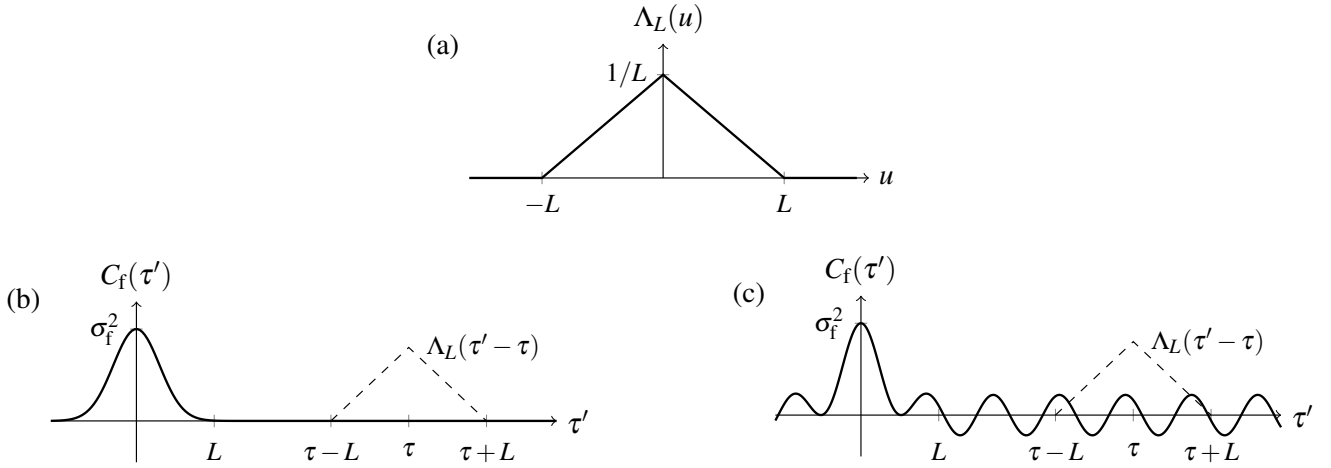


Figure S4. (a) The triangular weight $\Lambda_L(u)$ of Eq. (S11). (b) The autocovariance $C_f(\tau')$ of non-periodic fast variability (solid), which vanishes beyond lags of about L , and the window $\Lambda_L(\tau' - \tau)$ (dashed) over which Eq. (S26) integrates $C_f(\tau')$ at the lag τ . For $\tau - L > L$ the window and C_f do not overlap, and $C_{f,L}(\tau) = 0$. (c) As (b), but with a periodic component added to the fast part. Its contribution to C_f oscillates with undiminished amplitude at every lag, the window overlaps it at every τ , and $C_{f,L}(\tau) \neq 0$.

The fast part holds the variability on timescales up to about L . If this variability is not periodic, none of its fluctuations lasts longer than about L , so the fast part is uncorrelated between instants more than about L apart:

$$C_f(\tau') \approx 0 \quad \text{for } |\tau'| \gtrsim L. \quad (\text{S27})$$

As shown in Fig. S4b, the window in Eq. (S26) covers the lags from $\tau - L$ to $\tau + L$. If its lower edge lies beyond L , the
 390 integrand vanishes throughout the window:

$$C_{f,L}(\tau) \approx 0 \quad \text{for } \tau \gtrsim 2L, \quad (\text{S28})$$

and by Eq. (S27), $C_f(\tau) \approx 0$ at these lags too, so the second bracket of Eq. (S22) vanishes for $\tau \gtrsim 2L$. Inserting Eqs. (S25) and (S28) into Eq. (S22),

$$C_L(\tau) \approx C(\tau) \quad \text{for } \tau \gtrsim 2L, \quad (\text{S29})$$

395 for any exposure time L , provided the variability of $a(t)$ on timescales up to about L is not periodic. So the condition $C_{L_{\text{ref}}}(\tau) = C_{L_0}(\tau)$ holds when $\tau > \max(2L_0, 2L_{\text{ref}})$, because then by Eq. (S29), $C_{L_0}(\tau) \approx C_{L_{\text{ref}}}(\tau) \approx C(\tau)$.

S5.5 Periodic variability

Equation (S27) fails for periodic variability. A periodic component of $a(t)$ with a period P no longer than L varies on a timescale no longer than L and so belongs to the fast part, but it does not decorrelate: its contribution to C_f oscillates with

400 the period P and with undiminished amplitude at every lag (Fig. S4c). The window of Eq. (S26) therefore overlaps it at every τ , $C_{f,L}(\tau) \neq 0$ at every lag, and Eq. (S29) does not hold. In the current application of INP measurements with exposure time $6\text{ h} \leq L_0 \leq 14\text{ d}$, the only relevant candidate for a significant periodic cycle is the diurnal cycle.

Let $a_p(t)$ denote the periodic component of $a(t)$, and let σ_p^2 denote its variance. The non-periodic component of $a(t)$ satisfies Eq. (S29), so for $\tau \gtrsim \max(2L_0, 2L_{\text{ref}})$ this component contributes equally to $C_{L_0}(\tau)$ and $C_{L_{\text{ref}}}(\tau)$, thus causing no difference
 405 between them. The contribution of a_p to both $C_{L_0}(\tau)$ and $C_{L_{\text{ref}}}(\tau)$ is bounded by σ_p^2 , because the autocovariance of any series is bounded in magnitude by its variance. Hence

$$|C_{L_0}(\tau) - C_{L_{\text{ref}}}(\tau)| \lesssim 2\sigma_p^2 \quad \text{for } \tau \gtrsim \max(2L_0, 2L_{\text{ref}}). \quad (\text{S30})$$

To see how the difference $C_{L_0}(\tau) - C_{L_{\text{ref}}}(\tau) = \delta(\tau)$ enters the adjustment of the autocorrelation, use that the adjusted autocorrelation of Eq. (B26), $A\rho_0(\tau) + S$, is by Eqs. (B23) and (B12) and the definition of S identical to the right-hand side of
 410 Eq. (B24) with C_{L_0} in place of $C_{L_{\text{ref}}}$. Inserting $C_{L_0}(\tau) = C_{L_{\text{ref}}}(\tau) + \delta(\tau)$ there,

$$A\rho_0(\tau) + S = \frac{C_{L_{\text{ref}}}(\tau) + \delta(\tau) - \sigma^2 + V(R_{\text{ref}})}{V(R_{\text{ref}}) - V(L_{\text{ref}})} = \rho_{\text{ref}}(\tau) + \frac{\delta(\tau)}{\langle s_{\text{ref}}^2 \rangle}, \quad (\text{S31})$$

where the last step uses Eq. (B24) and $\langle s_{\text{ref}}^2 \rangle = V(R_{\text{ref}}) - V(L_{\text{ref}})$ from Eq. (B12). The adjusted autocorrelation thus deviates from the true $\rho_{\text{ref}}(\tau)$ by $\delta(\tau)$ divided by the expected variance of a reference record. With $|\delta(\tau)| \lesssim 2\sigma_p^2$ from Eq. (S30), the deviation is at most $2\sigma_p^2/\langle s_{\text{ref}}^2 \rangle$, twice the variance of the periodic component relative to the variance of a reference record.
 415 Since ρ_{ref} is a correlation coefficient, bounded by one in magnitude, the assumption $C_{L_{\text{ref}}}(\tau) = C_{L_0}(\tau)$ is adequate for $\tau > \max(2L_0, 2L_{\text{ref}})$ if $\sigma_p^2 \ll \langle s_{\text{ref}}^2 \rangle$, i.e., if the periodic variability with periods up to $\max(L_0, L_{\text{ref}})$, in practice the diurnal cycle, holds only a small share of the variance of a reference record.

S6 Effect of the adjustment on the deviation of a record from its expected value

Table S1. Symbols used in Sects. S6 and S7 that are not used in Appendix B of the main text, with the equations or sections in which they are defined. For the symbols introduced in Appendix B, refer to its symbol table, Table B1.

<i>Deviations from expected values, Sect. S6</i>	
$\delta_s, \delta_f(\Delta t), \delta_r(\tau), \delta_y$	deviations of $s_0^2, f_0(\Delta t), r_0(\tau)$, and $\bar{y}_0$ from their expected counterparts, Eq. (S33)
<i>Measurement noise, Sect. S7</i>	
$z_i, \eta_i, \sigma_\eta^2$	measured sample anomaly, its measurement noise, and the variance of the noise, Sect. S7
$\bar{z}, \bar{b}, \bar{\eta}$	means of z_i, b_i , and η_i over the samples inside a window, Eqs. (S40) and (S41)
$s_{\Delta t, z}^2$	variance of the measured sample anomalies inside a window of width Δt , Eq. (S40)
$w_p(\Delta t), W_z(\Delta t)$	measured accumulated variance of record p (Sect. 3.5), here the mean of $s_{\Delta t, z}^2$ over all windows of width Δt , and its expected value, Eq. (S44)
$s_{0, z}^2, s_{\text{ref}, z}^2$	measured variance of a whole record including the noise, and its value adjusted to the reference sampling, $s_{0, z}^2/A_z$, Eqs. (S47) and (S49)
A_z	A computed with W_z in place of V , Eq. (S48)

The notation follows Appendix B of the main text, whose symbols are listed in Table B1. The symbols introduced in this section and in Sect. S7 are listed in Table S1. The boxed relations of Sects. B3 to B6 hold between expected values. Applied to a single record, they are used in the forms

$$s_{\text{ref}}^2 = \frac{s_0^2}{A}, \quad f_{\text{ref}}(\Delta t) = A f_0(\Delta t) + B, \quad r_{\text{ref}}(\tau) = A r_0(\tau) + S, \quad \bar{y}_{\text{ref}} = \bar{y}_0 - \kappa B s_{\text{ref}}^2. \quad (\text{S32})$$

This section shows how Eq. (S32) changes the deviation of a record from its expected value. Let these deviations be denoted

$$\delta_s \equiv s_0^2 - \langle s_0^2 \rangle, \quad \delta_f(\Delta t) \equiv f_0(\Delta t) - F_0(\Delta t), \quad \delta_r(\tau) \equiv r_0(\tau) - \rho_0(\tau), \quad \delta_y \equiv \bar{y}_0 - \langle \bar{y}_0 \rangle, \quad (\text{S33})$$

for the variance, normalized accumulated variance, autocorrelation, and the mean value of $\log_{10} \text{INP}(t)$, respectively. The first term on the right-hand side of each definition in Eq. (S33) is computed from the samples of the record and therefore depends on its start date t_1 , whereas the second term is an average over all start dates and thus independent of t_1 . The purpose of this section is to derive the relations between the deltas of Eq. (S33), i.e., the deviations from the expected values at the native sampling, and the deviations from the expected values after adjustment to the reference sampling.

For the variance, inserting $s_0^2 = \langle s_0^2 \rangle + \delta_s$ into Eq. (S32) and using $\langle s_{\text{ref}}^2 \rangle = \langle s_0^2 \rangle / A$ from Eq. (B12) gives

$$s_{\text{ref}}^2 = \frac{\langle s_0^2 \rangle}{A} + \frac{\delta_s}{A} = \langle s_{\text{ref}}^2 \rangle + \frac{\delta_s}{A}. \quad (\text{S34})$$

Subtracting $\langle s_{\text{ref}}^2 \rangle$ and dividing by $\langle s_{\text{ref}}^2 \rangle = \langle s_0^2 \rangle / A$ yields

$$\frac{s_{\text{ref}}^2 - \langle s_{\text{ref}}^2 \rangle}{\langle s_{\text{ref}}^2 \rangle} = \frac{s_0^2 - \langle s_0^2 \rangle}{\langle s_0^2 \rangle}, \quad (\text{S35})$$

i.e., the adjusted variance deviates from its expected value by the same relative amount as the measured variance deviates from
435 its expected value.

For the normalized accumulated variance, inserting $f_0(\Delta t) = F_0(\Delta t) + \delta_f(\Delta t)$ into Eq. (S32) and using Eq. (B16) gives

$$f_{\text{ref}}(\Delta t) = A F_0(\Delta t) + B + A \delta_f(\Delta t) = F_{\text{ref}}(\Delta t) + A \delta_f(\Delta t) \Rightarrow f_{\text{ref}}(\Delta t) - F_{\text{ref}}(\Delta t) = A [f_0(\Delta t) - F_0(\Delta t)], \quad (\text{S36})$$

so the deviation of the adjusted curve from its expected value is the deviation of the measured curve multiplied by A , i.e., the deviation is conserved up to the factor A . This factor enters because the curve is normalized by the expected variance of
440 the record, Eq. (B14), which by Eq. (B12) is $\langle s_0^2 \rangle$ under the native sampling and $\langle s_0^2 \rangle / A$ under the reference sampling. One consequence is visible in Fig. 6b. For a record with $L_0 \leq L_{\text{ref}}$, the width $\Delta t = L_{\text{ref}}$ lies within the range of Eq. (B16), and since $F_{\text{ref}}(L_{\text{ref}}) = 0$, Eq. (S36) gives $f_{\text{ref}}(L_{\text{ref}}) = A \delta_f(L_{\text{ref}})$, so that the adjusted curves of single records start at zero only on average.

For the autocorrelation, inserting $r_0(\tau) = \rho_0(\tau) + \delta_r(\tau)$ into Eq. (S32) and using Eq. (B26) gives

$$445 \quad r_{\text{ref}}(\tau) = A \rho_0(\tau) + S + A \delta_r(\tau) = \rho_{\text{ref}}(\tau) + A \delta_r(\tau) \Rightarrow r_{\text{ref}}(\tau) - \rho_{\text{ref}}(\tau) = A [r_0(\tau) - \rho_0(\tau)], \quad (\text{S37})$$

so the deviation of the autocorrelation is likewise conserved up to the factor A , which enters because the autocorrelation is also normalized by the expected variance of the record, Eq. (B23).

For the mean, inserting $\bar{y}_0 = \langle y_0 \rangle + \delta_y$ and Eq. (S34) into Eq. (S32) and using Eq. (B34) gives

$$\bar{y}_{\text{ref}} = \langle y_0 \rangle + \delta_y - \kappa B \left[\langle s_{\text{ref}}^2 \rangle + \frac{\delta_s}{A} \right] = \langle y_{\text{ref}} \rangle + \delta_y - \frac{\kappa B}{A} \delta_s, \quad (\text{S38})$$

450 which, written out, reads

$$\bar{y}_{\text{ref}} - \langle y_{\text{ref}} \rangle = \left(\bar{y}_0 - \langle y_0 \rangle \right) - \frac{\kappa B}{A} \left(s_0^2 - \langle s_0^2 \rangle \right). \quad (\text{S39})$$

The deviation of the measured mean enters the deviation of the adjusted mean unscaled. Apart from the second term on the right-hand side, the adjusted mean therefore deviates from its expected value by the same absolute amount as the measured mean deviates from its own. The second term arises because the subtracted term is computed from the variance of the record,
455 $\kappa B s_0^2 / A$, rather than from its expected value, $\kappa B \langle s_0^2 \rangle / A$. It vanishes for $L_0 = L_{\text{ref}}$, where $B = 0$.

In summary, the adjustment conserves the relative deviation of the variance, Eq. (S35). It conserves the deviations of the normalized accumulated variance and of the autocorrelation up to the factor A , Eqs. (S36) and (S37), which enters because both are normalized by the expected variance of the record. The adjustment of the mean conserves the absolute deviation of the mean apart from a term proportional to the deviation of the variance, Eq. (S39). In this sense the adjustment of the measured
460 statistics to a reference sampling removes the systematic effect of (L, R) and conserves the influence of the start date of the record.

S7 An upper bound on the measurement noise

The notation follows Appendix B of the main text, whose symbols are listed in Table B1, and the symbols introduced in this section are listed in Table S1 of Sect. S6. This section bounds the measurement noise of the records of 6 h samples from which the curve of Fig. 4 is built, and the share of the reference variance that this noise represents. The idea is that noise inflates the variance inside every window of a record, and that this inflation can be read off the curve at its shortest width, where the geophysical variance is smallest and the noise most visible.

Let a measured sample anomaly of one of these 6 h records be denoted $z_i = b_i + \eta_i$, where b_i is the noise-free sample anomaly as defined in Sect. B1 and η_i is the contribution of the measurement noise. The noise is assumed to have zero mean and variance σ_η^2 , to be independent of $a(t)$, and to be uncorrelated between samples. Consider a window of width Δt inside the record, which holds $n = \Delta t/L_0$ samples. Let $\bar{z}$ denote the mean of the n measured sample anomalies inside the window and let $s_{\Delta t, z}^2$ denote their variance as defined in Eq. (B7). Expanding the square as in Eq. (B10) gives

$$s_{\Delta t, z}^2 = \frac{1}{n} \sum_{i=1}^n z_i^2 - \bar{z}^2, \quad \text{where} \quad \bar{z} = \frac{1}{n} \sum_{i=1}^n z_i. \quad (\text{S40})$$

Inserting $z_i = b_i + \eta_i$ into the two expressions of Eq. (S40) and letting $\bar{b}$ and $\bar{\eta}$ denote the means of b_i and η_i over the window yields

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n z_i^2 &= \frac{1}{n} \sum_{i=1}^n b_i^2 + \frac{2}{n} \sum_{i=1}^n b_i \eta_i + \frac{1}{n} \sum_{i=1}^n \eta_i^2, \quad \text{and} \quad \bar{z}^2 = \bar{b}^2 + 2\bar{b}\bar{\eta} + \bar{\eta}^2 \\ \Rightarrow \langle s_{\Delta t, z}^2 \rangle &= \frac{1}{n} \sum_{i=1}^n \langle b_i^2 - \bar{b}^2 \rangle + \frac{1}{n} \sum_{i=1}^n \langle \eta_i^2 - \bar{\eta}^2 \rangle + \frac{2}{n} \sum_{i=1}^n \langle b_i \eta_i \rangle - \langle 2\bar{b}\bar{\eta} \rangle, \end{aligned} \quad (\text{S41})$$

where the second line is the expected value of Eq. (S40) over all window positions. The first term of that line is the expected variance of the noise-free sample anomalies inside the window, which is $V(\Delta t) - V(L_0)$ by Eq. (B13). The last two terms vanish, because η_i is assumed to be independent of b_i and to have zero mean. The second term is the expected variance of the noise inside the window. Its first part is $\langle \eta_i^2 \rangle = \sigma_\eta^2$, and its second part is evaluated by writing $\bar{\eta}^2$ as a double sum:

$$\bar{\eta}^2 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \eta_i \eta_j. \quad (\text{S42})$$

Since the noise is uncorrelated between samples, only the n terms with $i = j$ have a non-zero expected value, and each of them has the expected value σ_η^2 , so that $\langle \bar{\eta}^2 \rangle = \sigma_\eta^2/n$. The second line of Eq. (S41) thus becomes

$$\langle s_{\Delta t, z}^2 \rangle = V(\Delta t) - V(L_0) + \sigma_\eta^2 - \frac{\sigma_\eta^2}{n}. \quad (\text{S43})$$

The left-hand side of Eq. (S43) is the accumulated variance of the record as measured, which is analogous to Eq. (B13) but with measurement noise included. This accumulated variance is denoted $W_z(\Delta t)$. For a single record p , let $w_p(\Delta t)$ denote the mean of $s_{\Delta t, z}^2$ over all windows of width Δt inside the record, which estimates $W_z(\Delta t)$. The gray curves of Fig. 4 are

$w_p(\Delta t)/w_p(7\text{d})$ of the individual records, and their mean, which Sect. 3.5 uses as the estimate of $\mathcal{W}(\Delta t)$, thus estimates
490 $W_z(\Delta t)/W_z(7\text{d})$, which coincides with $\mathcal{W}(\Delta t)$ for $\sigma_\eta^2 = 0$. With $n = \Delta t/L_0$, Eq. (S43) reads

$$W_z(\Delta t) \equiv \langle s_{\Delta t, z}^2 \rangle = V(\Delta t) - V(L_0) + \left(1 - \frac{L_0}{\Delta t}\right) \sigma_\eta^2. \quad (\text{S44})$$

The noise thus raises the accumulated variance by its own variance σ_η^2 , minus the part $L_0\sigma_\eta^2/\Delta t$, which is absorbed by the window mean. At $\Delta t = 2L_0$, where a window holds two samples, the factor $1 - L_0/\Delta t$ equals $1/2$ and Eq. (S44) gives

$$W_z(2L_0) = V(2L_0) - V(L_0) + \frac{\sigma_\eta^2}{2} \geq \frac{\sigma_\eta^2}{2}, \quad (\text{S45})$$

495 where the inequality holds because $V(2L_0) - V(L_0)$ is the expected variance of two consecutive noise-free sample anomalies by Eq. (B13), which cannot be negative. Hence

$$\sigma_\eta^2 \leq 2W_z(2L_0). \quad (\text{S46})$$

The bound on σ_η^2 is now carried to the reference variance that is reported for these records. Let $s_{0,z}^2$ denote the measured variance of a whole record, which is the variance inside the window of width R_0 that the record spans, so that its expected
500 value is $W_z(R_0)$ by the definition of W_z . Equation (S44) at $\Delta t = R_0$ gives

$$\langle s_{0,z}^2 \rangle = W_z(R_0) = V(R_0) - V(L_0) + \left(1 - \frac{L_0}{R_0}\right) \sigma_\eta^2 = \langle s_0^2 \rangle + \left(1 - \frac{L_0}{R_0}\right) \sigma_\eta^2, \quad (\text{S47})$$

where Eq. (B12) was used for $\langle s_0^2 \rangle$. The reference variance reported for the record is $s_{\text{ref},z}^2 = s_{0,z}^2/A_z$, where A_z is the constant A of Eq. (B12) computed from the measured curve, that is, with W_z in place of V . Since $W_z(L_0) = 0$ by Eq. (S44),

$$A_z = \frac{W_z(R_0) - W_z(L_0)}{W_z(R_{\text{ref}}) - W_z(L_{\text{ref}})} = \frac{W_z(R_0)}{W_z(R_{\text{ref}}) - W_z(L_{\text{ref}})}, \quad (\text{S48})$$

505 and the expected value of the reported reference variance follows from Eqs. (S47) and (S48), with Eq. (S44) evaluated at $\Delta t = R_{\text{ref}}$ and $\Delta t = L_{\text{ref}}$:

$$\langle s_{\text{ref},z}^2 \rangle = \frac{\langle s_{0,z}^2 \rangle}{A_z} = W_z(R_{\text{ref}}) - W_z(L_{\text{ref}}) = V(R_{\text{ref}}) - V(L_{\text{ref}}) + \left(\frac{L_0}{L_{\text{ref}}} - \frac{L_0}{R_{\text{ref}}}\right) \sigma_\eta^2. \quad (\text{S49})$$

The first two terms on the right are $\langle s_{\text{ref}}^2 \rangle$ by Eq. (B12). In the last term, $L_0/R_{\text{ref}} = 1/1460$ for $L_0 = 6\text{h}$ and $R_{\text{ref}} = 365\text{d}$, which is neglected to obtain

$$510 \quad \langle s_{\text{ref},z}^2 \rangle = \langle s_{\text{ref}}^2 \rangle + \frac{L_0}{L_{\text{ref}}} \sigma_\eta^2, \quad (\text{S50})$$

which shows that the reported reference variance is the noise-free reference variance plus the noise variance of one reference sample, since a reference sample averages L_{ref}/L_0 of the z_i (see Eq. (S42)). The share of it that is measurement noise is the second term of Eq. (S50) divided by the whole, and Eq. (S46) bounds it:

$$\boxed{\frac{L_0 \sigma_\eta^2 / L_{\text{ref}}}{\langle s_{\text{ref},z}^2 \rangle} \leq \frac{L_0}{L_{\text{ref}}} \frac{2W_z(2L_0)}{W_z(R_{\text{ref}}) - W_z(L_{\text{ref}})}}. \quad (\text{S51})$$

515 The curve W_z enters only as a ratio, so the bound does not depend on its scaling. The curve of Fig. 4, which is built from records with $L_0 = 6$ h, gives $W_z(2L_0) = W_z(12\text{ h}) = 0.226$, $W_z(L_{\text{ref}}) = W_z(24\text{ h}) = 0.433$, and $W_z(R_{\text{ref}}) = W_z(365\text{ d}) = 1.804$, all in units of $W_z(7\text{ d})$. Inserting into Eq. (S51) yields 0.082; i.e., at most 8.2% of the variance of a record of 6 h samples under the reference sampling is measurement noise, and at least $\sqrt{1 - 0.082} \approx 95.8\%$ of its spread is geophysical.

References

- Canzi, A., Freney, E., Grzegorzczak, P., Baray, J., Lacher, L., and Planche, C.: Unraveling ice nucleating particle concentration variability: Insights into source emissions origin and parameterizations, *Journal of Geophysical Research: Atmospheres*, 130, e2024JD041258, <https://doi.org/10.1029/2024jd041258>, 2025.
- Creamean, J. M., Hume, C. C., Vazquez, M., and Theisen, A.: Long-term measurements of ice nucleating particles at Atmospheric Radiation Measurement (ARM) sites worldwide, *Earth System Science Data*, 17, 6943–6963, <https://doi.org/10.5194/essd-17-6943-2025>, 2025.
- Gjelsvik, A. B., David, R. O., Carlsen, T., Hellmuth, F., Hofer, S., McGraw, Z., Sodemann, H., and Storelvmo, T.: Using a region-specific ice-nucleating particle parameterization improves the representation of Arctic clouds in a global climate model, *Atmospheric Chemistry and Physics*, 25, 1617–1637, <https://doi.org/10.5194/acp-25-1617-2025>, 2025.
- Herbert, R. J., Lacher, L., Böhmländer, A., Tarn, M. D., Canzi, A., Pantoya, A., Freney, E., Höhler, K., Bogert, P., Planche, C., et al.: GloPINE dataset: model-ready measurements of INP concentrations using PINE instruments, *Earth System Science Data*, 18, 6465–6483, <https://doi.org/10.5194/essd-18-6465-2026>, 2026.
- Lacher, L., Adams, M. P., Barry, K., Bertozzi, B., Bingemer, H., Boffo, C., Bras, Y., Büttner, N., Castarede, D., Cziczo, D. J., et al.: The Puy de Dôme ICe Nucleation Intercomparison Campaign (PICNIC): comparison between online and offline methods in ambient air, *Atmospheric Chemistry and Physics*, 24, 2651–2678, <https://doi.org/10.5194/acp-24-2651-2024>, 2024.
- Möhler, O., Adams, M., Lacher, L., Vogel, F., Nadolny, J., Ullrich, R., Boffo, C., Pfeuffer, T., Hobl, A., Weiß, M., et al.: The Portable Ice Nucleation Experiment (PINE): a new online instrument for laboratory studies and automated long-term field observations of ice-nucleating particles, *Atmospheric Measurement Techniques*, 14, 1143–1166, <https://doi.org/10.5194/amt-14-1143-2021>, 2021.
- Wex, H., Huang, L., Zhang, W., Hung, H., Traversi, R., Becagli, S., Sheesley, R. J., Moffett, C. E., Barrett, T. E., Bossi, R., et al.: Annual variability of ice-nucleating particle concentrations at different Arctic locations, *Atmospheric Chemistry and Physics*, 19, 5293–5311, <https://doi.org/10.5194/acp-19-5293-2019>, 2019.
- Wilks, D. S.: *Statistical methods in the atmospheric sciences*, vol. 627, Academic press London, 2006.
- Zhou, R., Perkins, R., Juergensen, D., Barry, K., Ayars, K., Dutton, O., DeMott, P., and Kreidenweis, S.: Seasonal variability, sources, and parameterization of ice-nucleating particles in the Rocky Mountain region: importance of soil dust and biological contributions, *Atmospheric Chemistry and Physics*, 26, 1515–1535, <https://doi.org/10.5194/acp-26-1515-2026>, 2026.