Nano Ranking Analysis: determining NPF event occurrence and intensity based on the concentration spectrum of formed (sub-5 nm) particles

Diego Aliaga[1], Santeri Tuovinen[1], Tinghan Zhang[1], Janne Lampilahti[1], Xinyang Li[1], Lauri Ahonen[1], Tom Kokkonen[1], Tuomo Nieminen[1,2], Simo Hakala[1], Pauli Paasonen[1], Federico Bianchi[1], Doug Worsnop[1,3], Veli-Matti Kerminen[1], Markku Kulmala[1,4,5]

[1]Institute for Atmospheric and Earth System Research (INAR) / Physics, Faculty of Science, University of Helsinki, Finland
[2]Department of Physics, Faculty of Science, University of Helsinki, Finland
[3]Aerodyne Research Inc., Billerica, Massachusetts, United States
[4]Aerosol and Haze Laboratory, Beijing Advanced Innovation Center for Soft Matter Sciences and Engineering, Beijing University of Chemical Technology (BUCT), Beijing, China
[5]Joint International Research Laboratory of Atmospheric and Earth System Sciences, School of Atmospheric Sciences, Nanjing University, Nanjing, China

Correspondence to: Markku Kulmala (markku.kulmala@helsinki.fi)

Abstract. Here we introduce a new method, termed "Nano Ranking Analysis," for characterizing new particle formation (NPF) from atmospheric observations. Using daily variations of the particle number concentration at sizes immediately above the continuous mode of molecular clusters, here in practice 2.5-5 nm - $\Delta N_{2.5-5}$, we can determine the occurrence and estimate the strength of atmospheric NPF events. After determining the value of $\Delta N_{2.5-5}$ for all the days during a period under consideration, the next step of the analysis is to rank the days based on this simple metric. The analysis is completed by grouping the days either into a number of percentile intervals based on their ranking or into a few modes in the distribution of $log(\Delta N_{2.5-5})$ values. Using five years (2018-2022) of data from the SMEAR II station in Hyytiälä, Finland, we found that the days with higher (lower) ranking values had, on average, both higher (lower) probability of NPF events and higher (lower) particle formation rates. The new method provides probabilistic information about the occurrence and intensity of NPF events and is expected to serve as a valuable tool to define the origin of newly formed particles at many types of environments that are affected by multiple sources of aerosol precursors.

1 Introduction

Atmospheric new particle formation (NPF) events take place relatively frequently in most of the continental environments (e.g Wang et al., 2017; Kerminen et al., 2018; Nieminen et al., 2018; Chu et al., 2019; Bousiotis et al., 2021), and this phenomenon appears to be connected with both regional cloud condensation nuclei production (e.g. Peng et al., 2014; Petäjä et al., 2022) and urban haze formation (Guo et al., 2014; Kulmala et al., 2021, 2022b). The various influences of atmospheric NPF events depend essentially on their frequency of occurrence and intensity, the latter determined by time-averaged particle formation and growth rates. Quantifying the main characteristics of NPF

in different environments, and connecting these characteristics to emissions and atmospheric processes, is therefore vital for our better understanding on the potential influences of atmospheric NPF on air quality, climate and weather. Field measurements have been, and will likely continue to be, the primary source of data on the frequency and intensity

40 of atmospheric NPF (Kerminen et al., 2018; and references therein). Traditionally, the most common way to estimate the NPF event frequency from atmospheric observations is to classify individual measurement days into a small number of categories, from which one then calculates the fraction of days during which NPF events occurs (Dal Maso et al., 2005; Kulmala et al., 2012; Dada et al., 2018). Such NPF event classification methods, while widely applied in the scientific literature, tend to be subjective and time consuming, and often result in a large fraction of days for which

45 it is difficult to estimate whether a NPF event took place or not. The subjectivity and time consumptions issues can be alleviated using automatic or semi-automatic methods applied to field measurement data (e.g. Joutsensaari et al., 2018; Zaidan et al., 2018; Su et al., 2022), but the problem of having a larger fraction of days difficult to classify tends to remain.

An alternative way to approach the NPF event frequency is to define some indicator, based on quantities obtained

50 from field measurements, that predicts NPF events in a more probabilistic way (Hyvönen et al., 2005; McMurry et al., 2005; Kuang et al., 2010; Jayaratne et al., 2015; Cai et al., 2021; Olin et al., 2022). The benefits of such approaches compared to the traditional NPF event classification methods are that they are usually faster and more easily applicable to all measurement days. However, the indicators developed so far tend to be sensitive to the dominant NPF pathway and possibly other site-specific factors, casting doubts about their general applicability to different atmospheric

55 environments.

In the vast majority of studies, the intensity of NPF was estimated only for days showing clear signs of both particle formation and subsequent particle growth. This is an undesirable feature, since atmospheric NPF appears to proceed on other types of days as well, albeit typically with weaker intensities (Kulmala et al., 2022a). Currently available tools have a hard time in quantifying these weak-intensity, yet often non-negligible, periods of NPF because of

60 instrumental limitations and heterogeneities in measured air masses, limiting determination of both particle formation and growth rates.

In this manuscript, we present a novel approach, the Nano Ranking Analysis, for characterizing NPF from atmospheric observations. In the following sections, we begin by introducing and detailing the Nano Ranking Analysis. Subsequently, we offer an overview of the measurement site and the instruments employed in our study. Finally, we

65 outline the procedure for calculating the formation rate, as well as the utilization of the traditional classification method. Both components will then be used to demonstrate the effectiveness of our novel approach.

2 Methods

2.1 Description of the Nano Ranking Analysis

The Nano Ranking Analysis is designed to characterize NPF events in an objective, quantifiable and replicable

70 manner. Our foundational supposition, which is in alignment with earlier observations of atmospheric ions (e.g.

Tammet et al., 2014; Leino et al., 2016), asserts that the daily fluctuation of the particle number concentrations within the 2.5–5 nm diameter range ($\Delta N_{2.5-5}$) is acutely sensitive to the presence of atmospheric NPF.

Drawing from this premise, our method quantifies NPF events using their corresponding $\Delta N_{2.5-5}$ value. This value practically represents the daily difference between the maximum and minimum concentrations of these particles, and conveniently serves as a unique and continuous metric. The approach yields a single representative value for each measurement day.

Subsequently, we employ a two-fold approach: firstly, the derived $\Delta N_{2.5-5}$ values are used to rank NPF events, and secondly, we scrutinize the logarithmic distribution of these values to discern any dominant modes. These modes can be further modelled using Gaussian curves, thereby serving as a useful tool to differentiate between varying intensity levels, or "modes", of NPF.

2.1.1 Steps to calculate $\Delta N_{2.5-5}$

The following steps outline our approach for analyzing the days based on the $\Delta N_{2.5-5}$ spectrum:

1. *Extract the timeseries of the particle number concentration ($N_{d_1-d_2}$) in the size interval immediately above the continuous mode of molecular clusters from the particle number size distribution (Fig. 1A-B). In our case we use the particles in the size range of 2.5–5 nm.*

2. *Smooth the $N_{2.5-5}$ timeseries* to mitigate the influence of potential spurious signals on the ranking value. In our case, we applied a rolling median over 2-hour intervals. This approach reduces the impact of noise or outliers, ensuring a more reliable and accurate ranking assessment.

3. *Identify diurnal background and active regions.* Background regions are generally characterized by times of day that have minimal diurnal values, whereas active regions exhibit maximal diurnal values of $N_{2.5-5}$. To identify these regions more accurately, we recommend dividing the dataset into seasons and examine the diurnal behavior in each season separately (Fig. 1C). This is particularly important in environments with high levels of particle emissions (particularly nanoparticle emissions), such as urban environments, in which poorly chosen time regions could result in values of $\Delta N_{2.5-5}$ considerably affected by these emissions.

4. *Find the background number concentration for each event ($N_{B;2.5-5}$).* The background concentration corresponding to an event is determined based on the median value of $N_{2.5-5}$ in the so-called background region. For Hyytiälä, this time window is between 21:00 and 06:00 (Fig 1.D). This median is obtained after the 2-hour rolling smoothing of the timeseries (step 2).

5. *Find the active peak daytime number concentration ($N_{A;2.5-5}$) for each event.* The active peak concentration corresponding to an event is determined based on the max value of $N_{2.5-5}$ in the so-called active region. Note that the $N_{2.5-5}$ timeseries has been previously smoothed to a 2-hour rolling median (step 2), and this will impact the maximum value. For Hyytiälä, the active time window is between 06:00 and 18:00 (Fig 1.D)

6. *Determine the change in number concentration ($\Delta N_{2.5-5}$) for each event.* This value is defined as

$$\Delta N_{2.5-5} = N_{A;2.5-5} - N_{B;2.5-5} \qquad (1)$$

and it is the metric used to characterize the strength of potential NPF event occurrence for the corresponding day (Fig. 1D).

7. *Rank the days in percentiles* based on their corresponding value of $\Delta N_{2.5-5}$ and group the days based on 5% intervals (Fig. 2) to assess the corresponding potential NPF pattern of each interval.

2.1.2 NPF "mode" fitting

110   The $\log(\Delta N_{2.5-5})$ distribution can be used to identify NPF modes based on their intensity. The procedure is as follows: first, the $\log(\Delta N_{2.5-5})$ distribution is depicted (Fig. 3A), and a visual assessment is made to determine the number of Gaussian curves needed to describe the distribution— in our case, three curves. Next, the distribution is fitted using three Gaussian functions (g1, g2, g3). Initial guesses for the Gaussian's center, width and amplitude and their allowed range in the fitting algorithm are provided based on visual inspection. The division between these groups is determined

115   by finding the midpoint between the centers of two subsequent Gaussians (Fig. 3, dashed line). The intensity of NPF events is assessed within each group by plotting the diurnal median particle number size distribution (Fig. 4, third row), so that both visual (Fig. 4, first row) and statistical (Fig. 4, second row) inspections of the diurnal variation of $N_{2.5-5}$ can be performed for each event.

2.2 Description of the dataset

120   **2.2.1 Site description**

All the measurements were conducted at the Station for Measuring Ecosystem–Atmosphere Relations (SMEAR) II, in Hyytiälä, southern Finland (61°51′N, 24°17′E; 181 m A.S.L.; Hari and Kulmala, 2005; Hari et al., 2013). The SMEAR II station is located in a boreal pine forest, in a pristine rural environment. The nearest large city is Tampere, located approximately 60 km southwest of the station, with a population of around 200 000 residents. Additionally,

125   comprehensive observations of trace gases, soil-atmosphere fluxes, as well as meteorological variables, have been concurrently conducted at the site. More details about the station can be found from Hari and Kulmala (2005).

2.2.2 Neutral cluster and Air Ion Spectrometer (NAIS)

We used data from the NAIS (Neutral cluster and Air Ion Spectrometer, Airel Ltd.; Mirme and Mirme, 2013). The NAIS measures the number size distributions of ions and total particles in the electrical mobility diameter ranges 0.8-

130   42 nm and 2.5-42 nm respectively.

The NAIS has two measurement columns operating in parallel, one for each polarity. During ion measurements, the positive and negative ions are simultaneously measured in the two columns. During particle measurements, aerosol particles are charged to opposite polarities using corona chargers and simultaneously measured in both columns. The particle data below about 2 nm is contaminated by charger ions and is not included in the measured size range for

135   particles. Air ions and charged particles are separated based on their electrical mobilities, and detected in a multichannel differential mobility analyzer (DMA).

Particle number size distribution data from the negative polarity was used and the number concentrations of total particles between 2.5 and 5 nm were determined based on interpolation. The number concentrations of 2.5–5 nm particles were used in the Nano Ranking Analysis presented in this study. A visual screening process was performed

140   to identify and remove faulty or rain or snow contaminated measurements from the dataset (11 days).

2.2.3 Differential Mobility Particle Sizer

The Differential Mobility Particle Sizer (DMPS) measures both the neutral and charged particle number size distributions (PNSD) (Aalto et al., 2001). A typical DMPS setup contains a bipolar charger to assess the equilibrium charge distribution, a DMA (Differential Mobility Analyzer) to sample particles at specific particle sizes, and a CPC

145 (Condensation Particle Counter) to measure particle number concentrations. At the Hyytiälä site, the DMPS measurement system contains two setups, consisting of a cylindrical DMA and a CPC for each setup. The CPCs can detect particle number concentrations ranging from 3 nm to 10 nm, whereas the detection of particle size range for the DMPS is 3–1000 nm. The DMPS system contains a dryer at the inlet, with a continuous sheath flow with a relative humidity below 30% to enable measuring the particle size under relatively dry conditions. The measurement height

150 of DMPS is at ground level at the station (~2 m), and time resolution of the measurements is 10 minutes. The DMPS outputs were inverted using kernel inversion method. The inverted outputs were used for visualization of the Nano Ranking Analysis percentile bin separation (Fig. 2), as well as for the traditional NPF event classification and calculations of particle formation rates.

2.2.4 Traditional event classification

155 We compared the ranks from the Nano Ranking Analysis with the traditional NPF event classification (see section 3, Fig 7.) introduced by Dal Maso et al. (2005). The traditional event classification categorizes days into NPF event days, undefined days and non-event days by visually analyzing the particle number size distribution data from a DMPS on a day-to-day basis. This classification procedure characterized NPF events by the growth of a new mode of particles in sub-25 nm over a time span of hours. Additionally, the classified NPF events were divided into three sub-classes

160 (event Ia, Ib and II) based on the level of confidence determined by particle growth and formation rates. A detailed description of the traditional NPF event classification can be found in Dal Maso et al. (2005).

2.2.5 Particle formation rate (*J*) calculations

The particle formation rates (*J*) at 3 nm were calculated following the scheme described in Kulmala et al. (2012), and the formula is shown below:

165 $$J_{Dp} = \frac{dN_{Dp}}{dt} + CoagS_{Dp} \cdot N_{Dp} + \frac{GR_{Dp}}{\Delta Dp} \cdot N_{Dp} \qquad (2)$$

where dN_{Dp}/dt stands for observed time derivative of particle number concentration in the size range of 3–7 nm, calculated from particle number size distributions (PNSD) from DMPS measurements. The second term describes the particle losses due to coagulation to larger size particles with corresponds to their number concentrations (N_{Dp}). The calculations of coagulation sink considered the sum of coagulation sinks of particles from each size bin (Eq. 3).

170 $$CoagS_{Dp} = \sum_{Dp_1}^{Dp_2} K(Dp_1, Dp_2)N_{Dp_2} \qquad (3)$$

where $Dp_1(Dp)$ is 3 nm, Dp_2 is 7 nm, and K is the coagulation coefficient between particles of size Dp_1 and Dp_2. The correction factor for particle hygroscopic growth is applied when calculating coagulation sinks.

The last term in Eq. (2) accounts for particle losses due to growth into larger sizes. We used the maximum concentration method described in Kulmala et al. (2012) to calculate particle growth rates (GR$_{3-7}$) from days that were

175   classified as having NPF events (event Ia, Ib and II) based on the guideline illustrated in Dal Maso et al. (2005) briefly discussed in the previous section. According to the recent discovery of ongoing, non-negligible atmospheric NPF on non-event days, termed quiet NPF (Kulmala et al., 2022a), GRs calculated on normalized non-event days were found to have similar values as on traditional NPF event days. This finding allows us to assume that the median value from all GR_{3-7} is applicable when calculating J_3 from non-event days.

3 Results and Discussion

180

In this section, we present the findings of our research derived from the Nano Ranking Analysis. We focus on the comparison between the $\Delta N_{2.5-5}$ metric—either directly or via its ranks—and traditional parameters in the study of NPF. Specifically, we examine the comparison of $\Delta N_{2.5-5}$ and diurnal patterns (Fig. 5), new particle formation rates (Fig. 6), traditional classification of events (Fig. 7), and seasonality (Fig. 8). Finally, we apply the normalization

185   proposed by Kulmala et al. (2022) to reveal quiet NPF patterns in the percentile interval bins (Fig. 9). These results provide valuable insights into the effectiveness and applicability of our proposed metric in capturing key aspects of NPF dynamics.

Figure 5 shows the median particle number concentrations $N_{2.5-5}$ for different hours of the day and for different intensity rank values. We can see that the concentration profile throughout the day is relatively similar for rank values

190   approximately below 50 % with little variation in $N_{2.5-5}$. However, for ranks above 50 % slight increases in $N_{2.5-5}$ during daytime hours are observed. As the rank increases, values of $N_{2.5-5}$ are correspondingly higher. In addition, increased daytime particle concentrations last longer for days with higher intensity ranks. For example, for rank values of 85–90 % the increased concentrations are mainly present between 12:00 and 18:00, while for rank values of 95–100 % they last two to three hours longer, from approximately 11:00 until 19:00–20:00. Fig. 5 shows how the analysis method

195   presented can be used to study the particle concentrations as well as hours of the day during which NPF potentially takes place.

Figures 6A and 6B plot the particle formation rate, J_3, as a function of $\Delta N_{2.5-5}$ and of the percentile ranking, respectively. The value of J_3 clearly increases with increasing $\Delta N_{2.5-5}$ and, as a result, with an increasing percentile ranking. J_3 is the highest in group g3 and the lowest in group g1. In group g1, corresponding to percentile rankings

200   smaller than about 70 %, average values of J_3 slowly increase with increased rank. In group g2, and especially in group g3, the increase of J_3 with increasing rank is stronger. The majority of cases with J_3 larger than 0.1 cm⁻³s⁻¹ occur when the percentile ranking is larger than 80 %, with these values mostly belonging to groups g2 and g3. Therefore, Fig. 6 illustrates the clear connection of ranking with intensity of atmospheric NPF. If the percentile rank of the day is high, the intensity of NPF can also be expected to be relatively high. In addition, Fig. 6 demonstrates that the novel Nano

205   Ranking Analysis presented here can be compared with continuous variables, such as the particle formation rate, in investigating NPF. Previously, studies of continuous variables and their effect on NPF have been constrained by a binary division of days into NPF event and non-event days. It should be noted that in addition to NPF, the value of $\Delta N_{2.5-5}$, and thus, the intensity rank, can be affected by factors such as coagulation sink, other sources of sub-5 nm particles, and ion or particle production associated with e.g., rain. In our case, the days included in our analysis were

210   visually screened for precipitation to account for this. As a result, while the presented method characterizes NPF

intensity statistically, some individual days might have lower NPF intensity than their corresponding rank would suggest.

Figure 7 shows the distribution of different NPF event classes, as per the traditional classification scheme, for the different percentile rankings. The number of days classified as non-events is the highest when the rank is close to zero, and most of the days with ranks below 50 % are classified as non-event days. The number of non-event days decreases with an increasing rank value, such that only a couple of days above the percentile rank of 90 % are classified as non-event days, representing a marginal fraction of all the days. The fraction of days classified as NPF event days starts to clearly grow after the percentile ranking is over 65 %. However, in Hyytiälä most days are classified as NPF event days only for percentile rankings larger than 85 %. In conclusion, while Fig. 6 illustrates the connection of the proposed ranking method with the intensity of particle formation, Fig. 7 illustrates its connection with the probability of the occurrence of atmospheric NPF events.

Figure 8 demonstrates the advantage of having a continuous variable that characterizes the formation of new particles. In the figure, each red point corresponds to a day, and the blue curve is the product of applying a 30-day rolling median. In this way, the seasonality of the NPF events can be clearly identified, which in this case shows a higher intensity and/or frequency in the spring. This analysis would be significantly more complex if a categorical classification of NPF events were used.

Lastly, as illustrated in Figure 9, the integration of the normalization method proposed by Kulmala et al. (2022a) and the percentile bin intervals (as shown in Fig. 2), significantly enhances the visibility of NPF event patterns, even in ranges that precede those depicted in Figure 2. A well-defined NPF event pattern becomes observable starting from the rank of about 55 %, with a less obvious pattern discernible starting from the ranks of about 25 %. Moreover, the observation that each successively higher interval reveals an increasingly distinct NPF event pattern strongly suggests that the Nano Ranking Analysis effectively orders these events, even when they initially seem indistinct.

Now we address potential limitations of the proposed NPF metric ($\Delta N_{2.5-5}$). It is important to acknowledge that the high values in this metric could be influenced by factors such as pollution, or by ions or particles produced by either rain (Wimmer et al., 2018) or blowing snow (Chen et al., 2017). On the other hand, a similar formation rate of new particles will result in a lower concentration of 2.5-5 nm particles when the coagulation sink is higher, as a larger fraction of the formed particles are scavenged by the pre-existing particles (e.g. Kerminen and Kulmala, 2002; Lehtinen et al., 2007). Consequently, an event with a high NPF ranking may not necessarily correspond to an atmospheric NPF event (in case of rain or other particle sources) and, likewise, both intensity and rank of any NPF event may be influenced by varying background aerosol conditions. However, no NPF event classification scheme or metric method is entirely exempt from errors, and the observation of NPF events is always dependent on measurement instruments (their size range and noise level) and whether the formation rate is high enough to produce an observable increase in particle concentrations over the influence of coagulation sink and other sub-5 nm particle sources. Depending on the dataset, one may want to filter out such influencing factors before applying the Nano Ranking Analysis.

The application of our new method to different scenarios requires understanding of the specific dynamics unique to each site. Variations in the time windows of active and background regions are expected, and these might even vary

within the same site depending, e.g., on the time of the year. Additionally, the metric utilized in this study (Eq. 1) may not be optimally suited for other sites. For example, in environments with high pollution, the background concentration

250 of $N_{2.5-5}$ may consistently remain elevated. In such cases, it may be more appropriate to calculate the ratio between the active and background time windows rather than simply subtracting them.

Finally, the mode derived from the distribution of $\log(\Delta N_{2.5-5})$ may indicate distinct sources, including regional, meteorological, or emissions-related factors. It is crucial to contrast these findings with available precursor gases, transport patterns, and meteorological variables. Further studies applying this methodology in various scenarios will

255 provide valuable insights into, and guidance for, effectively implementing this new metric.

4 Conclusions

Formation of fresh atmospheric aerosol particles is a worldwide phenomenon. Here we present a new method to analyze NPF events. Instead of traditional binary NPF event day vs non-event day analysis, we have developed the Nano Ranking Analysis, which ranks the days based on the concentration of 2.5–5 nm particles seamlessly from very

260 low values to high ones. At the same time, the frequency, or mode, of high-ranking values – earlier referred as clear NPF event days – and low-ranking values (clear non-event days) can be obtained.

The new Nano Ranking Analysis is an automatic and objective way to find out the two important characteristic of new particle formation, namely its intensity and frequency. This continues and deepens the recent finding by Kulmala et al. (2022a) that – in practice – new aerosol particles are formed in the atmosphere during all the days. The days with

265 high (low) ranking values show typically higher (lower) particle formation rates.

Our finding enables new ways to investigate connections between different days. This includes studying how factors such as vapor concentrations, precursor gases, condensation sink, and meteorology impact the intensity and frequency of NPF events. In the future it will be important to investigate connections of the $\Delta N_{2.5-5}$ metric—or its ranks—with atmospheric conditions in different environments. We hypothesize that the new method is applicable to many other

270 types of environments and, besides providing probabilistic information about the occurrence and intensity of NPF, it has potential to provide valuable insight into the origin of newly formed particles at sites affected by multiple sources of aerosol precursors.

Code availability

https://github.com/daliagachc/ranking-hy

Author contribution

275 MK conceptualized the original idea. DA developed the method with help from MK, VMK and PP. DA performed the analysis with help from ST, TZ, JL, XL, and SH. JL and TN collected the data. DA prepared and edited the manuscript with contributions from all co-authors.

Competing interests

MK is member of the editorial board of the Aerosol Research journal. The authors have no other competing interests to declare.

Acknowledgements

We acknowledge the following projects: ACCC Flagship funded by the Academy of Finland grant number 337549 (UH) and 337552 (FMI); Academy professorship funded by the Academy of Finland (grant no. 302958); Academy of Finland projects no. 1325656, 311932, 334792, 316114, 325647, 325681, 347782, 337549; "Quantifying carbon sink, CarbonSink+ and their interaction with air quality"; INAR project funded by Jane and Aatos Erkko Foundation; "Gigacity" project funded by Wihuri foundation; European Research Council (ERC) project ATM-GTP Contract No. 742206; European Commission, H2020 Research Infrastructures (CHAPAs grant no. 850614; FORCeS grant no. 821205), and Horizon Europe Infrastructures (FOCI grant no. 101056783). Doctoral School in Atmospheric Sciences at the University of Helsinki (ATM-DP) is acknowledged. University of Helsinki support via ACTRIS-HY is acknowledged. Support of the technical and scientific staff in Hyytiälä are acknowledged. ChatGPT (GPT-3.5;4, OpenAI's large-scale language-generation model) has been used to improve the writing style of some excerpts in this article. DA reviewed, edited, and revised the ChatGPT generated texts to his own liking and takes ultimate responsibility for the content of this publication.

References

Aalto, P., Hämeri, K., Becker, E., Weber, R., Salm, J., Mäkelä, J. M., Hoell, C., O'dowd, C. D., Karlsson, H., Hansson, H.-C., Väkevä, M., Koponen, I. K., Buzorius, G., and Kulmala, M.: Physical characterization of aerosol particles during nucleation events, Tellus B Chem. Phys. Meteorol., 53, 344, https://doi.org/10.3402/tellusb.v53i4.17127, 2001.

Bousiotis, D., Pope, F. D., Beddows, D. C. S., Dall'Osto, M., Massling, A., Nøjgaard, J. K., Nordstrøm, C., Niemi, J. V., Portin, H., Petäjä, T., Perez, N., Alastuey, A., Querol, X., Kouvarakis, G., Mihalopoulos, N., Vratolis, S., Eleftheriadis, K., Wiedensohler, A., Weinhold, K., Merkel, M., Tuch, T., and Harrison, R. M.: A phenomenology of new particle formation (NPF) at 13 European sites, Atmospheric Chem. Phys., 21, 11905–11925, https://doi.org/10.5194/acp-21-11905-2021, 2021.

Cai, R., Yan, C., Worsnop, D. R., Bianchi, F., Kerminen, V.-M., Liu, Y., Wang, L., Zheng, J., Kulmala, M., and Jiang, J.: An indicator for sulfuric acid–amine nucleation in atmospheric environments, Aerosol Sci. Technol., 55, 1059–1069, https://doi.org/10.1080/02786826.2021.1922598, 2021.

Chen, X., Virkkula, A., Kerminen, V.-M., Manninen, H. E., Busetto, M., Lanconelli, C., Lupi, A., Vitale, V., Del Guasta, M., Grigioni, P., Väänänen, R., Duplissy, E.-M., Petäjä, T., and Kulmala, M.: Features in air ions measured by an air ion spectrometer (AIS) at Dome C, Atmospheric Chem. Phys., 17, 13783–13800, https://doi.org/10.5194/acp-17-13783-2017, 2017.

Chu, B., Kerminen, V.-M., Bianchi, F., Yan, C., Petäjä, T., and Kulmala, M.: Atmospheric new particle formation in China, Atmospheric Chem. Phys., 19, 115–138, https://doi.org/10.5194/acp-19-115-2019, 2019.

Dada, L., Chellapermal, R., Buenrostro Mazon, S., Paasonen, P., Lampilahti, J., Manninen, H. E., Junninen, H., Petäjä, T., Kerminen, V.-M., and Kulmala, M.: Refined classification and characterization of atmospheric new-particle formation events using air ions, Atmospheric Chem. Phys., 18, 17883–17893, https://doi.org/10.5194/acp-18-17883-2018, 2018.

Dal Maso, M., Kulmala, M., Riipinen, I., Wagner, R., Hussein, T., Aalto, P. P., and Lehtinen, K.: Formation and growth of fresh atmospheric aerosols: eight years of aerosol size distribution data from SMEAR II, Hyytiälä, Finland, Boreal Environ. Res., 10, 323–336, 2005.

Guo, S., Hu, M., Zamora, M. L., Peng, J., Shang, D., Zheng, J., Du, Z., Wu, Z., Shao, M., Zeng, L., Molina, M. J., and Zhang, R.: Elucidating severe urban haze formation in China, Proc. Natl. Acad. Sci., 111, 17373–17378, https://doi.org/10.1073/pnas.1419604111, 2014.

Hari, P. and Kulmala, M.: Station for Measuring Ecosystem-Atmosphere Relations (SMEAR II), Boreal Environ. Res., 10, 315–322, 2005.

Hari, P., Nikinmaa, E., Pohja, T., Siivola, E., Bäck, J., Vesala, T., and Kulmala, M.: Station for Measuring Ecosystem-Atmosphere Relations: SMEAR, in: Physical and Physiological Forest Ecology, edited by: Hari, P., Heliövaara, K., and Kulmala, L., Springer Netherlands, Dordrecht, 471–487, https://doi.org/10.1007/978-94-007-5603-8_9, 2013.

Hyvönen, S., Junninen, H., Laakso, L., Dal Maso, M., Grönholm, T., Bonn, B., Keronen, P., Aalto, P., Hiltunen, V., Pohja, T., Launiainen, S., Hari, P., Mannila, H., and Kulmala, M.: A look at aerosol formation using data mining techniques, Atmospheric Chem. Phys., 5, 3345–3356, https://doi.org/10.5194/acp-5-3345-2005, 2005.

Jayaratne, E. R., Clifford, S., and Morawska, L.: Atmospheric Visibility and PM $_{10}$ as Indicators of New Particle Formation in an Urban Environment, Environ. Sci. Technol., 49, 12751–12757, https://doi.org/10.1021/acs.est.5b01851, 2015.

335

Joutsensaari, J., Ozon, M., Nieminen, T., Mikkonen, S., Lähivaara, T., Decesari, S., Facchini, M. C., Laaksonen, A., and Lehtinen, K. E. J.: Identification of new particle formation events with deep learning, Atmospheric Chem. Phys., 18, 9597–9615, https://doi.org/10.5194/acp-18-9597-2018, 2018.

Kerminen, V.-M. and Kulmala, M.: Analytical formulae connecting the "real" and the "apparent" nucleation rate and

340

the nuclei number concentration for atmospheric nucleation events, J. Aerosol Sci., 33, 609–622, https://doi.org/10.1016/S0021-8502(01)00194-X, 2002.

Kerminen, V.-M., Chen, X., Vakkari, V., Petäjä, T., Kulmala, M., and Bianchi, F.: Atmospheric new particle formation and growth: review of field observations, Environ. Res. Lett., 13, 103003, https://doi.org/10.1088/1748-9326/aadf3c, 2018.

345

Kuang, C., Riipinen, I., Sihto, S.-L., Kulmala, M., McCormick, A. V., and McMurry, P. H.: An improved criterion for new particle formation in diverse atmospheric environments, Atmospheric Chem. Phys., 10, 8469–8480, https://doi.org/10.5194/acp-10-8469-2010, 2010.

Kulmala, M., Petäjä, T., Nieminen, T., Sipilä, M., Manninen, H. E., Lehtipalo, K., Dal Maso, M., Aalto, P. P., Junninen, H., Paasonen, P., Riipinen, I., Lehtinen, K. E. J., Laaksonen, A., and Kerminen, V.-M.: Measurement of the

350

nucleation of atmospheric aerosol particles, Nat. Protoc., 7, 1651–1667, https://doi.org/10.1038/nprot.2012.091, 2012.

Kulmala, M., Dada, L., Daellenbach, K. R., Yan, C., Stolzenburg, D., Kontkanen, J., Ezhova, E., Hakala, S., Tuovinen, S., Kokkonen, T. V., Kurppa, M., Cai, R., Zhou, Y., Yin, R., Baalbaki, R., Chan, T., Chu, B., Deng, C., Fu, Y., Ge, M., He, H., Heikkinen, L., Junninen, H., Liu, Y., Lu, Y., Nie, W., Rusanen, A., Vakkari, V., Wang, Y., Yang, G., Yao, L., Zheng, J., Kujansuu, J., Kangasluoma, J., Petäjä, T., Paasonen, P., Järvi, L., Worsnop, D., Ding, A., Liu, Y., Wang,

355

L., Jiang, J., Bianchi, F., and Kerminen, V.-M.: Is reducing new particle formation a plausible solution to mitigate particulate air pollution in Beijing and other Chinese megacities?, Faraday Discuss., 226, 334–347, https://doi.org/10.1039/D0FD00078G, 2021.

Kulmala, M., Junninen, H., Dada, L., Salma, I., Weidinger, T., Thén, W., Vörösmarty, M., Komsaare, K., Stolzenburg, D., Cai, R., Yan, C., Li, X., Deng, C., Jiang, J., Petäjä, T., Nieminen, T., and Kerminen, V.-M.: Quiet New Particle

360

Formation in the Atmosphere, Front. Environ. Sci., 10, 2022a.

Kulmala, M., Cai, R., Stolzenburg, D., Zhou, Y., Dada, L., Guo, Y., Yan, C., Petäjä, T., Jiang, J., and Kerminen, V.-M.: The contribution of new particle formation and subsequent growth to haze formation, Environ. Sci. Atmospheres, 2, 352–361, https://doi.org/10.1039/D1EA00096A, 2022b.

Lehtinen, K. E. J., Dal Maso, M., Kulmala, M., and Kerminen, V.-M.: Estimating nucleation rates from apparent

365

particle formation rates and vice versa: Revised formulation of the Kerminen–Kulmala equation, J. Aerosol Sci., 38, 988–994, https://doi.org/10.1016/j.jaerosci.2007.06.009, 2007.

Leino, K., Nieminen, T., Manninen, H., Petäjä, T., Kerminen, V., and Kulmala, M.: Intermediate ions as a strong indicator of new particle formation bursts in a boreal forest, Boreal Environ. Res., 2016.

McMurry, P. H., Fink, M., Sakurai, H., Stolzenburg, M. R., Mauldin, R. L., Smith, J., Eisele, F., Moore, K., Sjostedt, S., Tanner, D., Huey, L. G., Nowak, J. B., Edgerton, E., and Voisin, D.: A criterion for new particle formation in the sulfur-rich Atlanta atmosphere, J. Geophys. Res., 110, D22S02, https://doi.org/10.1029/2005JD005901, 2005.

Mirme, S. and Mirme, A.: The mathematical principles and design of the NAIS – a spectrometer for the measurement of cluster ion and nanometer aerosol size distributions, Atmospheric Meas. Tech., 6, 1061–1071, https://doi.org/10.5194/amt-6-1061-2013, 2013.

Nieminen, T., Kerminen, V.-M., Petäjä, T., Aalto, P. P., Arshinov, M., Asmi, E., Baltensperger, U., Beddows, D. C. S., Beukes, J. P., Collins, D., Ding, A., Harrison, R. M., Henzing, B., Hooda, R., Hu, M., Hõrrak, U., Kivekäs, N., Komsaare, K., Krejci, R., Kristensson, A., Laakso, L., Laaksonen, A., Leaitch, W. R., Lihavainen, H., Mihalopoulos, N., Németh, Z., Nie, W., O'Dowd, C., Salma, I., Sellegri, K., Svenningsson, B., Swietlicki, E., Tunved, P., Ulevicius, V., Vakkari, V., Vana, M., Wiedensohler, A., Wu, Z., Virtanen, A., and Kulmala, M.: Global analysis of continental boundary layer new particle formation based on long-term measurements, Atmospheric Chem. Phys., 18, 14737–14756, https://doi.org/10.5194/acp-18-14737-2018, 2018.

Olin, M., Okuljar, M., Rissanen, M. P., Kalliokoski, J., Shen, J., Dada, L., Lampimäki, M., Wu, Y., Lohila, A., Duplissy, J., Sipilä, M., Petäjä, T., Kulmala, M., and Dal Maso, M.: Measurement report: Atmospheric new particle formation in a coastal agricultural site explained with binPMF analysis of nitrate CI-APi-TOF spectra, Atmospheric Chem. Phys., 22, 8097–8115, https://doi.org/10.5194/acp-22-8097-2022, 2022.

Peng, J. F., Hu, M., Wang, Z. B., Huang, X. F., Kumar, P., Wu, Z. J., Guo, S., Yue, D. L., Shang, D. J., Zheng, Z., and He, L. Y.: Submicron aerosols at thirteen diversified sites in China: size distribution, new particle formation and corresponding contribution to cloud condensation nuclei production, Atmospheric Chem. Phys., 14, 10249–10265, https://doi.org/10.5194/acp-14-10249-2014, 2014.

Petäjä, T., Tabakova, K., Manninen, A., Ezhova, E., O'Connor, E., Moisseev, D., Sinclair, V. A., Backman, J., Levula, J., Luoma, K., Virkkula, A., Paramonov, M., Räty, M., Äijälä, M., Heikkinen, L., Ehn, M., Sipilä, M., Yli-Juuti, T., Virtanen, A., Ritsche, M., Hickmon, N., Pulik, G., Rosenfeld, D., Worsnop, D. R., Bäck, J., Kulmala, M., and Kerminen, V.-M.: Influence of biogenic emissions from boreal forests on aerosol–cloud interactions, Nat. Geosci., 15, 42–47, https://doi.org/10.1038/s41561-021-00876-0, 2022.

Su, P., Joutsensaari, J., Dada, L., Zaidan, M. A., Nieminen, T., Li, X., Wu, Y., Decesari, S., Tarkoma, S., Petäjä, T., Kulmala, M., and Pellikka, P.: New particle formation event detection with Mask R-CNN, Atmospheric Chem. Phys., 22, 1293–1309, https://doi.org/10.5194/acp-22-1293-2022, 2022.

Tammet, H., Komsaare, K., and Hõrrak, U.: Intermediate ions in the atmosphere, Atmospheric Res., 135–136, 263–273, https://doi.org/10.1016/j.atmosres.2012.09.009, 2014.

Wang, Z., Wu, Z., Yue, D., Shang, D., Guo, S., Sun, J., Ding, A., Wang, L., Jiang, J., Guo, H., Gao, J., Cheung, H. C., Morawska, L., Keywood, M., and Hu, M.: New particle formation in China: Current knowledge and further directions, Sci. Total Environ., 577, 258–266, https://doi.org/10.1016/j.scitotenv.2016.10.177, 2017.

Wimmer, D., Buenrostro Mazon, S., Manninen, H. E., Kangasluoma, J., Franchin, A., Nieminen, T., Backman, J., Wang, J., Kuang, C., Krejci, R., Brito, J., Goncalves Morais, F., Martin, S. T., Artaxo, P., Kulmala, M., Kerminen,

405   V.-M., and Petäjä, T.: Ground-based observation of clusters and nucleation-mode particles in the Amazon, Atmospheric Chem. Phys., 18, 13245–13264, https://doi.org/10.5194/acp-18-13245-2018, 2018.

Zaidan, M. A., Haapasilta, V., Relan, R., Junninen, H., Aalto, P. P., Kulmala, M., Laurson, L., and Foster, A. S.: Predicting atmospheric particle formation days by Bayesian classification of the time series features, Tellus B Chem. Phys. Meteorol., 70, 1530031, https://doi.org/10.1080/16000889.2018.1530031, 2018.

410

Figures

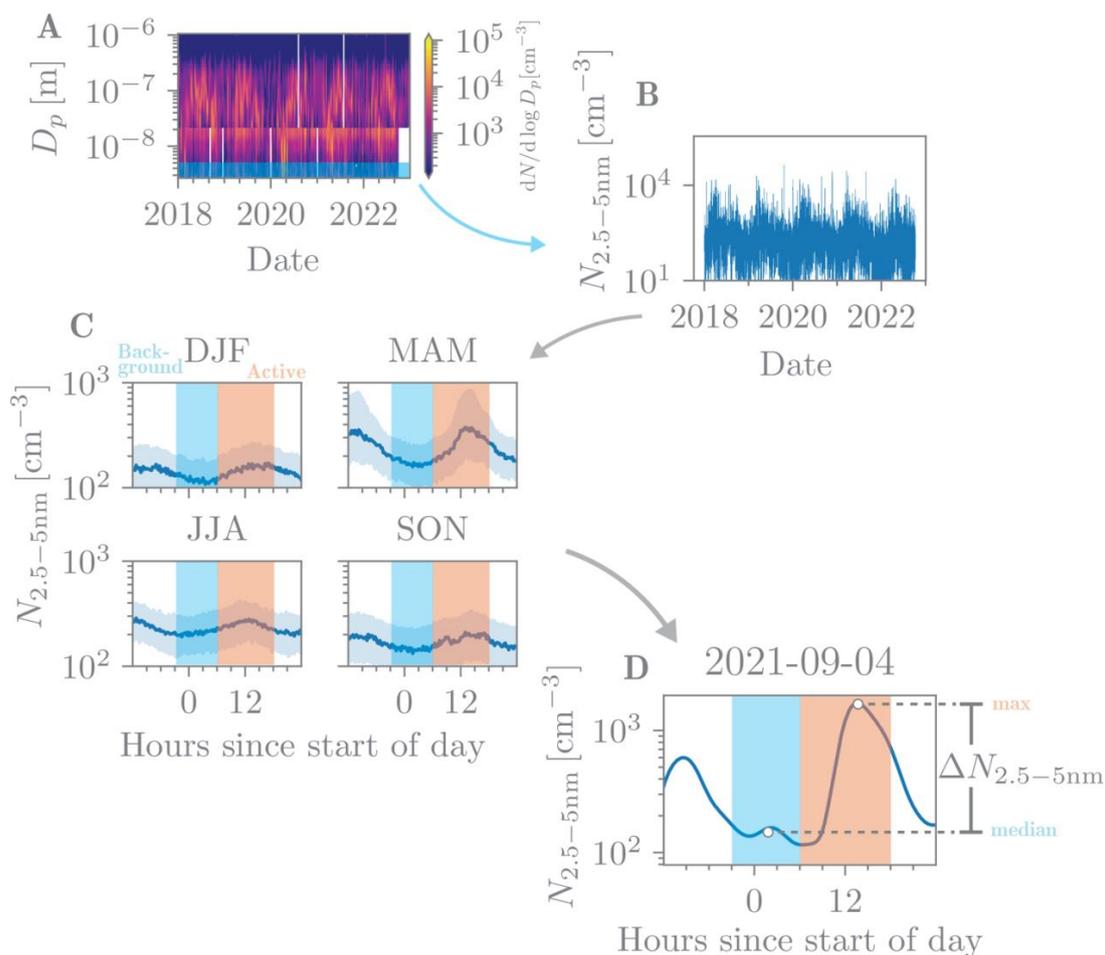



Figure 1: Diagram illustrating the process used to calculate the $\Delta N_{2.5-5}$ metric. The methodology involves extracting particle concentrations in the 2.5 to 5nm range (B) from the particle number size distribution timeseries (A), followed by grouping the timeseries by season and plotting daily patterns identifying the background and active zones (C). The final step calculates the daily difference for each day—here exemplified by the 4th of September 2021— between the maximum concentration in the active region and the median concentration in the background region (D).

415

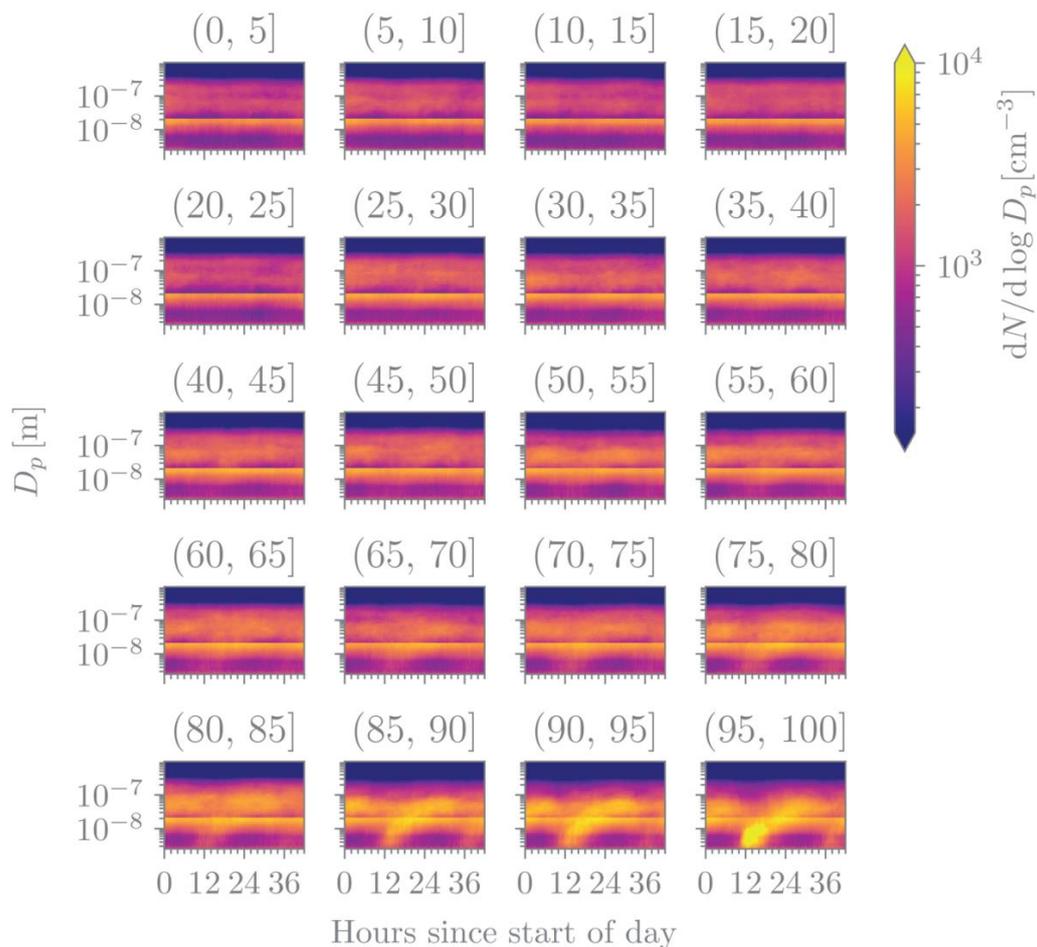
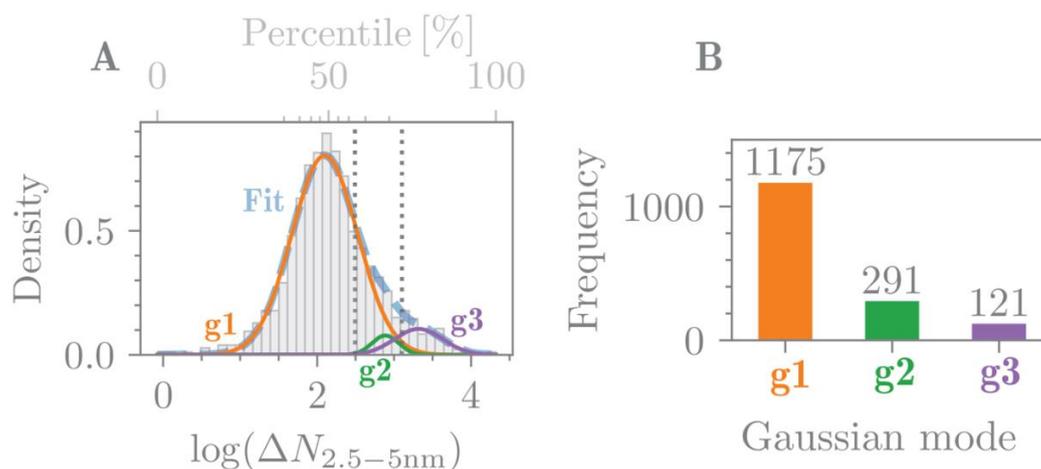
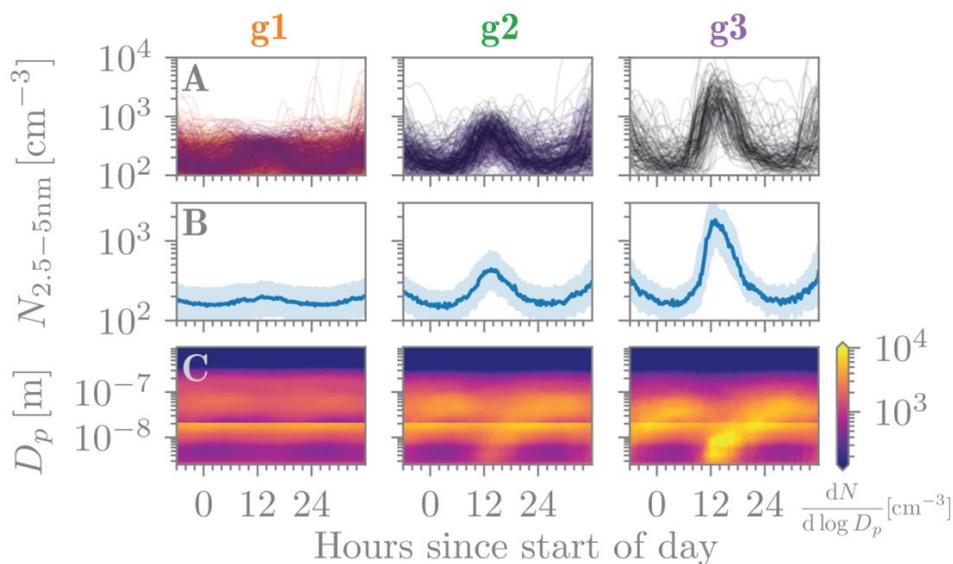



Figure 2: Daily median number particle size distribution grouped into 5% intervals based on the $\Delta N_{2.5-5}$ ranks. The diameter limits (y-axis) in the surface plots are 2.5 and 1000 nm and are used to illustrate the shape of the potential NPF events in each interval. For illustration purposes, the particle number concentrations obtained from NAIS (2.5-20nm) and DMPS (20-1000nm) are combined.

420

425 **Figure 3: (A) Density histogram (gray) based on the daily $log(\Delta N_{2.5-5})$ values. Additionally, three Gaussian curves (g1, g2, g3) are fitted (sky blue) to the distribution. (B) Number of days belonging to each of the identified Gaussian curves in (A). The regions corresponding to each group are determined by calculating the middle distance (dashed lines) between adjacent Gaussian curve centers.**

430 **Figure 4: The first row displays the daily curve of particle concentrations in the range of 2.5 to 5nm, grouped into three columns: g1, g2, and g3. The second row shows the median and interquartile range of the daily curves shown in the first row. The third row presents the median particle number distribution for the days belonging to each group g.**

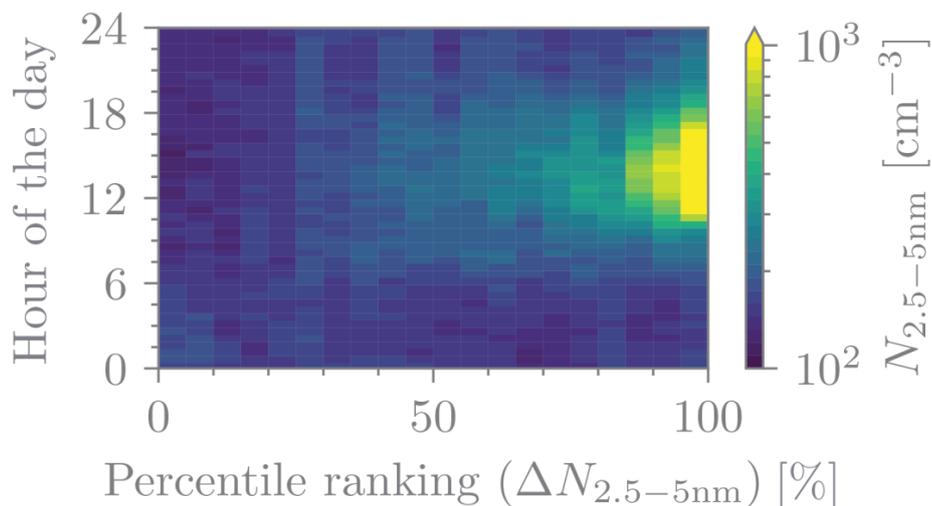



Figure 5: Median 2.5-5 nm particle number concentrations (N$_{2.5-5\ nm}$) for different times of the day and for different percentile ranking values, which were based on ΔN$_{2.5-5\ nm}$. The percentile rankings have been divided into 5% intervals, while hourly time resolution was used.

435

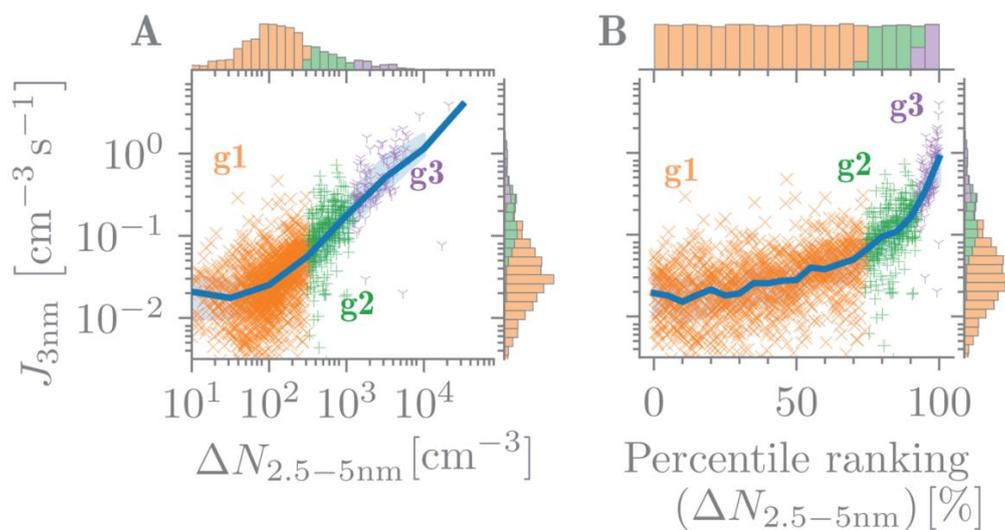



Figure 6: (A) Daily $\Delta N_{2.5-5}$ (x) vs J (y) color-coded by group g. The blue line represents the median, and the shaded region indicates the interquartile range. At the top of the panel, a histogram of $\Delta N_{2.5-5}$ values is presented, while on the right side, a histogram of J values is shown, both color-coded by group g. (B) Similar to panel A but using percentile ranking instead of $\Delta N_{2.5-5}$.

440

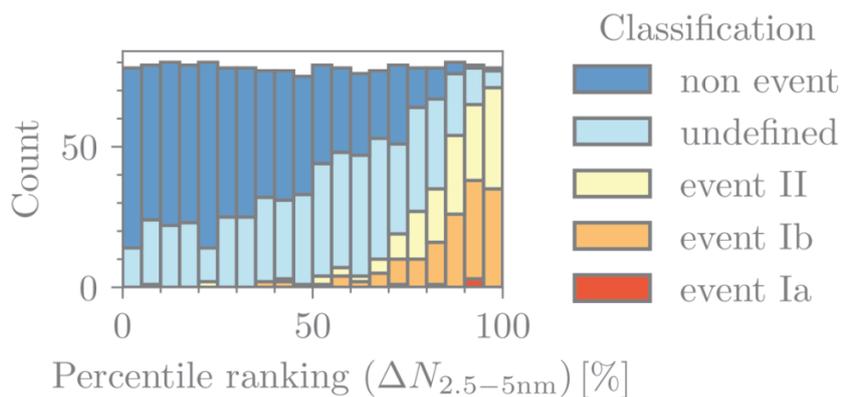



Figure 7: Comparison between percentile ranking and traditional classification. This histogram displays the percentile rankings divided into 5% bins and color-coded based on the classification: Blue (non-event), light blue (undefined), yellow (event II), orange (event Ib), and red (event Ia).

445

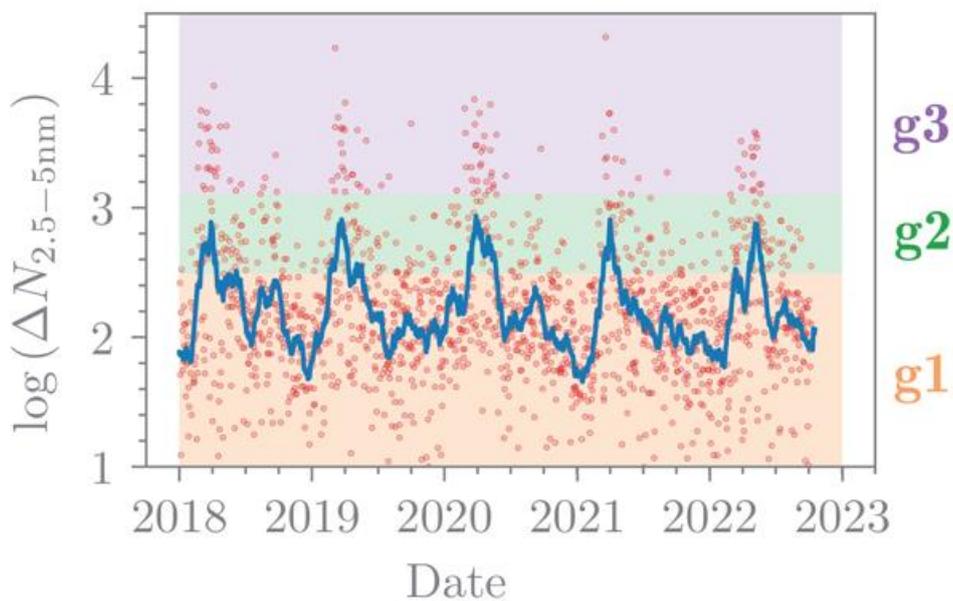



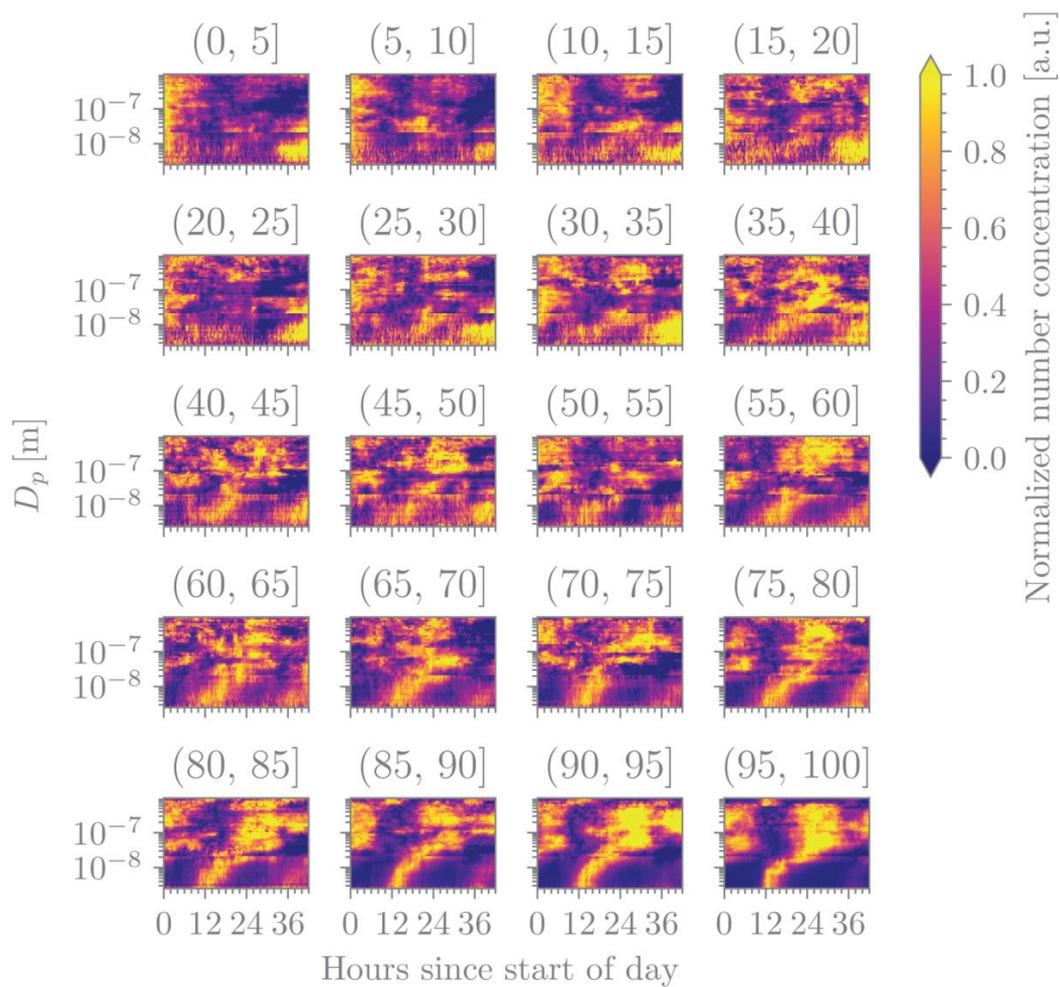
Figure 8: Daily time series of $\Delta N_{2.5-5}$. The red dots represent the daily values, while the blue line shows the 30-day rolling median. The shaded areas indicate the regions corresponding to the modes g1, g2, and g3.

450

Figure 9: Similar to Fig. 2 but using the normalization proposed by (Kulmala et al., 2022a). In short, for every 5%-interval, each size bin is linearly scaled based on its median maximum and minimum so that values span from 0 to 1.