We thank the Editor, Co-editor and reviewers for their patient and careful revision of the manuscript. Besides the changes in the manuscript to address both reviewers' comments, we slightly reorganized sub-sections and sub-sub-sections in section 2. We also added Katrianne Lehtipalo to the author list as she was essential in revising the ACPD paper. Below we include point-by-point answers to the comments as well as a marked version of the manuscript.

Reply to Referee #1

We thank Referee #1 for their helpful suggestions. We replied to the comments below. The bold text refers to the referee's comments, and the text in italics are additions to the manuscript. The line numbers mentioned in the text below refer to the ACPD version of the manuscript.

I. General comments

1. The research follows other studies considering the effect of cloud cover on NPF events (Baranizadeh et al in Boreal Env. Research) and several criteria for predicting new particle formation (papers by Kuang (ACP 10 8469), and Hyvonen, Nieminen as cited in the manuscript), some of which have also considered solar radiation levels. This research shows that cloudiness, condensation sink and temperature, when used together, can effectively predict the probability of a nucleation event at this important field station in spring, but rather less effectively in summer, autumn and winter. The authors should be more explicit in motivating their work. Yes, aerosols are important for climate and nucleation is an important source of aerosols, but it should be explained more clearly why predicting the probability of a nucleation event in springtime at Hyytiala will help people improve their understanding of the atmosphere. For example, the authors could point out that Hyytiala is reasonably representative of semiclean forested environments in the Northern Hemisphere, and that it is a suitable site without too many highly localised sources of aerosol that are difficult to model. And, since the authors present a criterion that is only effective in spring, explain that this is the most important season for NPF. And that 20 years of detailed observation data are not readily available at other sites.

We agree with the reviewer that it is important to mention the characteristics of Hyytiälä that make the location interesting and important for studying NPF. Accordingly, the following has been added to the manuscript (Line 55):

The Station for Measuring Forest Ecosystem-Atmosphere Relations (SMEAR II) located in Hyytiälä, southern Finland, compiles up to 21 years of particle number size distribution and extensive complementary data, providing the longest size distribution time series in the world, and hence allows for robust NPF analysis which is not readily possible at other sites. The station is located in a homogenous Scots pine forest far from major pollution sources. Hyytiälä, thus, is classified as a background site representative of the semi-clean northern hemisphere boreal forests.

Our focus on springtime is explained more thoroughly following the reviewer's suggestion (line 235):

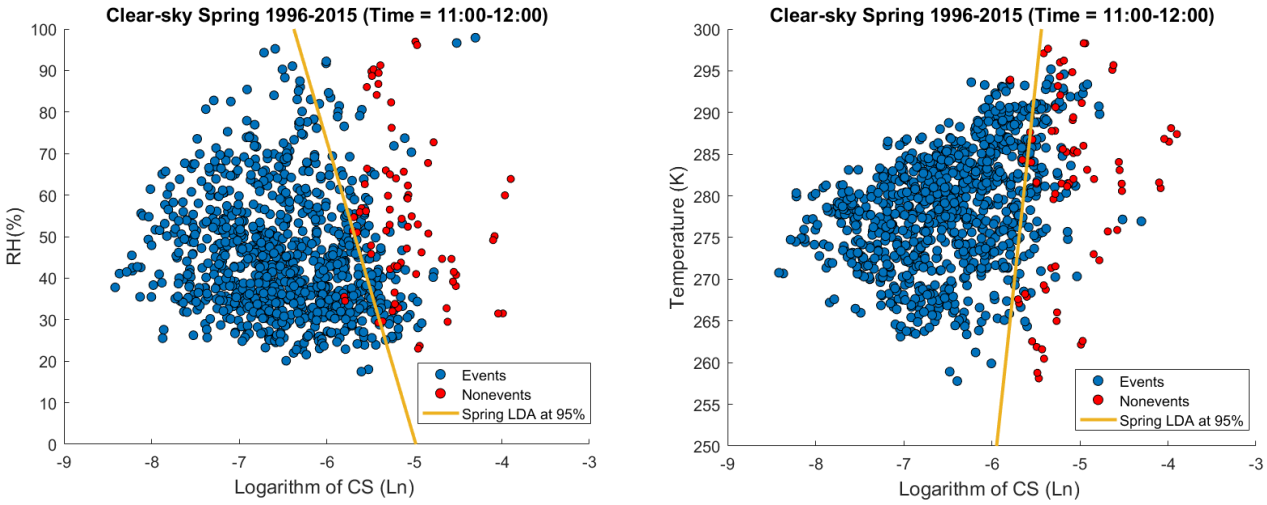
Since NPF is most frequent in spring, we dedicate our focus on this season (Figure 3a).

2. The authors investigate several other variables that should be correlated to NPF event probability, but they do not explain why CS and temperature, and not the other variables, feature in their final criterion. For example, the box-and-whisker plots (where the cloudiness parameter is than 0.7, indicating clear skies) do not clearly suggest that on clear sky days, T offers a better separation between NPF events and non-events than RH. The correlation matrix shows RH and CS are less correlated than T and CS, so RH might have more discriminating power. Moreover, RH on non-event days is almost always higher than on event days, while temperature on non-event days is higher in winter and spring but lower in autumn and winter. Therefore, in principle one might expect RH to be a better second variable than T when all seasons are considered, in line

with Hyvonen et al, even after one has separated clear sky days and cloudy days as suggested by Baranizadeh et al already. The authors should quantitatively demonstrate why their criterion offers better discriminating power than a few other obvious possibilities, such as RH/CS.

Looking at the median values presented in RH monthly box plots in Figure 5c, might give the idea that RH values are clearly separated between events and non-events. Considering the wider spread in the RH data (25%-75% percentiles) for events and nonevents as seen in the boxplots, we feel this parameter is less conclusive in separating the two classes. However, based on the reviewer's suggestion, we plot RH vs CS (spring time window 11:00-12:00) below and compare it to T vs CS (spring time window 11:00-12:00) plot. The line in the figure is the Linear Discriminant Analysis (LDA) at 95% confidence that all the nonevent points are outside the line (to the right). The plots show that RH does not result in better separation than temperature (events from nonevents) as CS sink seems to be the main controlling factor. We then conclude that during clear-sky conditions the results are somewhat different from what Hyvönen et al. (2005) who did not consider clear-sky conditions only. Based on the aforementioned results, and following the reviewer's suggestion, we add the following to the text to line 369:

Furthermore, we analyzed the effect of RH in separating the events from nonevents, similar to the study done on RH by Hyvönen et al. 2005. We found that compared with CS vs temperature data, depicting CS vs RH (data not presented) did not work better in separating NPF events from non-events during clear-sky conditions.



3. In addition to the criterion for new particle formation events, the paper also aims to quantify the effect of cloudiness on NPF event frequency. However, this was done already by Baranizadeh et al and it should be made clearer what this manuscript adds to this relatively comprehensive previous work. The authors should either remove all but a very brief summary of this from the paper, or state clearly how their analysis relates to that of Baranizadeh et al with a sentence like "Our work confirms the conclusions of Baranizadeh et al with a complementary dataset".

We added the suggested sentence to the manuscript to confirm the parallel between our study and Baranizadeh et al. 2014 to line (393): *Our work confirms, with a complementary dataset, the conclusions of Baranizadeh et al. (2014) that NPF events and non-events are typically associated with clear-sky and cloudy conditions, respectively.*

4. The paper also aims to "find out the connection between nucleating precursor vapours and new particle formation rates". The authors should re-think this part of the manuscript. The analysis has the potential to provide interesting conclusions, but currently it is not well connected to the rest of the paper and the approach of the authors does not match the stated aim. The sentence I quote here is misleading because the connection is assumed by the manuscript, not "found": the new particle formation rates presented in the paper are not calculated from the rate of change of particle concentration, but from a parameterisation of the nucleating precursor vapour proxy concentrations. This is not a bad approach but just needs to be described more carefully: the comparison of the probability of a new particle formation event to the parameterized nucleation rate is still a useful exercise.

5.

We modified the aim mentioned by the referee into the following form:

iii) Explore the connections between new particle formation rates calculated from precursor vapor proxies and the occurrence of NPF events.

6. To connect this to the rest of the paper, the authors could consider presenting this study as a comparison of the effectiveness of their condensation sink based criterion and their nucleation rate parameterisation at determining whether or not a nucleation event will occur on a given day. Then the conclusion might indicate explicitly that the parameterisation is a poor criterion for NPF compared to the condensation sink-temperature criterion (except perhaps in winter, from Figure 10?). Since parametrisations of this form might be considered reasonable starting points to determine whether or not NPF should occur (naively, a high parameterised rate ought to imply a nucleation event is likely), this would seem to be an interesting message. While Figure 10 is helpful in providing this message, further evidence could be obtained by re-plotting some of the data so that the criteria can be compared more directly: Figures 5a and 9a can be compared, but it would be better if the combined criterion including temperature were plotted on the y axis of a new version of Fig. 5a, since this should further improve the separation.

According to the suggestion of both reviewers and to clarify the message of Figure 10, we re-plot figure 10 into a diurnal cycle showing the median CS and the parameterized formation rate. The replacement aids the reader in visualizing the influence of the CS on $J_{3,C}$ as well as the negative correlation between the two. The plot will also improve and connect the parts of the paper together as the reviewer implied. Accordingly, we modify the related text in the manuscript to the following (Line 342):

On NPF event days, the median approximated formation rate of 3 nm particles had its maximum value at about midday and was significantly higher than on non-events days (Figures 9b and 10). A clear negative relation could be seen between the median seasonal diurnal cycles of CS and $J_{3,C}$ on NPF event days (specifically during spring daytime) (Figure 10). This kind of relation was not observed during non-event days when these two quantities seemed to be independent of each other (Figure 10). In summer, the median value of $J_{3,C}$ was roughly similar between NPF events and non-events, whereas the median value of CS was almost ten times higher during the non-event days compared with event days. The high values of $J_{3,C}$ for the non-event days in summer, despite the high CS values, seem to suggest that some other factor limits the actual NPF rate. One option is that the freshly formed clusters are rapidly evaporated due to higher ambient temperatures (see Fig. 5b). This will be discussed in more detail in the following section.

7. The careful statistical summaries presented in this manuscript do convince the reader that the underlying dataset is valuable. The large size allows statistically significant results to be extracted. A csv table (or similar) containing the full dataset, or at a minimum a list of dates studied over the 20 years with, on each day, the condensation sink, temperature, RH, cloudiness parameter, and whether or not the day contained a nucleation event, should be included in supplementary materials. While much of this information is already available, via the smartSMEAR website for example, a carefully compiled dataset specific to this paper would still be very useful. It would allow, for example, modellers comparing event frequency in models and observations to split up the dataset into individual years and compare model to measurements selected from the overall dataset to match their model simulations. A brief explanation of where subsets of the data have been published before should accompany the data file.

As the reviewer mentioned, the data used for our analysis can be downloaded from smartSMEAR. Because compiling 20 years of data results in a massive file, we do not think that it is necessary. However, the authors are happy to collaborate and send the needed data to modelers to improve or add valuable data to the literature.

II. Specific comments

Summary: the sentence "This study serves as basis for scientists aiming at improving their understanding towards new particle formation" should be rephrased to improve the written English, for example "This study serves as a basis for scientists aiming to improve their understanding of new particle formation."

1. Abstract: "utilizing"->"building on".

We replaced the word utilizing by building on.

2. "In this comparison we considered, for example, the effects of calculated particle formation rates, condensation sink, trace gas concentrations and various meteorological quantities." -> considered the effect on what?

We reformulated the sentence according to the reviewer's suggestion in line 20 by adding the following continuation of the sentence: *in discriminating NPF events from non-events*.

3. "The formation rate of 1.5 nm particles was calculated by using proxies for gaseous sulfuric acid and oxidized products of low volatile organic compounds"-> add "and a nucleation rate parameterization" after "compounds".

We added: "*and a nucleation rate parameterization factor*" to line 22.

4. "As expected, our results indicate an increase in the frequency of NPF events under clear-sky conditions."-> "increase under clear-sky conditions compared to cloudy conditions".

We added the suggested to the sentence "*in comparison to cloudy ones"* to line 23.

5. "The calculated formation rate of 3 nm particles showed a notable difference between the NPF event and non-event days during clear-sky conditions, especially in winter and spring"-> so in cloudy conditions do you get high NPF rates but no events? Please be more explicit here.

The objective of the paper is to select the clear-sky days and compare the events and nonevents within this dataset and we only discuss the clear-sky days in this article. As this is implicit based on the previous sentences, we removed the "during clear-sky conditions" from the sentence in order not to cause confusion such as for the

referee. The sentence quoted by the reviewer above means that within the selected clear-sky dataset, the formation rate between events and non-events is different. However, for the interest of the reviewer, we present below a table of statistical analysis of formation rates at 1.5nm between clear-sky (P>0.7) and cloudy (P<0.3) events and non-events.

$J_{1.5}$ (cm^{-3} s^{-1})	5th percentile	25th percentile	Median	75th percentile	95th percentile
Clear-sky events	0.10	0.56	1.37	2.89	8.76
Clear-sky non-events	0.03	0.28	0.77	1.82	5.72
Cloudy events	0.01	0.21	0.72	1.93	7.50
Cloudy nonevents	0.01	0.06	0.17	0.46	2.00

Results appear as expected as the $J_{1.5}$ is calculated directly from the concentrations of sulfuric acid and OxOrg (equation 5) which are both produced photochemically. Thus, it is, as shown, expected that the parametrized formation rate is higher in clear-sky conditions than cloudy ones, in general. The pattern is the same for the J_3. However, in both cases, particle formation rate is higher on event days than non-events if we take either clear-sky or cloudy conditions, separately.

6. Line 59: "That study" -> which of the three cited?

The sentence "That study" was replaced with: "They" for clarity. The Baranizadeh et al. are the only authors being actively addressed in the previous sentence.

7. Line 88: The title of section 2.1.1 should be amended to make it clear that it is this section which explains how events are categorised, and this section should be extended with a very brief summary of how Dal Maso et al decide whether a day is an event, non-event or undefined day.

The title of section 2.1.1 is modified to: *"New Particle Formation Events Classification"*.

The following is added to this section: *"The latter uses a decision criterion based on the presence of particles < 25 nm in diameter and their consequent growth to Aitken mode. Event days are days on which sub 25 nm particle formation and growth are observed. Non-event days are days on which neither modes are present. Undefined days are the days which do not fit either criterion."*

8. Line 127: At least four possible MT proxies are presented in Kontkanen et al. Which one did you use? Is it the recommended proxy MTproxy1,doy? Please specify.

We added the missing information to the manuscript.

9. Line 140-147: this section needs more detail on the data analysed and the characteristics of NPF in Hyytiala. Table 1 caption implies all of the data analysed are from the months of March to May, but this seems not to be true. However, it is clear that the instruments would not be running every single day between 1 January 1996 and 31 December 2015. While Figure 2 is helpful here, it also needs some additional explanation and referencing early in the text. The brief statements about the seasonal cycle at lines 227-232 are confusing without this additional context.

While the analysis included in the manuscript is comprehensive and includes data between 1996 and 2015, some of the figures/tables include only part of the data. For instance, as the reviewer pointed out that Table 1 relates only to the spring time correlation. The table includes only spring time as there are the months with the highest

frequency of events. Including all months in one correlation calculation would not give reasonable results. The number of classified clear-sky data points are included in Figure 2.

10. Line 160: Somewhere here it would be good to state why you calculate J3, why not just use J1.5?

In previous studies which did not consider clear-sky conditions, it has been observed that there is a clear difference in J_3 between event and non-event days while $J_{1.5}$ is more similar (Kulmala et al. 2013). Furthermore, we have more direct J_3 measurements to which we can compare the calculated values. Classification of events around the world are based on DMPS data at 3 nm and above, rather than on data below 3 nm diameter.

Line 320 now reads*: "Since previous studies have shown that there is a clear difference in J_3 between the event and non-event days, and much less difference in $J_{1.5}$ (Kulmala et al. 2013), we decided to focus on J_3 in our event to non-event discrimination."*

11. Line 174: Please summarise very briefly the improvement made by Kontkanen (2016).

The main improvements in OxOrg proxy by Kontkanen et al. (2016) compared to the previous version (Lappalainen et al., 2009) of the proxy are: 1) Monoterpene concentrations measured during the whole day were used for the proxy. 2) The mixing within the boundary layer, diluting monoterpene concentration, was considered. (3) The oxidation of monoterpenes by nitrate radical (NO_3) was included.

The explanation of the improvements by Kontkanen et al. 2016 are discussed in the method's section 2.1.4 lines 122-127.

12. Line 197: It would be helpful to state the number of undefined days here, so the reader does not wonder how it can be that 877 days are events, 229 are non-events, and 55% of days are event days.

Based on the reviewer's suggestion we add the missing number of undefined days to line 198. "*with 877 days classified as NPF events, 560 undefined days and only 229 as non-events."*

13. Line 205 "days having less" -> "days with fewer"

See following comment.

14. Lines 204-209: Please rewrite or combine with the previous paragraph to ensure the message of this paragraph does not repeat the message of the previous paragraph

We modified the text in lines 204-209 based on the reviewer's suggestion.

15. Line 213: "In order to find out clear results and conclusions, we will focus on comparison between NPF events and non-events in following sections."-> this is long-winded, could shorten to "Undefined events are not considered further in the analysis"

We considered the reviewer's suggestion.

16. Line 230: If the annual trend is important to note, state explicitly what is the annual trend.

As stated in the next sentences, there was no clear trend, but the number of events varied from year to year. We modified the sentence to: *The total number of NPF events varied from year to year between 1996 and 2015.*

17. Line 235: What is the median and percentile of a trajectory? The median compass direction at the point on the trajectory where it arrives at Hyytiala, or the median compass direction of some kind of average over the length of the trajectory? Does "at every half hour" mean for the arrival of the air masses at Hyytiala every half hour or for one trajectory, moving back along it by half an hour at a time?

Based on the reviewer's comment and for clarity we modified the section explaining the median and percentile of the trajectories. *The medians and similarly the percentiles were calculated by taking the median compass direction at every point on the trajectory (1 hour between every two points), arriving every half an hour at Hyytiälä.*

18. Line 252: "However, the monthly cycle of CS on non-event days had two maxima, one in spring and another one in autumn"- what is the reader meant to conclude from this sentence?

We added a continuation to the sentence in line 252 to ensure clarity: *However, the monthly cycle of CS during non-event days had two maxima, one in spring and another one in autumn, which might suggest that during these seasons, high values of CS prevented NPF to occur during those particular days.*

19. Line 257: "The temperature at which clear-sky NPF events occurred was different for each month" - > The following sentences are not really 'examples' of this sentence. I would delete this.

We endorse the reviewer's suggestion.

20. Line 266: "even though it might also be attributed to the presumable increase" -> but it might also be attributed to the increase"

We did the change.

21. Line 279-281: this sentence needs a verb outside the "while" clause

See the following comment.

22. Line 281: Increased RH leads to increased production of H2SO4. Additionally, even with constant H2SO4 concentrations, nucleation rates increase with RH in flow tube or chamber studies (e.g. Duplissy et al, JGR 2015) and are expected to from theory (Merikanto et al, JGR 2015, Vehkamaki et al 2002). However, it is indeed clear from Fig 5c that RH is negatively correlated with nucleation. This could be due to any number of reasons, but it seems odd to point out the Boy & Kulmala study on RH limiting VOC ozonolysis without discussing the far more robust and well-established evidence from atmospheric chemistry that RH should promote nucleation of sulphuric acid.

Although increased RH leads to enhanced production of H_2SO_4 even with constant H_2SO_4 concentrations, nucleation rates increase with RH in flow tube or chamber studies (e.g.Duplissy and Flagan, 2016) and are expected to from theory (Merikanto et al., 2016;Vehkamäki et al., 2002). Our results show that RH is negatively correlated with nucleation, mainly because pure H_2SO_4 nucleation is not expected in Hyytiälä. However, in correspondence to the reviewer's suggestions we add more details to the paragraph referring to the effect of RH on NPF:

Previous studies in Hyytiälä have found that events are accompanied with lower RH values in comparison to non-events (Hamed et al., 2011). Other studies have proposed that increased RH limits some VOC (Volatile Organic Compounds) ozonolysis reactions, preventing the formation of come condensable vapors necessary for nucleation (Boy and Kulmala, 2002). This might partially explain the observed anti-correlation between RH and particle formation rates. Therefore, it seems plausible that RH affects NPF via atmospheric chemistry rather than via changing the sink term for condensing vapors and small clusters. Additionally, we found clear differences in how trace gas concentrations were associated with RH between the NPF event and non-event days (Table 1). For instance, O_3 showed a strong negative correlation with RH during events and non-events. However, during non-event days, a positive correlation appears between RH and each of CO, SO_2 and NOx while the correlation between those seems to be absent during event days. Our results show that air masses coming from central Europe and passing over the Baltic Sea tend to have higher values of RH.

23. Line 303: Specify that the OxOrg proxy concentration depends on temperature via the MT proxy in Kontkanen et al. Also see previous comment concerning this proxy.

A brief summary of the derivation of the OxOrg proxy concentration is present on lines 122 – 127.

24. Line 323: The comparison between J1.5 and J3 is interesting but should be made more explicit – how much later is the peaking time of J3 than J1.5? Are Figures 8b and 9b the same, or are there differences? Would you expect differences, based on how long it takes particles to grow from 1.5 to 3nm in general?

Differences could arise also if the growth from 1.5-3nm is the critical step for NPF, rather than the initial clustering forming 1.5nm particles, as is suggested by some studies (Kulmala et al., 2013). During clear-sky conditions, Table below summarizes the time delays until 3 nm particles are formed. Time delay is calculated as d_{dp}/GR. The peak times varies thus based on the GR and differs on each day. We modify the text in the manuscript (Line 329) to include the discussion related to the delay. Figure 9 is also replaced with the corrected one which includes the delay.

Time Delay	Median
Events	0.6 h
Non-events	0.4 h

'*On event days, in comparison to springtime $J_{1.5,C}$ which peaked at around 10:45 (Figure 8b), $J_{3,C}$ peaked typically about half-an hour later. This time delay indicates how long it takes for the particles grow from 1.5 nm to 3 nm. This growth is a critical step of NPF (Kulmala et al. 2013), and depends on concentrations of available vapour precursors.*'

25. Line 331 It is stated that figure 10 represents "median diurnal cycles". The description of what this actually means is currently a bit hard to follow, and it needs to be repeated in the figure caption. If I understand correctly, the median CS is calculated for each half hour, and plotted against the corresponding median J3 value. Perhaps it would be clearer to describe the plot by saying "the J3 and CS data were divided into 30 groups according to the time of day at which the data were recorded, and the median J3 and CS values for each group were calculated. The first group of data were recorded between 5am and 5.30am, the next from 5.30am to 6am, and so on until 8pm local time" (with adjustments for the precise times/numbers of groups). The figure would also be clearer if the scales of the axes were better optimised so that the data extend closer to the extremes of the axis ranges.

Figure 10 was modified into a clearer version. See comment 4 in the General comments.

26. Line 342. From Figure 10, it is interesting that in summer, autumn and winter the highest J3 on non-event days is almost as high as on event days, and one would therefore expect the J1.5 to be very similar to the J1.5 on event days. This is in sharp contrast to the large differences between event and non-event days shown in Figure 9b for spring. For autumn, winter and summer, figure 10 would imply that the J rate by itself is a poor predictor for whether or not an event will occur. This is surely a useful message for your paper: it could be used to emphasise the importance of your new discriminating variable, based on condensation sink, which from Figure 10 clearly should perform better than the nucleation rate, which naively sounds like a more obvious variable to determine whether or not a nucleation event is occurring.

Figure 10 is modified into a clearer version as well as the text accompanying it. See comment 4 in the General comments.

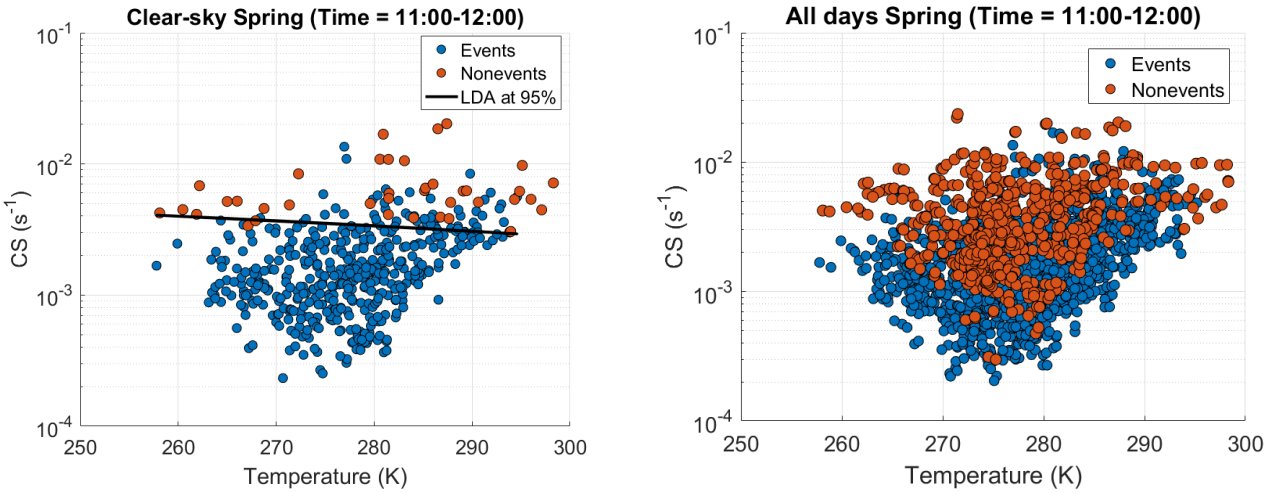
27. Figure 2: In addition to the helpful rows of numbers presented below the box plot, it would be helpful to state the number of event and non-event days with P > 0.7 in two additional rows.

Based on the suggestion, we added the corresponding data to Figure 2 and modified the figure caption.

28. Figure 12: why does the criterion for NPF you have developed perform badly in summer, autumn and winter? It would be possible to determine quantitatively the benefit of the clear-sky classification by applying the NPF criterion in the clear-sky case and also without first separating clear sky events from non-events. Please state the effectiveness of the criterion in the case where you do not distinguish clear-sky and cloudy events, in order to prove the usefulness of the clear-sky distinction by showing that the NPF criterion is less effective without it.

The differences between meteorological parameters is clearly lower in all other seasons in comparison to spring, which explains the limitation of our criterion. Although precursor vapors have high concentrations in summer, however, the concentrations are similar between event and non-event days which makes it hard to separate days into events and non-events based on vapor concentrations. We plotted below figure 12 (left) accompanied by a similar figure where the clear-sky classification had not been taken into account. In total, the number of total events and non-events were 1458 and 2118, respectively, while in clear-sky conditions the numbers were 877 and 229, respectively. Accordingly, and as shown in the plot below, if no clear-sky selection is done, it is

basically almost impossible to separate the events from the non-events. While in clear-sky conditions, it was possible to very well separate events from the non-events in spring, which makes clear-sky distinction useful.



29. References: should cite Kuang et al, ACP 2010, "An improved criterion for new particle formation in diverse atmospheric environments" somewhere

Kuang et al. (2010) developed a criterion for NPF probability based on a dimensionless parameter (ratio of particle loss rate to growth rate), which determines when the newly formed clusters are likely to grow to detectable sizes. They conclude this criterion to work in diverse environments, however, they did not explore the dependency of their parameter on atmospheric conditions. Line 383 now reads "Although previous studies have developed criteria for NPF probability which could work in diverse environments (Kuang et al., 2010), they did not explore the dependency of their criterion on atmospheric conditions."

30. Table 1: +/-0.45 does not really indicate "high correlation": for this description to be justified I think you need +/-0.7 at least! Also, since the tables are symmetric about the diagonal, please remove the lower triangle (or replace with "-") so the reader does not have to check the upper and lower triangles are the same.

We changed the coloring criteria in Table 1 so that only values higher than 0.7 are indicated as high.

31. Figures 1/2: what is the "relevant statistical limit"?

The default box plots refer to 99.3% statistical limit. For clarity the figure caption accompanying all boxplots is modified to: *The lines extending from the central box represent 1.5 x interquartile range which includes 99.3% of the data inclusive. Data outside this statistical limit are considered outliers and are marked with red crosses.*

32. Figure 3 caption: "5.4%, (add comma) making the classification biased." Please state more explicitly what you mean here: do you have only global radiation data for 5.4% of the days in 1998?

We did the change.

References

Boy, M., and Kulmala, M.: Nucleation events in the continental boundary layer: Influence of physical and meteorological parameters, Atmospheric Chemistry and Physics, 2, 1-16, 2002.

Duplissy, J., and Flagan, R.: Effect of ions on sulfuric acid-water binary particle formation: 2. Experimental data and comparison, Journal of Geophysical Research. Atmospheres, 121, 1752-1775, 2016.

Hamed, A., Korhonen, H., Sihto, S. L., Joutsensaari, J., Järvinen, H., Petäjä, T., Arnold, F., Nieminen, T., Kulmala, M., and Smith, J. N.: The role of relative humidity in continental new particle formation, Journal of Geophysical Research: Atmospheres, 116, 2011.

Hyvönen, S., Junninen, H., Laakso, L., Maso, M. D., Grönholm, T., Bonn, B., Keronen, P., Aalto, P., Hiltunen, V., and Pohja, T.: A look at aerosol formation using data mining techniques, Atmospheric Chemistry and Physics, 5, 3345-3356, 2005.

Kuang, C., Riipinen, I., Sihto, S.-L., Kulmala, M., McCormick, A., and McMurry, P.: An improved criterion for new particle formation in diverse atmospheric environments, Atmospheric Chemistry and Physics, 10, 8469-8480, 2010.

Kulmala, M., Kontkanen, J., Junninen, H., Lehtipalo, K., Manninen, H. E., Nieminen, T., Petäjä, T., Sipilä, M., Schobesberger, S., and Rantala, P.: Direct observations of atmospheric aerosol nucleation, Science, 339, 943-946, 2013.

Lappalainen, H., Sevanto, S., Bäck, J., Ruuskanen, T., Kolari, P., Taipale, R., Rinne, J., Kulmala, M., and Hari, P.: Day-time concentrations of biogenic volatile organic compounds in a boreal forest canopy and their relation to environmental and biological factors, Atmospheric Chemistry and Physics, 9, 5447-5459, 2009.

Merikanto, J., Duplissy, J., Määttänen, A., Henschel, H., Donahue, N. M., Brus, D., Schobesberger, S., Kulmala, M., and Vehkamäki, H.: Effect of ions on sulfuric acid-water binary particle formation: 1. Theory for kinetic- and nucleation-type particle formation and atmospheric implications, J. Geophys. Res. Atmos, 121, 1736-1751, 2016.

Vehkamäki, H., Kulmala, M., Napari, I., Lehtinen, K. E., Timmreck, C., Noppel, M., and Laaksonen, A.: An improved parameterization for sulfuric acid–water nucleation rates for tropospheric and stratospheric conditions, Journal of Geophysical Research: Atmospheres, 107, 2002.

Reply to Referee #2

We thank Referee #2 for their helpful suggestions. We replied to the comments below. The bold text refers to the referee's comments, and the text in italics are additions to the manuscript. The line numbers mentioned in the text below refer to the ACPD version of the manuscript.

I. General comments:

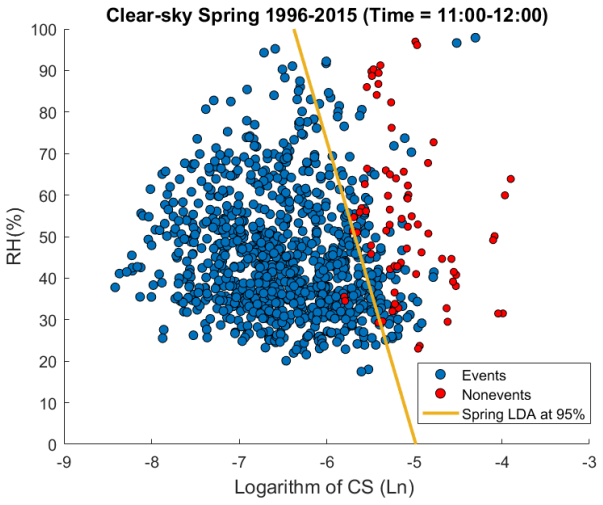
In their manuscript, the authors present an in-depth analysis of a long dataset of aerosol, meteorology, trace gas and irradiation measurements at the SMEAR station in Finland. The analysis is performed to find the key parameters that would explain new particle formation.

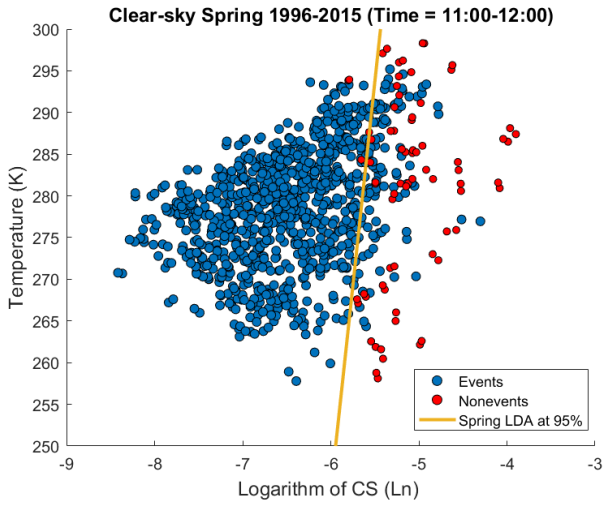
Similar analyses with the same datasets have already been performed several times, as explained by the authors. However, in this analysis the authors focus on eliminating the effect of cloudiness in the analysis, which is an interesting approach and merits publication in ACP. The data aquisition methods are described in good details, and the data analysis mostly follows the procedures described in the cited literature. Some of the specific methods for this paper could be described in more detail and the choices and justification for them should be explained in the text (see detailed comments).

1. A similar analysis without the cloudiness parameter has been performed earlier, it would be nice to see a direct comparison of the analysis of regarding the separation of events and non-events (Hyvönen et al., 2005). It should be quite straightforward to perform the same linear discriminant analysis as the Hyvönen paper for the CS and RH data (Fig 4 in the Hyvönen et al paper), and compare whether the result has changed.

Our results show that although the relative humidity seems to be a variable that influences NPF, when only clear-sky conditions are considered, the variation of RH between events and non-events does not seem to explain the occurrence of NPF. For instance, looking at figure 5c, we notice that although there might seem to be a difference in the median value of the RH when comparing event days and non-event days within each month, the percentiles seem to coincide minimizing the overall separation. However, we agree with the reviewer that it is important to compare with the suggested publication. Accordingly, we plotted clear-sky RH vs CS (below). The plots include spring clear-sky events and non-events within the time window 11:00-12:00 which has been proven shown to be the peak time of NPF Figures 8b and 9b. We also performed Linear Discriminant Analysis (LDA) according to Hyvönen *et al.* 2005 and added the 100% confidence level line to the plot. The 100% confidence limit corresponds to separating 95% of the non-events (to the right of the line in this case). We compare the corresponding spring RH vs CS plot with that of Temperature vs CS (below). The plots show that RH is as good as temperature under clear-sky conditions however it does not aid the separation (events from nonevents) further as CS sink seems to be the main controlling factor. We then conclude that during clear-sky conditions the results are somewhat different from what Hyvönen et al. 2005 who did not consider clear-sky conditions only. Based on the aforementioned results, and following the reviewer's suggestion, we add the following to the text to line 369:

Furthermore, we analyzed the effect of RH in separating the events from nonevents, similar to the study done on RH by Hyvönen et al. 2005. However, when plotting CS vs RH (data not presented), our results show no enhanced separation of events from non-events based on RH when only clear-sky conditions are considered.





2. Also, I think it should be made clear that the event probability described in Figure 13 and in section 3.3.4 is different from the equation 6, and also different from the event probability introduced in the Hyvänen et al paper. In the latter, the event probablility is computed from the LDA analysis, while in the current paper the probability seems to be directly calculated from data, and thus it is not a predictive equation. I suggest that the authors revise this part of the paper. Also, if no real propability-giving predictive equation is given, I think that aim IV in the Introduction (line 66) should be revised.

As the reviewer mentioned, the event probability presented in figure 13 and in section 3.3.4 is calculated directly using the current data set. However, introducing such results best explains the direct effect of extreme temperatures and condensation sink on classifying days as events or non-events. The aim IV in the introduction, refers to equation 6 which sets the line for variable separation during clear-sky events and non-events. First, to improve our analysis, we perform LDA analysis to our dataset, similar to the analysis presented by Hyvönen et al. 2005. That made rather equation (6) more reasonable. Accordingly, equation (6) and the corresponding figures 11 and 12 are improved and replaced. Second, to make the aims clearer, following the reviewer's comments, aim IV in line 66 is divided into two to show the independence of equation (6) accompanied by figures 11 and 12 and the figure 13 modified to "*iv) formulate an equation that predicts whether a clear-sky day with specific temperature and CS is classified as an event; v) use the clear-sky data set to calculate the NPF probability distribution based on temperature and CS*".

However, overall I think that the paper is a potentially good addition to the literature of understanding NPF, and its topic is certainly appropriate for ACP. Therefore, if the above corrections and the detailed notes given below can be considered by the authors, I would suggest publication. The corrections and revisions are, in my opinion, minor.

II. <u>Detailed notes:</u>

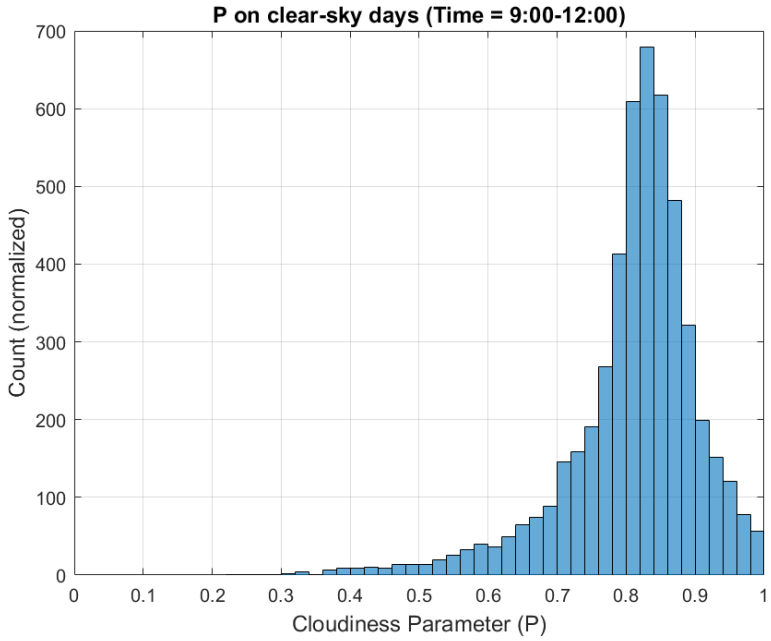
1. line 150-158 and 223-225: I do not fully understand the definition of the clear-sky day presented by the authors. Generally, it is known that particle formation occurs around noon, and that especially the mixing of the residual layer in the morning seems to have an influence. From that, I can follow that using the morning value is useful in the analysis. However, only the median P value for three hours is used. This raises the following questions: *i) Were only events that started during this three-hour window included in the analysis? *ii) Why was the median used? In this case, a time period that is for example 1 hour 29 minutes cloudy and 1 hour 31 minutes sunny gets classified as a sunny (clear-sky) day. Does the result change when the mean is used? *iii) what is the basis of using the value 0.7?

i) NPF occurs usually in the morning hours and peaks around noon (Dada et. al 2017, In Preparation). For that, favorable conditions for clustering should be available to initiate the process as well as ensure its continuation. Since the sun-cycle varies widely in Hyytiälä between seasons which affects the NPF cycle also, the time window 9:00-12:00 seems to cover all seasons equally. Also, although all events are included in this classification, the ones that occur outside our selected time window are rather very few. For consistency, the variables compared in our study are taken between the same time window 9:00-12:00. Also, we do not aim at studying exhaustively the day-by-day results, rather formulate a picture on the variables that are very different on event days in comparison to non-events. And since up to our knowledge, no clear method has been detected to identify the start time of an event, doing the data mining manually for event days of such a characteristic (start time between 9:00 and 12:00) is very time consuming and adds no heavy value for the current paper.

ii) We agree with the reviewer that it is tricky to define the day as clear/cloudy period by a median or a mean value; however, let's assume that 2 hours have P = 0.6 and the remaining 1 hour is P=0.9, this will result in a mean of 0.7, clear-sky day while this is not the reality. However, we recalculated the difference in frequency of event occurrence in case we choose the mean P value for these three hours instead of the median, the differences are acceptable demonstrated in the table below:

	Median	**Mean**	**Percentage difference**
Clear-sky events and non- events	1106	1045	5.5%

We consider this difference insignificant for our analysis. The median value is useful also because NPF is a regional-scale phenomenon, so for instance scattered clouds on an otherwise sunny day affecting the local radiation measurements (and leading to a momentarily drop in P) do not usually interrupt the regional NPF process. The histogram (Count Normalized) below also shows that within the clear-sky days, the P values calculated every half hour between 9:00 and 12:00 almost never reach a value below 0.3 (which is the threshold for complete sky cover). This result advocates the fact that if any clouds appear on a day classified as clear-sky they are mostly scattered and not thick.



iii) The value of 0.7 is found from previous studies which are mentioned in line 155. This has proven to be a value that works for different seasons, and is strict enough to exclude all totally or partly cloudy days, but not to eliminate sunny days with occasional scattered clouds passing over the station.

2. The reasoning between this central points in the methodology should be explained in much more detail, as I expect that similar analyses will be performed in the future for other sites, and therefore the method should be as robustly implemented as possible. Also, can the authors give insight on how sensitive the method is on the limit value of P chosen?

The sensitivity of the method on the limit P value is shown in Figure 1a which shows the variation of the fraction of days when using different P values. We added more details to the method section 2.2.2:

In Hyytiälä, the great majority of NPF events are initiated during the morning hours after the sunrise, yet before the noon (Dada et. al 2017, In Preparation). Since the time of the sunrise varies widely in Hyytiälä between the different seasons, the time window 9:00-12:00 seems a reasonable compromise for considering whether NPF did occur or not. We found that NPF events occurring outside our selected time window were very few. Accordingly, in this work the days were classified as cloudy or clear-sky days based on the median value of P during 9:00-12:00 each day, corresponding to the time window for new particle formation. Clear-sky days were those with a median of P > 0.7 between 9:00 and 12:00 and are the focus of this study. The median value ensures that at least half of our selected time window is clear-sky while the rest can vary between clear-sky and minor scattered clouds. The median is useful also because NPF is a regional-scale phenomenon, so for instance scattered clouds on an otherwise sunny day affecting the local radiation measurements (and leading to a momentarily drop in P) do not usually interrupt the regional NPF process. Clear-sky days were those with a median of P > 0.7 between 9:00 and 12:00 and are the focus of this study. For consistency, the variables compared in our study are taken between the same time window 9:00-12:00.

3. Line 198: ". . . radiation is essential for NPF as these events occur mainly during daylight hours." If radiation was essential, no NPF could be observed during nighttime. In the literature, several examples of NPF during nighttime can be found. Please rephrase.

We modified to *"radiation seems essential for NPF at this site, as the events occur almost solely during daylight hours."*

4. line 200: is SA really the main component of freshly formed particles? If heteromolecular nucleation is the prevailing mechanism, the the organic compound is as important. Both are still likely to be formed photochemically, so I think that this sentence can be fixed by just by rewording (e.g. '..because the main components of freshly formed particles are likely formed photochemically. . .')

As suggested by the reviewer, we did the change.

5. line 235-245: Please clarify also in the text and in the caption of Figure 4 that these results refer to clear-sky events only.

Line 235 is modified to *"The springtime medians are percentiles of air-mass trajectories arriving at Hyytiälä during clear-sky NPF events and non-events…."*

6. Line 251-254: As the CS is highest for event days, but not so for non-event days, does the presented conclusion that the CS is the reason for the minimum in events in summer really follow? It seems to me that in summer, events may occur despite high CS, and the actual reason for non-events is not the inhibiting effect of CS. If the authors disagree, this could be clarified.

After modifying figure 10 and the accompanying text based on both reviewers' suggestions (See comment 9 below), it appears clearer that in summer, the calculated formation rates are high also during nonevent days, yet an event is not happening. This might be explained by higher temperatures in summer which leaves the freshly formed clusters rather unstable.

7. Line 270: with monthly I think that the authors mean yearly

The whole section is rearranged to fit both reviewers' suggestions.

8. Line 280-281, '. . .low or almost no correlation. . .' something seems to be missing in this sentence.

Line 280 is modified to "*However, during non-event days, a positive correlation appears between RH and each of CO, SO_2 and NOx while the correlation between those seems to be absent during event days.*"

9. lines 331-350: I don't really understand what is shown in figure 10, and therefore also don't follow the explanation in this paragraph. What is meant by diurnal cycle here? By definition it means a repeating pattern that occurs every 24 hours, and I don't see how this could result in Figure 10. Please clarify and rewrite, or replace with the correct figure.

Figure 10 is replaced with median diurnal cycles of J_3 and CS during different seasons. The diurnal time frame is limited to 5:00-20:00 due to the incapability of calculating SA concentrations in the absence of UVB, therefore no $J_{3,C}$ values are calculated outside this time window.

10. Line 357: The procedure of finding the separating line in Fig 11-12 is described very poorly. Is this done by linear discriminant analysis (such as e.g. in the cited Hyvönen et al., (2005) paper or some other method? The authors should describe this in more detail. I'm especially concerned about the sentence "the data points have been estimated by taking the non-events with the lowest possible CS which still fit the linear ˘ separation"; was some kind of data selection applied to produce the figure?

See comment 2 in the General comments section

11. Figures: Several figures have the sentence "The lines extending 1.5 times from the central box represent the remaining of the data yet still within the relevant statistical limit. " Please clarify what this means: firstly, what is 1.5 times from the central box (the lines seem to have different lengths, eg. in fig. 5. Also, clarify what is meant by relevant statistical limit.

The extended lines are equal to 1.5 x interquartile range, and the points beyond whiskers are outliers (> 1.5 x interquartile range). Outliers are the individual stars. Maximum length of the extended line, is defined as $q3 + w \times (q3 - q1)$ while the minimum is $q1 - w \times (q3 - q1)$ where w, q1 and q3 are the mean, the 25th and 75th percentiles of the sample data, respectively. The statistical limit is defined by default as the 99.3% coverage. Based on the reviewer's comment we modify the text corresponding to the box plots for clarity to the following:

The length of the whiskers represent 1.5 x interquartile range which includes 99.3% of the data. Data outside the whiskers are considered outliers and are marked with red crosses.

References

L. Dada, R. Chellapermal, S. Buenrostro Mazon, V.M. Kerminen, P. Paasonen And M. Kulmala (2017). Method for identifying NPF event start and end times as well as NPF types (ion-initiated, particle initiated, transported..) using characteristic nucleation-mode particles and air ions. *In Preparation.*

Long-term analysis of clear-sky new particle formation events and non-events in Hyytiälä

Lubna Dada[1], Pauli Paasonen[1], Tuomo Nieminen [1,2], Stephany Buenrostro Mazon[1], Jenni Kontkanen[1], Otso Peräkylä[1], Katrianne Lehtipalo [1,4], Tareq Hussein[1,45], Tuukka Petäjä[1], Veli-Matti Kerminen[1], Jaana Bäck[3], and Markku Kulmala[1]

[1]Department of Physics, University of Helsinki, P.O. Box 64, FIN-00014 Helsinki, Finland
[2]Department of Applied Physics, University of Eastern Finland, P.O. Box 1627, FI-70211 Kuopio, Finland
[3]Department of Forest Sciences, University of Helsinki, P.O. Box 27, FIN-00014 Helsinki, Finland
[4]~~Department~~ Laboratory of Atmospheric Chemistry, Paul Scherrer Institute (PSI), 5232 Villigen PSI, Switzerland
[5]Department of Physics, the University of Jordan, Amman 11942, Jordan
Correspondence to: Lubna Dada (lubna.dada@helsinki.fi)

Abstract. New particle formation (NPF) events have been observed all around the world and are known to be a major source of atmospheric aerosol particles. Here we combine 20 years of observations in a boreal forest at the SMEAR II station (Station for Measuring Ecosystem-Atmosphere Relations) in Hyytiälä, Finland, by ~~utilizing~~building on previously accumulated knowledge, and by focusing on clear-sky (non-cloudy) conditions. We first investigated the effect of cloudiness on NPF and then compared the NPF event and non-event days during clear-sky conditions. In this comparison we considered, for example, the effects of calculated particle formation rates, condensation sink, trace gas concentrations and various meteorological quantities~~.~~ in discriminating NPF events from non-events. The formation rate of 1.5 nm particles was calculated by using proxies for gaseous sulfuric acid and oxidized products of low volatile organic compounds~~.~~, together with an empirical nucleation rate coefficient. As expected, our results indicate an increase in the frequency of NPF events under clear-sky conditions~~.~~ in comparison to cloudy ones. Also, focusing on clear-sky conditions enabled us to find a clear separation of many variables related to NPF. For instance, oxidized organic vapors showed higher concentration during the clear-sky NPF event days, whereas the condensation sink (CS) and some trace gases had higher concentrations during the non-event days. The calculated formation rate of 3 nm particles showed a notable difference between the NPF event and non-event days during clear-sky conditions, especially in winter and spring. For spring time, we are able to find a threshold ~~value~~equation for the combined values of ambient temperature and CS, (CS (s-1) > $-3.091\times10^{-5} \times$ T (in Kelvin) + 0.0120), above which practically no clear-sky NPF event could be observed. Finally, we present a probability distribution for the frequency of NPF events at a specific CS and temperature.

Keywords: Boreal forest, formation rate, atmospheric aerosols, aerosol dynamics, condensation sink, cloudiness parameter

1 Introduction

The effects of atmospheric aerosol particles on the climate system, human health and environmental interactions have raised the interest in various phenomena associated with the formation, growth and loss of these particles (Pöschl, 2005; Seinfeld and Pandis, 2012; Apte et al., 2015). While primary emissions are a very important source of atmospheric aerosol particles, especially in terms of the aerosol mass loading, the particle number concentration is greatly affected by atmospheric new particle formation (NPF). During the last couple of decades, NPF has been observed to take place almost all over the world (Kulmala et al., 2004; Zhang et al., 2011; Bianchi et al., 2016; Kontkanen et al., 2016a; Kontkanen et al., 2017). Atmospheric NPF is thought to be the dominant source of the total particle number concentration (Kulmala et al., 2016), and a major source of cloud condensation nuclei, in the global troposphere (Merikanto et al., 2009; Yu et al., 2010; Kerminen et al., 2012; Salma et al., 2016).

Understanding the NPF phenomenon requires understanding ~~the~~its precursors and pathways involved under different atmospheric conditions. For instance, high concentrations of low-volatility vapors result in a higher probability for NPF (Nieminen et al., 2015), whereas a high relative humidity and condensation sink tend to suppress NPF (Hyvönen et al., 2005; Nieminen et al., 2014). Recent laboratory experiments have shown the importance of sulfuric acid and low-volatile oxidized organic vapors ~~on~~to NPF (Metzger et al., 2010; Kirkby et al., 2011; Petäjä et al., 2011; Kulmala et al., 2013; Ehn et al., 2014; Riccobono et al., 2014). Additionally, atmospheric observations confirm the importance of these precursor vapors in the initial steps of NPF and in the further growth of newly-formed particles (Kulmala et al., 1998; Smith et al., 2005; Kerminen et al., 2010; Paasonen et al., 2010; Ahlm et al., 2012; Bzdek et al., 2014; Nieminen et al., 2014; Vakkari et al., 2015)~~.~~. The Station for Measuring Forest Ecosystem-Atmosphere Relations (SMEAR II), located in Hyytiälä, southern Finland, compiles almost 21 years of particle number size distribution and extensive complementary data, providing the longest size distribution time series in the world, and hence allows for robust NPF analysis which is not readily possible at other sites. The station is located in a homogenous Scots pine forest far from major pollution sources. Hyytiälä is therefore classified as a background site representative of the semi-clean northern hemisphere boreal forests.

~~Amongst the~~Many studies ~~investigating~~ have investigated the role of different variables in causing, enhancing or preventing new particle formation (Hyvönen et al., 2005; Baranizadeh et al., 2014; Nieminen et al., 2014)~~,~~. In particular, Baranizadeh et. al (2014) studied the effect of cloudiness on NPF events observed at SMEAR II in Hyytiälä~~, a boreal forest environment in Southern Finland. That study~~. They concluded, in agreement with some other studies, ~~the role of~~that clouds ~~in attenuating and interrupting~~tend to attenuate or interrupt NPF events (Sogacheva et al., 2008; Boulon et al., 2010; Baranizadeh et al., 2014; Nieminen et al., 2015) . In this study, we ~~eliminate~~eliminated one variable that limits NPF (cloudiness), in order to provide a better insight into the other ~~parameters~~quantities related to atmospheric NPF~~ rate and its probability.~~. Based on 20 ~~-~~years of observations and data analysis for the SMEAR II station in Hyytiälä, we aim to i) quantify the effect of cloudiness on new particle formation frequency, ii) characterize the differences between NPF event and non-event days during clear-sky conditions~~,~~ iii) ~~find out~~explore the ~~connection~~connections between ~~nucleating precursor vapors and~~ new particle formation rates ~~and~~calculated from precursor vapor proxies and the occurrence of NPF events, iv) formulate an equation that predicts whether ~~NPF occurs or not during~~a clear-sky ~~conditions~~day with specific

temperature and CS is classified as an event; v) use the clear-sky data set to calculate the NPF probability distribution based on temperature and CS.

2 Materials and methods

2.1 Measurements

The data used for the analysis in this study is from the University of Helsinki SMEAR (Station for Measurement of Ecosystem –Atmosphere Relations) II station (Hari and Kulmala, 2005). The station provides long-term continuous comprehensive measurements of quantities describing ~~the~~ atmospheric-forest-ecosystem interactions. The SMEAR II station is located in the boreal forest in Hyytiälä, southern Finland (61°51N', 24°17E', 181 m a.s.l.), 220 km NW of Helsinki. Tampere (200,000 inhabitants) is the largest city nearest to the station and is located 60 km SW of the site. Being far from major human activities and surrounded by a homogenous scots pine belt, Hyytiälä is considered as a rural background site due to the low levels of air pollutants (Asmi et al., 2011). A more detailed overview of the measurements at the station can be found in Hari and Kulmala (2005) and Nieminen et al. (2014).

In this study, the data analysis is based on four types of measurements: (i) aerosol particle number size distributions, (ii) concentration of the trace gases (CO, NO, NO_2, NO_x, SO_2 and O_3), (iii) meteorological parameters (solar radiation, temperature and relative humidity), and (iv) precursor vapor concentrations from previously-developed proxies. The collection of data started in January 1996. Trace gas concentrations are measured at 6 different heights on a 74-m-high mast (extended to 126 m in summer 2010). Gas concentrations used in this study are collected from the middle level on the mast above the forest (at 16.8 m).

~~2.1.1 Particle number size distributions~~

The aerosol number concentration size distributions were measured with a twin-DMPS (Differential Mobility Particle Sizer) system (Aalto et al., 2001) for the size ranges 3-500 nm until year 2004 and 3-1000 nm from 2005 onwards. ~~The~~These data ~~measured was~~were used to classify days as NPF events and non-events following the method proposed by Dal Maso et al. (2005). The size distributions obtained from the DMPS measurements were used to calculate the condensation sink, CS, which is equal to the rate at which non-volatile vapors condense onto a pre-existing aerosol particle population (Kulmala et al., 2012).

~~2.1.2 Trace Gases~~

The CO concentration is measured with one infrared light absorption analyzer (API 300EU, Teledyne Monitor Labs, Englewood, CO, USA). The NO and NOx concentrations are monitored with a chemiluminescence analyzer (TEI 42C TL, Thermo Fisher Scientific, Waltham, MA, USA). The NO_2 concentration is calculated from the difference NOx–NO. The detection limit is about 0.05 ppb. SO_2 measurements are made through a UV-fluorescence analyzer (TEI 43 CTL, Thermo Fisher Scientific, Waltham, MA, USA) that has a detection limit of 0.1 ppb. The O_3 concentration is measured with an UV light absorption analyzer (TEI 49C, Thermo Fisher Scientific, Waltham, MA, USA) that has a detection limit of about 1 ppb. The data for trace gases are available as 30-minute arithmetic means.

2.1.3 Radiation

Solar radiation in the wavelengths of UV-B (280 – 320nm) and global radiation (0.30 - 4.8 µm) are monitored using pyranometers (SL 501A UVB, Solar Light, Philadelphia, PA, USA; Reeman TP 3, Astrodata, Tõravere, Tartumaa, Estonia until June 2008, and Middleton Solar SK08, Middleton Solar, Yarraville, Australia since June 2008) above the forest at 18 m. The air temperature is measured with 4-wired PT-100 sensors, and the relative humidity (in percent) is measured with relative humidity sensors (Rotronic Hygromet MP102H with Hygroclip HC2-S3, Rotronic AG, Bassersdorf, Switzerland). These data are provided as 30-minute averages.

2.1.4 Sulfuric Acid and Oxidized Organics Proxies

The gaseous sulfuric acid concentration is estimated from a pseudo-steady-state approximation proxy developed by Petäjä et al. (2009). This proxy takes into consideration the sulfuric acid source and sink terms as

$$[H_2SO_4]_{proxy} = k \cdot \frac{[SO_2] \cdot UVB}{CS} \quad (1).$$

Here, UVB (W m^{-2}) is the fraction of the UV radiation reaching earth after being screened by ozone (280 – 320 nm) and the coefficient k (m^2 W^{-1} s^{-1}) is obtained from the comparison of the proxy concentration to the available measured H_2SO_4 data, and has a median value of 9.9×10^{-7} m^2 W^{-1} s^{-1}.

The concentration of monoterpene oxidation products, called oxidized organic compounds (OxOrg) here, is estimated using a proxy developed by . This proxy is calculated by using the concentrations of different oxidants (the measured ozone concentration [O_3] and parameterizations for the hydroxyl and nitrate radical concentration, [OH] and [NO_3], respectively) and their reaction rates, k_i, with the monoterpenes. The MT_{proxy} is calculated by taking into account the effect of temperature-driven emissions, mixing of the boundary layer and the oxidation of monoterpenes, .

$$[OxOrg]_{proxy} = \frac{(k_{OH+MT}[OH] + k_{O3+MT}[O_3] + k_{NO3+MT}[NO_3]) \cdot MT_{proxy}}{CS} \quad (2).$$

2.1.5 Backward Air-mass Trajectories

Air mass trajectories were calculated using Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT_4) Model at 96-hour backward trajectories at 100, 250 and 500 m arrival heights once per hour. Free access to transport model is developed and provided by NOAA (). Input meteorological data required for the model were collected from GDAS (Global Data Assimilation System) archives.

2.2 Data analysis

2.2.1 New particle formation events classification

Formation of new aerosol particles in Hyytiälä is typically observed in the time window of several hours around noon, while this phenomenon seems to be rare during nighttime (Junninen et al., 2008; Buenrostro Mazon et al., 2016). Accordingly, aerosol number size distributions data from the DMPS measurements at around this time window are used for classifying individual days as new particle formation event or non-event days. The classification follows the guidelines presented by Kulmala et al. (2012), and the procedure presented in Dal Maso et al. (2005)~~.~~. The latter uses a decision criterion based on the presence of particles < 25 nm in diameter and their consequent growth to Aitken mode. Event days are days on which sub 25 nm particle formation and growth are observed. Non-event days are days on which neither modes are present. Undefined days are the days which do not fit either criterion.

2.2.2 Selecting non-cloudy days

Cloudiness parameter (P) is the ratio of measured global radiation (Rd) divided by the theoretical global irradiance (Rg):

$$P = \frac{\mathrm{Rd}}{\mathrm{Rg}} \quad (1)$$

The theoretical maximum of global radiation (Rg) is calculated by taking into consideration the latitude of the measurement station and the seasonal solar cycle. While a complete cloud coverage is classified as $P < 0.3$, a clear-sky is classified as $P > 0.7$ (Perez et al., 1990; Sogacheva et al., 2008; Sánchez et al., 2012). In Hyytiälä, the great majority of NPF events are initiated during the morning hours after the sunrise, yet before the noon (Dada et al., 2017, in preparation)~~.~~. Since the time of the sunrise varies widely in Hyytiälä between the different seasons, the time window 9:00-12:00 seems a reasonable compromise for considering whether NPF did occur or not. We found that NPF events occurring outside our selected time window were very few. Accordingly, in this work the days were classified as cloudy or clear-sky days based on the median value of P during 9:00-12:00 each day, corresponding to the time window for new particle formation. Clear-sky days were those with a median of $P > 0.7$ between 9:00 and 12:00 and are the focus of this study. The median value ensures that at least half of our selected time window is clear-sky while the rest can vary between clear-sky and minor scattered clouds. The median is useful also because NPF is a regional-scale phenomenon, so for instance scattered clouds on an otherwise sunny day affecting the local radiation measurements (and leading to a momentarily drop in P) do not usually interrupt the regional NPF process. Clear-sky days were those with a median of $P > 0.7$ between 9:00 and 12:00 and are the focus of this study. For consistency, the variables compared in our study are taken from the same time window, 9:00-12:00.

2.2.3 Sulfuric acid and oxidized organics proxies

The gaseous sulfuric acid concentration is estimated from a pseudo-steady-state-approximation proxy developed by Petäjä et al. (2009). This proxy takes into consideration the sulfuric acid source and sink terms as

$$[H_2SO_4]_{proxy} = k \cdot \frac{[SO_2] \cdot UVB}{CS} \quad (2).$$

Here, UVB (W m^{-2}) is the fraction of the UV radiation reaching earth after being screened by ozone (280 – 320 nm) and the coefficient k (m^2 W^{-1} s^{-1}) is obtained from the comparison of the proxy concentration to the available measured H_2SO_4 data, and has a median value of 9.9×10^{-7} m^2 W^{-1} s^{-1}.

The concentration of monoterpene oxidation products, called oxidized organic compounds (OxOrg) here, is estimated using a proxy developed by Kontkanen et al. (2016b). This proxy is calculated by using the concentrations of different oxidants (the measured ozone concentration [O_3] and parameterizations for the hydroxyl and nitrate radical concentration, [OH] and [NO_3], respectively) and their reaction rates, k_i, with the monoterpenes. The MT_{proxy} (in this case $MT_{proxy1,doy}$) is calculated by taking into account the effect of temperature-driven emissions, mixing of the boundary layer and the oxidation of monoterpenes, (Kontkanen et al., 2016b).

$$[OxOrg]_{proxy} = \frac{(k_{OH+MT}[OH] + k_{O3+MT}[O_3] + k_{NO3+MT}[NO_3]) \cdot MT_{proxy}}{CS} \quad (3).$$

~~2.2.3~~2.2.4 Particle formation rates

The formation rate of nucleation mode particles ($J_{3,C}$, particle diameter > 3 nm) was calculated based on the method suggested by Kerminen and Kulmala's equation (Kerminen and Kulmala, 2002). This quantity is a function of the calculated formation rate of 1.5 nm sized particles ($J_{1.5,C}$), their growth rate (GR) and the condensation sink (CS):

$$J_{3,C} = J_{1.5,C} \exp\left(-\gamma \frac{CS'}{GR_{1.5-3}}\left(\frac{1}{1.5} - \frac{1}{3}\right)\right), \quad (4)$$

where γ is a coefficient with an approximate value of 0.23 m^3 nm^2 s^{-1}. The value of $J_{1.5,C}$ was calculated by assuming heteromolecular nucleation between SA and OxOrg as follows:

$$J_{1.5,C} = K_{het}[H_2SO_4]_{proxy}[OxOrg]_{proxy}, \quad (5)$$

The heterogeneous nucleation coefficient used in Equation 5 is the median estimated coefficient for Hyytiälä scaled from Paasonen et al. (2010): K_{het} = 9.2×10^{-14} cm^3 s^{-1}. The scaling was made in order to fit the current data. The median value of [OxOrg] during the event days in April and May was found to be 1.~~3×10^8~~6×10^7 cm^{-3}(Paasonen et al., 2010), whereas the ~~improved~~revised median value of [OxOrg] by Kontkanen et al. (~~2016~~2016b) is 1.~~6×10^7~~3×10^8 cm^{-3}. The scaling factor is the ratio between new and original [OxOrg] (0.1194~~);~~). Accordingly, while the value of K_{het} from Paasonen et al. (2010) is 1.1×10^{-13}~~14~~ cm^3 s^{-1}, after the scaling by 0.1194 we obtain the revised K_{het} = 9.2×10^{-14} cm^3 s^{-1}.

The particle growth rate over the particle diameter range of 1.5−3 nm was calculated by taking into account the size of the condensing vapor molecule size and the thermal speed of the particle (Nieminen et al., 2010). The growth rates (1.5 – 3 nm) were calculated as 30-minute averages and as the sum of the growth rates due to the sulfuric acid (SA) vapor and OxOrg vapor condensation. The density of the particle was assumed to be constant (1440 kg/m^3). For SA, we first

determined the SA concentration needed to make the particles grow at the rate of 1 nm/h by taking into account the mass of hydrated SA at the present RH and its density (Kurtén et al., 2007). Then, we calculated the GR of the particles due to SA condensation by using the SA proxy concentration. The same method was used for GR due to OxOrg condensation, where the vapor density was assumed to be 1200 kg/m^3 (Hallquist et al., 2009; Kannosto et al., 2008). Similarly, the GR due to OxOrg was calculated by using OxOrg proxy concentrations divided by the concentration needed for 1 nm/h GR.

2.2.5 Calculation of backward air-mass trajectories

Air mass trajectories were calculated using Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT_4) Model at 96-hour backward trajectories at 100, 250 and 500 m arrival heights once per hour. Free access to transport model is developed and provided by NOAA (http://www.ready.noaa.gov/HYSPLIT.php). Input meteorological data required for the model were collected from GDAS (Global Data Assimilation System) archives.

3 Results and discussion

3.1 Effect of cloudiness on NPF

We studied NPF events as a function of cloudiness. Figure 1a shows the fraction of event, non-event and undefined days as a function of cloudiness parameter. We can see that clear-sky conditions favor the occurrence of NPF: the less clouds there were, the higher was the fraction of NPF event days. For instance, for days with the cloudiness parameter of 0.3 or less, the fraction of event days was less than 0.1 of the total classified days. However, the fraction of NPF event days reached a maximum of around 0.55 during complete clear-sky conditions (P >0.7), with 877 days classified as NPF events, 560 undefined days and only 229 as non-events. ~~The pattern found in Figure 1a follows from the fact that radiation is essential for NPF as these events occur mainly during daylight hours . NPF is favored under abundant radiation conditions, since sulfuric acid, which is the main component of freshly formed particles, is mainly formed photochemically . The fraction of undefined days, however, remained constant regardless of cloudiness conditions.~~

~~In Figure 1b we show the medians and percentiles of cloudiness parameters during the time window 9:00-12:00. As expected, NPF events tended to occur preferentially on the days having less clouds.~~ On the NPF event days, the median cloudiness parameter P during the time window 9:00-12:00 was found to be 0.75~~,~~ (Fig. 1b), while the non-event days were characterized by lower values of ~~this parameter~~P (a median of around 0.25). Also, 75% of the NPF event days were found to have a cloudiness parameter larger than 0.5. The pattern found in Figure 1a follows from the fact that radiation seems essential for NPF at this site, as the events occur almost solely during daylight hours (Kulmala et al., 2004b). NPF is favored under abundant radiation conditions since the main components of freshly formed particles, are mainly formed photochemically (Petäjä et al., 2009; Ehn et al., 2014)~~Therefore,~~. The fraction of undefined days, however, remained constant regardless of cloudiness conditions.

Our results emphasize the fact that radiation favors NPF to occur ~~is emphasized and correspondingly cloud cover decreases~~, while clouds tend to decrease the probability of NPF. Undefined days were observed under cloudiness conditions that fell between those for NPF events and non-events. In general, undefined days can be interrupted NPF

events, or unclassified plumes of small particles due to pollution (Buenrostro Mazon et al., 2009). The interruption of a NPF event can be due to a change in the measured air mass, or due to attenuation of solar radiation caused by the appearance of a cloud during the event. ~~In order to find out clear results and conclusions, we will focus on comparison between NPF events and non-events in following sections.~~ We will not consider undefined days further in our analyses. The monthly variation of daily median cloudiness parameter within the time window of 9:00-12:00 during the classified days is shown in Figure 2. ~~While~~ Spring showed the best separation between the events and non-events in terms of the cloudiness parameter, while the separation became weaker during the summer and specifically for June and July. Taken together, Figures 1 and 2 emphasize the observation that the presence of clouds decreases the probability of NPF events.

3.2 General character of NPF on clear-sky days

Upon visualizing the cloudiness conditions during events and non-events, we chose a fixed constraint for clear-sky conditions ($P > 0.7$) during the time window of NPF (9:00-12:00) and will next focus on other parameters that distinguish NPF events from non-events.

The monthly distribution of the event fraction on clear-sky days appeared as double peaks in spring and autumn, with spring having a higher fraction of events (Figure 3a). The minimum fraction of NPF events was recorded in December. The fraction of non-event days peaked during winter with another peak in summer. ~~It is important to note the annual variation of the~~The total number of NPF events ~~during the years~~varied from year to year between 1996~~-~~ and 2015. However, this variation did not show any specific trend of frequency (Figure 3b), which is in agreement with previous statistics reported from studies that did not consider clear-sky classification (Nieminen et al., 2014).

3.2.1 Backward air mass trajectories during clear-sky NPF events and non-events

Since NPF is most frequent in spring, we dedicated our focus into this season (Figure 3a). The springtime medians and percentiles of air-mass trajectories arriving at Hyytiälä during ~~the~~clear-sky NPF events and non-events were calculated 96 hours backward in time at the 100-m, 250-m and 500-m arrival heights for the years 1996-2015. ~~The trajectories arriving to~~The medians and similarly the percentiles were calculated by taking the median compass direction at every point on the trajectory (1 hour between every two points), arriving every half an hour at Hyytiälä. The trajectories arriving at Hyytiälä at these three heights were quite similar, and those arriving at the 500-m height are shown in Figure 4. Medians and percentiles of the routes were calculated by taking the median of the trajectories at every half hour for spring time NPF event days and non-event days separately~~; spring time is chosen as it is the peak time of NPF.~~. During the NPF event days, the measured air masses were found to originate mainly from the north and to pass over Scandinavia before arriving ~~to~~at Hyytiälä. Similar to previously reported results, air masses arriving from the north and north-west directions result in clean air with low pollutant (particulate matter and trace gas) concentrations (Nieminen et al., 2015)~~. On the other hand,~~. During NPF the non-event days, air masses ~~originate~~originated from more polluted areas in Europe and Russia, resulting in elevated levels of condensation sink and other air pollutants in Hyytiälä, as ~~reported by~~also seen in previous studies (Sogacheva et al., 2005).

3.2.2 Influences of CS, meteorological parameters and trace gases

In Figure 5a we present the monthly variation of condensation sink during NPF events and non-events under daytime clear-sky conditions. NPF events tended to be favored by low values of CS throughout the year. In all the months~~,~~ except during summer, the 75th percentile of the event day values of CS was lower than the 25th percentile of the non-event day values of CS. On the NPF event days, CS had its maximum in summer, which might be one of the main reasons for the local minimum in the NPF event frequency during the summer months (Figure 3a). However, the monthly cycle of CS ~~on~~during non-event days had two maxima, one in spring and another one in autumn, which might suggest that during these seasons, high values of CS prevented NPF to occur on particular days. The difference in the value of CS between the NPF event and non-event days was the highest in March and the lowest during the summer months.

Figure 5b shows the monthly temperature conditions (*T*) during the daytime NPF events and non-events. ~~The temperature at which clear-sky NPF events occurred was different for each month. For example,~~While higher temperatures ~~were favorable for~~favored NPF during the months when the ~~overall~~average temperature was below 273.15 K (0º C) (months 1, 2, 3, 11 and 12~~). On~~), the ~~other hand, NPF events tend to occur at lower~~opposite was true at average temperatures ~~when the overall temperature was~~ above 273.15 K (0º C). The highest recorded temperature at which an event occurred during $P > 0.7$ sky was 300 K (25 ºC) and the minimum temperature was 252 K (-21 ºC). Accordingly, both very high ~~or~~and very low temperatures ~~are~~were not favorable conditions for NPF to occur. Although an increase in the ambient temperature results in higher concentrations of monoterpenes due to increased emissions, thereby favoring new particle formation and growth (Kulmala et al., 2004), Figure 5b shows that very high temperatures tend to suppress NPF. This latter feature is at least partly related to the positive relation between the ambient temperature and pre-existing aerosol loading (and hence CS) in Hyytiälä (Liao et al., 2014), even though it might also be attributed to the ~~presumable~~ increase in vapor evaporation coefficients, which results in less stable clusters at high temperatures (Paasonen et al., 2012).

Consistent with an earlier study (Hamed et al., 2011), our results indicate that NPF is favored by low values of ambient relative humidity in Hyytiälä (Fig. 5c). This observation does not conflict with chamber experiments (e.g.Duplissy and Flagan, 2016) or theory (Merikanto et al., 2016; Vehkamäki et al., 2002)~~The relative humidity (RH) appeared to be lower on NPF event days (9:00-12:00) compared with non-events (9:00-12:00)days (Figure 5c). The RH has a monthly cycle with highest values in winter and lowest ones in summer, opposite to that of *T* and CS. The difference in RH medians between the NPF events and non-events was the highest in winter and became almost negligible in August. Many speculations have been presented on the effect of high RH on suppressing NPF. For instance, a high RH tends to increase the sinks of aerosol particles and their precursor vapors (CS and coagulation sinks), lowering the NPF probability . However, in this study no positive correlation between CS and RH was observed during the clear-sky conditions (Table 1). Previous studies in Hyytiälä~~, which suggest higher nucleation rates at higher values of RH, because binary H_2SO_4–water nucleation is not expected take place in Hyytiälä. Other studies have proposed that increased RH limits some VOC (Volatile Organic Compounds) ozonolysis reactions, preventing the formation of ~~certain~~come condensable vapors necessary for nucleation (Boy and Kulmala, 2002)~~.~~. This might partially explain the observed anti-correlation between RH and particle formation rates. Therefore, it seems plausible that RH affects NPF via atmospheric chemistry rather than via changing the sink term for condensing vapors and small clusters. Additionally, we found clear differences in how trace gas concentrations were associated with RH between the NPF event and non-event days (Table 1). For instance, O_3

showed a strong negative correlation with RH during events and non-events. However, ~~while a very strong~~during non-event days, a positive correlation ~~appeared~~appears between RH and each of CO, SO_2 and NOx ~~during non-event days, low or almost no~~ while the correlation ~~with~~between those seems to be absent during event days. ~~Therefore, it seems plausible~~Our results show that ~~RH affects NPF via atmospheric chemistry rather than via changing the sink term for condensing vapors~~air masses coming from central Europe and ~~small clusters~~passing over the Baltic Sea tend to have higher values of RH.

After looking at the characteristics of clear-sky NPF event and non-event days in terms of meteorological parameters and CS, we looked at the variation of trace gas (CO, SO_2, NOx and O_3) concentrations during these conditions (Figure 6). Out of these gases, at least SO_2 and O_3 are expected to enhance NPF, SO_2 as a precursor for sulfuric acid and O_3 as an oxidant forming ELVOCs (Extremely Low Volatile Organic Compounds) (Donahue et al., 2012; Ehn et al., 2014). However, none of these vapors seemed to have higher concentrations during NPF event days. This suggests that, as tracers of pollution, these gases are strongly linked with high anthropogenic CS, so air masses having high trace gas concentrations often do not result in NPF in Hyytiälä.

3.3 Connection of nucleating precursor vapors with new particle formation rate

3.3.1 Precursor vapor proxies

In this study, we determined $J_{1.5,C}$ using the proxies for both SA and OxOrg. The monthly variations of these precursors (in the time window 9:00-12:00) are shown in the Figure 7. During clear-sky conditions, the SA proxy tended to have the highest median daytime values during the winter months with a maximum in February (Figure 7a). Contrary to this, the seasonal distribution of the SA proxy reported in Hyytiälä appears as double peaks with an absolute maximum in spring and a smaller one in autumn when presenting the data without excluding cloudy days (Nieminen et al., 2014). During winter, both condensation sink and boundary layer height are lower than in the summer (Paasonen et al., 2013), which might explain the higher concentrations of SA during the winter months.

Being a function of temperature, the OxOrg proxy concentration was generally found to follow the monthly cycle of the ambient temperature. The median value of [OxOrg] was higher on NPF events days during every month compared with non-event days (Figure 7b). The ~~biggest~~largest difference in [OxOrg] between the NPF events and non-events, in terms of its median value, was recorded for January and the least difference for May. It is to be noted that the proxy values represent the measured values less accurately during the winter than during the other periods (Kontkanen et al., ~~2016~~2016b).

3.3.2 Particle formation rates

The calculated new particle formation rate, $J_{1.5,C}$, approximated with Eq. (5) shows a similar behavior as the [OxOrg ~~precursor~~] (see Figures 7 and 8), being higher for the clear-sky NPF event days in comparison with non-event days. Also, the difference in the value of $J_{1.5,C}$ between the NPF events and non-events was the highest in the winter, and the lowest in summer. The monthly cycle of $J_{1.5,C}$ followed closely that of ~~the~~ [OxOrg ~~concentration~~], as the latter ~~has~~had a higher

seasonal variability ~~and is therefore~~than the sulfuric acid proxy concentration, being thereby capable of affecting the seasonal pattern of $J_{1.5,C}$ (Figure 8a). ~~Also,~~ The diurnal cycle of $J_{1.5,C}$ during the NPF event days showed an increase along with sunrise, a peak at midday and decrease along with sunset. However, for non-event days the $J_{1.5,C}$ value was relatively constant throughout the day and had clearly lower values than during the NPF event days (Figure 8b).

Since previous studies have shown that there is a clear difference in observed J_3 between the event and non-event days, and much less difference in observed $J_{1.5}$ (Kulmala et al. 2013), we decided to focus on J_3 in our event to non-event discrimination. Previous studies which did not consider clear-sky conditions have reported values of observed spring time J_3 between 0.01 and 5 cm^{-3} s^{-1} (median = 0.94 cm^{-3} s^{-1}) during the period of active NPF (Kulmala et al. 2013). Our values of $J_{3,C}$ fit between the extremes of these values for the spring time and time window 9:00 to 12:00, with a slightly higher median value of 1.9 cm^{-3} s^{-1} (Figures 9 a,b). The formation rate of 3 nm particles is affected not only by the new particle formation rate ($J_{1.5}$) but also by the scavenging of newly-formed particles by coagulation into pre-existing particles. We found that, in general, the values of $J_{3,C}$ calculated using equations (4) and (5) were higher on NPF event days compared with non-event days in all months (Figure 9a). The difference between the event and non-event days was the largest in winter and ~~then~~ decreased towards summer. However, the diurnal cycles of percentiles and medians of $J_{3,C}$ during each month peaked around noon for both NPF events and non-events. One example is presented in Figure 9b, showing that $J_{3,C}$ tended to increase after the sunrise, to peak at about midday and to diminish after sunset. This kind of diurnal cycle was similar for all the months. Hourly values of $J_{3,C}$ calculated during the NPF event days were higher than those during the non-event days. During the spring months, the difference in the ~~media~~median $J_{3,C}$ between the NPF events and non-events, calculated for every half an hour, appeared to increase at about 10:00 and then it started to decrease again at about 13:00 (Figure 9b). On NPF event days, in comparison to springtime $J_{1.5,C}$ which peaked at around 10:45 (Figure 8b), $J_{3,C}$ peaked typically about half-an hour later. This time delay indicates how long it takes for the particles grow from 1.5 nm to 3 nm. This growth is a critical step of NPF (Kulmala et al. 2013), and depends on concentrations of available vapor precursors.

In Figure 10 we present the median diurnal cycles of $J_{3,C}$ plotted against the median diurnal cycles of CS during classified clear-sky NPF events and non-events. The diurnal cycle was calculated by taking the median CS at every half hour throughout the season. On the NPF event days, the CS had higher values during the nighttime and lower values during daytime with a minimum at noon. It is important to remember that J was calculated only for daytime when the SA proxy was available (UV-B radiation is needed for the proxy). On non-event days, the values of CS showed no clear diurnal pattern, had practically no difference between the daytime and nighttime hours, and were roughly twice those recorded during the clear-sky NPF event days. The difference in CS between NPF events and non-events follows from the distinctly different air masses arriving at Hyytiälä. For instance, it has been shown that air-masses originating from the north and passing over Scandinavia have, on average, lower values of CS than the air masses passing over Russia and central Europe (Sogacheva et al., 2005; Nieminen et al., 2015).

On NPF event days, the median approximated formation rate of 3 nm particles had its maximum value at about midday and was significantly higher than ~~that~~ on non-events days (~~Figure~~Figures 9b~~). While~~ and 10). A clear negative relation could be seen between the median seasonal diurnal cycles of CS and $J_{3,C}$ on NPF event days (specifically during ~~daytime), this kind of relation has not been observed during non-event days (Figure 10). Higher values of CS on non-event days~~

~~is~~spring daytime) (Figure 10). This kind of relation was not observed during non-event days when these two quantities seemed to be independent of each other (Figure 10). In summer, the median value of $J_{3,C}$ was roughly similar between NPF events and non-events, whereas the median value of CS was almost ten times higher during the non-event days compared with event days. The high values of $J_{3,C}$ for the non-event days in summer, despite the high CS values, seem to suggest that some other factor limits the actual NPF rate. One possibility is that freshly-formed clusters are rapidly evaporated due to higher ambient temperatures (see Fig. 5b). This will be discussed in a more detail in the following section. Higher values of CS on non-event days are expected, bearing in mind that these particles act as surfaces for scavenging precursor gases and freshly formed particles (Hussein et al., 2008). The association of a high CS with the lower NPF probability has been observed in many studies conducted in Hyytiälä (Boy and Kulmala, 2002; Hyvönen et al., 2005; Baranizadeh et al., 2014), as well as in other rural and urban areas, including Egbert and Toronto in Canada (Jun et al., 2014), Preila in Lithuania (Mordas et al., 2016), Po Valley in Italy (Hamed et al., 2007) and Budapest and K-puszta in Hungary (Salma et al., 2016).

3.3.3 Threshold separating the NPF events and non-events

Since quite a visible separation could be observed in the calculated values of $J_{3,C}$ between the spring-time clear-sky NPF events and non-events, and since $J_{3,C}$ had its maximum at around midday, the plot of CS versus temperature at midday (11:00-12:00) in spring provides an equation that effectively separates the NPF events from non-events during this season (Figure 11). ~~The~~This equation ~~is calculated~~was determined using a linear ~~fit that draws~~discriminant analysis (LDA) similar to Hyvönen et al. (2005). The equation provides a line ~~between the points that separate the maximum number of~~that separates NPF events ~~data points~~ from ~~the~~ non-events~~; the data points have been estimated by taking the~~ at 95% confidence towards non-events ~~with the lowest possible~~. Based on their midday CS ~~which still fit the linear separation~~and Temperature, the data point follows either classes. More specifically, the days with

$$\text{CS}\ (s^{-1}) > \text{~~9.034~~}\text{3.091}\times10^{-5} \times T\ (\text{in Kelvin}) \text{ ~~–~~+ 0.~~0213~~0120}, \quad (6)$$

lie above the threshold line ~~and were almost solely~~. Almost no non-event days ~~(<2% NPF event days), whereas the days classified as NPF events were mostly~~fall below this line~~.~~ (< 5%). The points above the line were also characterized with higher trace gases concentrations and lower calculated formation rates of 3 nm particles than the rest of the points.

The separation between the clear-sky NPF events and non-events in the CS versus T plot was less evident in autumn and disappeared completely in the summer and winter (Figure 12). Interestingly, ~~yet more than 95%~~a large number of ~~the~~ NPF event days during these seasons still fell below the threshold line given by Equation 6. Furthermore, we analyzed the effect of RH in separating the events from nonevents, similar to the study done on RH by Hyvönen et al. (2005). We found that compared with CS vs temperature data, depicting CS vs RH (data not presented) did not work better in separating NPF events from non-events during clear-sky conditions.

3.3.4 Probability of NPF events and non-events

Since the biggest difference in the calculated 3 nm particle formation rates between the NPF events and non-events was observed around noon (Figure 9b), and since CS and temperature showed promising threshold values for predicting the occurrence of NPF non-events during spring (up to ~~98~~95%) (Figure 11), Figure 13 presents the probability of having a NPF event in Hyytiälä at a specific CS and temperature within the time window 11:00-12:00. The probability ~~is~~was calculated taking the fraction of events to the total events and non-events in every cell which is ~~0.15 steps in CS and 2.5 steps in temperature.~~2.5 K on the x-axis and a ratio of 1.14 on the y-axis between every two consecutive CS values. The highest probability of having a NPF event corresponded to conditions having moderate temperatures and low values of CS. At high values of CS ~~levels~~, there was a zero probability for NPF regardless of the temperature. However, at moderate and low values of CS, the probability of having a NPF event decreases as we go to lower temperature. This could be explained by lower emissions of VOCs, and thus lower OxOrg concentrations, at lower temperatures. Similarly, the probability of NPF decreases as we go to higher temperatures at constant values of CS. This latter feature might be attributed to conditions unfavorable for clustering due to high temperatures. Although previous studies have developed criteria for NPF probability which could work in diverse environments(Kuang et al., 2010), they did not explore the dependency of their parameter on atmospheric conditions.

4 Conclusion

In this study we combined 20 years of data collected ~~in~~at the SMEAR II station in order to characterize the conditions affecting the frequency of NPF events in that location. By focusing only on clear-sky conditions, we were able to get new insight into differences between the NPF events and non-events. In clear-sky conditions, the meteorological conditions, trace gas concentrations and other studied variables on NPF event days appeared to be similar to those presented in the previous studies which did not consider clear-sky classification. Furthermore, the monthly data refined the analysis so that the differences caused by different quantities became more visible compared the previous studies conducted for this site. Our work confirms, with a complementary dataset, the conclusions of Baranizadeh et al. (2014) that NPF events and non-events are typically associated with clear-sky and cloudy conditions, respectively.

Our results showed that using SA and OxOrg proxies to calculate the apparent formation rates of 1.5 and 3 nm particles works well in differentiating the clear-sky NPF events from non-events. Moreover, during clear-sky conditions the effect of CS on attenuating or even preventing NPF was quite visible: CS was, on average, two times higher on the non-event days compared with the NPF event days. Similarly, many other meteorological variables affected NPF. By using CS and ambient temperature, we were able to find a threshold above which no clear-sky NPF events occurred. This threshold is described with an equation that is able to separate ~~almost 98~~97.4% of the NPF events from non-events during spring time. In clear sky conditions, when there is plenty of radiation available, NPF events ~~happen~~take place as long as the CS is low enough and temperature is moderate. Although a weaker separation was observed in the other seasons, considering only clear-sky conditions enabled us to form a map of the probability of having a NPF event within specific CS and temperature conditions. Using clear-sky conditions appears to bring us one step forward towards understanding NPF and predicting their occurrences in Hyytiälä. Our study serves as a basis to future detailed comparisons with observations to formulate even more robust conclusions.

5 Acknowledgements

This work was supported by the Academy of Finland Centre of Excellence program (grant no. 272041) and Nordic Top-level Research Initiative (TRI) Cryosphere-Atmosphere Interactions in a Changing Arctic Climate (CRAICC). The authors thank the division of atmospheric sciences at the University of Helsinki. We also thank Mrs. Ksenia Tabakova for providing air-mass trajectory data.

6 References

Aalto, P., Hämeri, K., Becker, E., Weber, R., Salm, J., Mäkelä, J. M., Hoell, C., O'dowd, C. D., Hansson, H.-C., Väkevä, M., Koponen, I. K., Buzorius, G., and Kulmala, M.: Physical characterization of aerosol particles during nucleation events, Tellus B, 53, 344-358 10.1034/j.1600-0889.2001.530403.x, 2001.

Ahlm, L., Liu, S., Day, D. A., Russell, L. M., Weber, R., Gentner, D. R., Goldstein, A. H., DiGangi, J. P., Henry, S. B., Keutsch, F. N., VandenBoer, T. C., Markovic, M. Z., Murphy, J. G., Ren, X., and Scott, S.: Formation and growth of ultrafine particles from secondary sources in Bakersfield, California, J. Geophys. Res. Atmos, 117, 10.1029/2011JD017144, 2012.

Apte, J. S., Marshall, J. D., Cohen, A. J., and Brauer, M.: Addressing global mortality from ambient PM2. 5, Environmental science & technology, 49, 8057-8066, 10.1021/acs.est.5b01236, 2015.

Asmi, A., Wiedensohler, A., Laj, P., Fjaeraa, A.-M., Sellegri, K., Birmili, W., Weingartner, E., Baltensperger, U., Zdimal, V., and Zikova, N.: Number size distributions and seasonality of submicron particles in Europe 2008–2009, Atmos. Chem. Phys., 11, 5505-5538, 10.5194/acp-11-5505-2011, 2011.

Baranizadeh, E., Arola, A., Hamed, A., Nieminen, T., Mikkonen, S., Virtanen, A., Kulmala, M., Lehtinen, K., and Laaksonen, A.: The effect of cloudiness on new-particle formation: investigation of radiation levels, Boreal Env. Res., 19, 343-354, 2014.

Bianchi, F., Tröstl, J., Junninen, H., Frege, C., Henne, S., Hoyle, C., Molteni, U., Herrmann, E., Adamov, A., Bukowiecki, N., Chen, X., Duplissy, J., Gysel, M., Hutterli, M., Kangasluoma, J., Kontkanen, J., Kürten, A., Manninen, H. E., Münch, S., Peräkylä, O., Petäjä, T., Rondo, L., Williamson, C., Weingartner, E., Curtius, J., Worsnop, D. R., Kulmala, M., Dommen, J., and Baltensperger, U.: New particle formation in the free troposphere: A question of chemistry and timing, Science, 352, 1109-1112, 10.1126/science.aad5456, 2016.

Boulon, J., Sellegri, K., Venzac, H., Picard, D., Weingartner, E., Wehrle, G., Collaud Coen, M., Bütikofer, R., Flückiger, E., and Baltensperger, U.: New particle formation and ultrafine charged aerosol climatology at a high altitude site in the Alps (Jungfraujoch, 3580 m asl, Switzerland), Atmos. Chem. Phys., 10, 9333-9349, 10.5194/acp-10-9333-2010, 2010.

Boy, M., and Kulmala, M.: Nucleation events in the continental boundary layer: Influence of physical and meteorological parameters, Atmos. Chem. Phys., 2, 1-16, 10.5194/acp-2-1-2002, 2002.

Buenrostro Mazon, S., Riipinen, I., Schultz, D., Valtanen, M., Maso, M. D., Sogacheva, L., Junninen, H., Nieminen, T., Kerminen, V.-M., and Kulmala, M.: Classifying previously undefined days from eleven years of aerosol-particle-size distribution data from the SMEAR II station, Hyytiälä, Finland, Atmos. Chem. Phys., 9, 667-676, 10.5194/acp-9-667-2009, 2009.

Buenrostro Mazon, S., Kontkanen, J., Manninen, H. E., Nieminen, T., Kerminen, V.-M., and Kulmala, M.: A long-term comparison of nighttime cluster events and daytime ion formation in a boreal forest, Boreal Env. Res., 21, 242-261, 2016.

Bzdek, B. R., Lawler, M. J., Horan, A. J., Pennington, M. R., DePalma, J. W., Zhao, J., Smith, J. N., and Johnston, M. V.: Molecular constraints on particle growth during new particle formation, Geophys. Res. Lett., 41, 6045-6054, 10.1021/ac100856j, 2014.

Dada, L., Chellapermal, R., Buenrostro Mazon, S., Junninen, H., Kerminen, V. M., Paasonen, P., and Kulmala, M.: Method for identifying NPF event start and end times as well as NPF types (ion-initiated, particle initiated, transported..) using characteristic nucleation-mode particles and air ions, 2017, in preparation.

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

Donahue, N. M., Kroll, J., Pandis, S. N., and Robinson, A. L.: A two-dimensional volatility basis set–Part 2: Diagnostics of organic-aerosol evolution, Atmos. Chem. Phys., 12, 615-634, 10.5194/acp-12-615-2012, 2012.

Duplissy, J., and Flagan, R.: Effect of ions on sulfuric acid-water binary particle formation: 2. Experimental data and comparison, J. Geophys. Res. Atmos, 121, 1752-1775, 10.1002/2015JD023539, 2016.

Ehn, M., Thornton, J. A., Kleist, E., Sipilä, M., Junninen, H., Pullinen, I., Springer, M., Rubach, F., Tillmann, R., Lee, B., Lopez-Hilfiker, F., Andres, S., Acir, I.-H., Rissanen, M., Jokinen, T., Schobesberger, S., Kangasluoma, J., Kontkanen, J., Nieminen, T., Kurtén, T., Nielsen, L. B., Jørgensen, S., Kjaergaard, H. G., Canagaratna, M., Maso, M. D., Berndt, T., Petäjä, T., Wahner, A., Kerminen, V.-M., Kulmala, M., Worsnop, D. R., Wildt, J., and Mentel, T. F.: A large source of low-volatility secondary organic aerosol, Nature, 506, 476-479, 10.1038/nature13032, 2014.

Hallquist, M., Wenger, J., Baltensperger, U., Rudich, Y., Simpson, D., Claeys, M., Dommen, J., Donahue, N., George, C., and Goldstein, A.: The formation, properties and impact of secondary organic aerosol: current and emerging issues, Atmos. Chem. Phys., 9, 5155-5236, 10.5194/acp-9-5155-2009, 2009.

Hamed, A., Joutsensaari, J., Mikkonen, S., Sogacheva, L., Maso, M. D., Kulmala, M., Cavalli, F., Fuzzi, S., Facchini, M., and Decesari, S.: Nucleation and growth of new particles in Po Valley, Italy, Atmos. Chem. Phys., 7, 355-376, 10.5194/acp-7-355-2007, 2007.

Hari, P., and Kulmala, M.: Station for measuring ecosystem-atmosphere relations, Boreal Env. Res., 10, 315-322, 2005.

Hussein, T., Martikainen, J., Junninen, H., Sogacheva, L., Wagner, R., Dal Maso, M., Riipinen, I., Aalto, P. P., and Kulmala, M.: Observation of regional new particle formation in the urban atmosphere, Tellus B, 60, 509-521, 10.1111/j.1600-0889.2008.00365.x, 2008.

Hyvönen, S., Junninen, H., Laakso, L., Maso, M. D., 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, Atmos. Chem. Phys., 5, 3345-3356, 10.5194/acp-5-3345-2005, 2005.

Jun, Y.-S., Jeong, C.-H., Sabaliauskas, K., Leaitch, W. R., and Evans, G. J.: A year-long comparison of particle formation events at paired urban and rural locations, Atmospheric Pollution Research, 5, 447-454, 10.5094/APR.2014.052, 2014.

Junninen, H., Hulkkonen, M., Riipinen, I., Nieminen, T., Hirsikko, A., Suni, T., Boy, M., LEE, S. H., Vana, M., Tammet, H., KERMINEN, V.-M., and KULMALA, M.: Observations on nocturnal growth of atmospheric clusters, Tellus B, 60, 365-371, 10.1111/j.1600-0889.2008.00356.x, 2008.

Kannosto, J., Virtanen, A., Lemmetty, M., Mäkelä, J. M., Keskinen, J., Junninen, H., Hussein, T., Aalto, P., and Kulmala, M.: Mode resolved density of atmospheric aerosol particles, Atmos. Chem. Phys., 8, 5327-5337, 10.5194/acp-8-5327-2008, 2008.

Kerminen, V.-M., and Kulmala, M.: Analytical formulae connecting the "real" and the "apparent" nucleation rate and the nuclei number concentration for atmospheric nucleation events, Journal of Aerosol Science, 33, 609-622, 10.1016/S0021-8502(01)00194-X, 2002.

Kerminen, V.-M., Petäjä, T., Manninen, H., Paasonen, P., Nieminen, T., Sipilä, M., Junninen, H., Ehn, M., Gagné, S., Laakso, L., Riipinen, I., Vehkamäki, H., Kurten, T., Ortega, I. K., Maso, M. D., Brus, D., Hyvärinen, A., Lihavainen, H., Leppä, J., Lehtinen, K. E. J., Mirme, A., Mirme, S., Hõrrak, U., Berndt, T., Stratmann, F., Birmili, W., Wiedensohler, A., Metzger, A., Dommen, J., Baltensperger, U., Kiendler-Scharr, A., Mentel, T. F., Wildt, J., Winkler, P. M., Wagner, P. E., Petzold, A., Minikin, A., Plass-Dülmer, C., Pöschl, U., Laaksonen, A., and Kulmala, M.: Atmospheric nucleation: highlights of the EUCAARI project and future directions, Atmos. Chem. Phys., 10, 10829-10848, 10.5194/acp-10-10829-2010, 2010.

Kerminen, V.-M., Paramonov, M., Anttila, T., Riipinen, I., Fountoukis, C., Korhonen, H., Asmi, E., Laakso, L., Lihavainen, H., Swietlicki, E., Svenningsson, B., Asmi, A., Pandis, S. N., Kulmala, M., and Petäjä, T.: Cloud condensation nuclei production associated with atmospheric nucleation: a synthesis based on existing literature and new results, Atmos. Chem. Phys., 12, 12037-12059, 10.5194/acp-12-12037-2012, 2012.

Kirkby, J., Curtius, J., Almeida, J., Dunne, E., Duplissy, J., Ehrhart, S., Franchin, A., Gagné, S., Ickes, L., Kürten, A., Kupc, A., Metzger, A., Riccobono, F., Rondo, L., Schobesberger, S., Georgios Tsagkogeorgas, Daniela

Wimmer, Antonio Amorim, Bianchi, F., Martin Breitenlechner, André David, Josef Dommen, Downard, A., Ehn, M., Flagan, R. C., Haider, S., Hansel, A., Hauser, D., Jud, W., Junninen, H., Kreissl, F., Kvashin, A., Laaksonen, A., Lehtipalo, K., Lima, J., Lovejoy, E. R., Makhmutov, V., Mathot, S., Mikkilä, J., Minginette, P., Sandra Mogo, Nieminen, T., Onnela, A., Pereira, P., Petäjä, T., Schnitzhofer, R., Seinfeld, J. H., Sipilä, M., Stozhkov, Y., Stratmann, F., Tomé, A., Vanhanen, J., Viisanen, Y., Vrtala, A., Wagner, P. E., Walther, H., Weingartner, E., Wex, H., Winkler, P. M., Carslaw, K. S., Worsnop, D. R., Baltensperger, U., and Kulmala, M.: Role of sulphuric acid, ammonia and galactic cosmic rays in atmospheric aerosol nucleation, Nature, 476, 429-433, 10.1038/nature10343, 2011.

Kontkanen, J., Järvinen, E., Manninen, H. E., Lehtipalo, K., Kangasluoma, J., Decesari, S., Gobbi, G. P., Laaksonen, A., Petäjä, T., and Kulmala, M.: High concentrations of sub-3nm clusters and frequent new particle formation observed in the Po Valley, Italy, during the PEGASOS 2012 campaign, Atmos. Chem. Phys., 16, 1919-1935, 10.5194/acp-16-1919-2016, 2016a.

Kontkanen, J., Paasonen, P., Aalto, J., Bäck, J., Rantala, P., Petäjä, T., and Kulmala, M.: Simple proxies for estimating the concentrations of monoterpenes and their oxidation products at a boreal forest site, Atmos. Chem. Phys., 16, 13291-13307, 10.5194/acp-16-13291-2016, 2016b.

Kontkanen, J., Lehtipalo, K., Ahonen, L., Kangasluoma, J., Manninen, H. E., Hakala, J., Rose, C., Sellegri, K., Xiao, S., Wang, L., Qi, X., Nie, W., Ding, A., Yu, H., Lee, S., Kerminen, V. M., Petäjä, T., and Kulmala, M.: Measurements of sub-3 nm particles using a particle size magnifier in different environments: from clean mountain top to polluted megacities, Atmos. Chem. Phys., 17, 2163-2187, 10.5194/acp-17-2163-2017, 2017.

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

Kulmala, M., Toivonen, A., Mäkelä, J. M., and Laaksonen, A.: Analysis of the growth of nucleation mode particles observed in Boreal forest, Tellus B, 50, 449-462, 10.3402/tellusb.v50i5.16229, 1998.

Kulmala, M., Vehkamäki, H., Petäjä, T., Dal Maso, M., Lauri, A., Kerminen, V.-M., Birmili, W., and McMurry, P. H.: Formation and growth rates of ultrafine atmospheric particles: a review of observations, Journal of Aerosol Science, 35, 143-176, 10.1016/j.jaerosci.2003.10.003, 2004.

Kulmala, M., Suni, T., Lehtinen, K. E. J., Dal Maso, M., Boy, M., Reissell, A., Rannik, Ü., Aalto, P., Keronen, P., Hakola, H., Bäck, J., Hoffmann, T., Vesala, T., and Hari, P.: A new feedback mechanism linking forests, aerosols, and climate, Atmos. Chem. Phys.
, 4, 557-562, 10.5194/acp-4-557-2004, 2004b.

Kulmala, M., Petäjä, T., Nieminen, T., Sipilä, M., Manninen, H. E., Lehtipalo, K., Dal Maso, M., Aalto, P. P., Junninen, H., and Paasonen, P.: Measurement of the nucleation of atmospheric aerosol particles, Nature protocols, 7, 1651-1667, 10.1038/nprot.2012.091, 2012.

Kulmala, M., Kontkanen, J., Junninen, H., Lehtipalo, K., Manninen, H. E., Nieminen, T., Petäjä, T., Sipilä, M., Schobesberger, S., Rantala, P., Franchin, A., Jokinen, T., Järvinen, E., Äijälä, M., Kangasluoma, J., Hakala, J., Aalto, P., Paasonen, P., Mikkilä, J., Vanhanen, J., Aalto, J., Hakola, H., Makkonen, U., Ruuskanen, T., Mauldin, R. r., Duplissy, J., Vehkamäki, H., Bäck, J., Kortelainen, A., Riipinen, I., Kurtén, T., Johnston, M., Smith, J., Ehn, M., Mentel, T., Lehtinen, K., Laaksonen, A., Kerminen, V., and Worsnop, D.: Direct observations of atmospheric aerosol nucleation, Science, 339, 943-946, 10.1126/science.1227385, 2013.

Kulmala, M., Luoma, K., Virkkula, A., Petäjä, T., Paasonen, P., Kerminen, V.-M., Nie, W., Qi, X., Shen, Y., and Chi, X.: On the mode-segregated aerosol particle number concentration load, Boreal Env. Res., 2016.

Kurtén, T., Torpo, L., Ding, C. G., Vehkamäki, H., Sundberg, M. R., Laasonen, K., and Kulmala, M.: A density functional study on water-sulfuric acid-ammonia clusters and implications for atmospheric cluster formation, J. Geophys. Res. Atmos, 112, 10.1029/2006JD007391, 2007.

Liao, L., Kerminen, V.-M., Boy, M., Kulmala, M., and Dal Maso, M.: Temperature influence on the natural aerosol budget over boreal forests, Atmos. Chem. Phys., 14, 8295-8308, 10.5194/acp-14-8295-2014, 2014.

Merikanto, J., Spracklen, D., Mann, G., Pickering, S., and Carslaw, K.: Impact of nucleation on global CCN, Atmos. Chem. Phys., 9, 8601-8616, 10.5194/acp-9-8601-2009, 2009.

Merikanto, J., Duplissy, J., Määttänen, A., Henschel, H., Donahue, N. M., Brus, D., Schobesberger, S., Kulmala, M., and Vehkamäki, H.: Effect of ions on sulfuric acid-water binary particle formation: 1. Theory for kinetic-and nucleation-type particle formation and atmospheric implications, J. Geophys. Res. Atmos, 121, 1736-1751, 10.1002/2015JD023538, 2016.

Metzger, A., Verheggen, B., Dommen, J., Duplissy, J., Prevot, A. S., Weingartner, E., Riipinen, I., Kulmala, M., Spracklen, D. V., and Carslaw, K. S.: Evidence for the role of organics in aerosol particle formation under atmospheric conditions, Proceedings of the National Academy of Sciences, 107, 6646-6651, 10.1073/pnas.0911330107, 2010.

Mordas, G., Plauškaitė, K., Prokopciuk, N., Dudoitis, V., Bozzetti, C., and Ulevicius, V.: Observation of new particle formation on Curonian Spit located between continental Europe and Scandinavia, Journal of Aerosol Science, 10.1016/j.jaerosci.2016.03.002, 2016.

Nieminen, T., Lehtinen, K., and Kulmala, M.: Sub-10 nm particle growth by vapor condensation–effects of vapor molecule size and particle thermal speed, Atmos. Chem. Phys., 10, 9773-9779, 10.5194/acp-10-9773-2010, 2010, 2010.

Nieminen, T., Asmi, A., Dal Maso, M., Aalto, P. P., Keronen, P., Petäjä, T., Kulmala, M., and Kerminen, V.-M.: Trends in atmospheric new-particle formation: 16 years of observations in a boreal-forest environment, Boreal Env. Res., 19, 2014.

Nieminen, T., Yli-Juuti, T., Manninen, H., Petäjä, T., Kerminen, V.-M., and Kulmala, M.: Technical note: New particle formation event forecasts during PEGASOS–Zeppelin Northern mission 2013 in Hyytiälä, Finland, Atmos. Chem. Phys., 15, 12385-12396, 10.5194/acp-15-12385-2015, 2015.

Paasonen, P., Nieminen, T., Asmi, E., Manninen, H., Petäjä, T., Plass-Dülmer, C., Flentje, H., Birmili, W., Wiedensohler, A., Horrak, U., Metzger, A., Hamed, A., Laaksonen, A., Facchini, M. C., Kerminen, V.-M., and Kulmala, M.: On the roles of sulphuric acid and low-volatility organic vapours in the initial steps of atmospheric new particle formation, Atmos. Chem. Phys., 10, 11223-11242, 10.5194/acp-10-11223-2010, 2010.

Paasonen, P., Olenius, T., Kupiainen, O., Kurtén, T., Petäjä, T., Birmili, W., Hamed, A., Hu, M., Huey, L., Plass-Duelmer, C., Smith, J. N., Wiedensohler, A., Loukonen, V., McGrath, M. J., Ortega, I. K., Laaksonen, A., Vehkamäki, H., Kerminen, V.-M., and Kulmala, M.: On the formation of sulphuric acid–amine clusters in varying atmospheric conditions and its influence on atmospheric new particle formation, Atmos. Chem. Phys., 12, 9113-9133, 10.5194/acp-12-9113-2012, 2012.

Paasonen, P., Asmi, A., Petäjä, T., Kajos, M. K., Äijälä, M., Junninen, H., Holst, T., Abbatt, J. P., Arneth, A., Birmili, W., Gon, H. D. v. d., Hamed, A., Hoffer, A., Laakso, L., Laaksonen, A., Leaitch, W. R., Plass-Dülmer, C., Pryor, S. C., Räisänen, P., Swietlicki, E., Wiedensohler, A., Worsnop, D. R., Kerminen, V.-M., and Kulmala, M.: Warming-induced increase in aerosol number concentration likely to moderate climate change, Nature Geoscience, 6, 438-442, 10.1038/ngeo1800, 2013.

Perez, R., Ineichen, P., Seals, R., and Zelenka, A.: Making full use of the clearness index for parameterizing hourly insolation conditions, Solar Energy, 45, 111-114, 10.1016/0038-092X(90)90036-C, 1990.

Petäjä, T., Mauldin Iii, R., Kosciuch, E., McGrath, J., Nieminen, T., Paasonen, P., Boy, M., Adamov, A., Kotiaho, T., and Kulmala, M.: Sulfuric acid and OH concentrations in a boreal forest site, Atmos. Chem. Phys., 9, 7435-7448, 10.5194/acp-9-7435-2009, 2009.

Petäjä, T., Sipilä, M., Paasonen, P., Nieminen, T., Kurtén, T., Ortega, I. K., Stratmann, F., Vehkamäki, H., Berndt, T., and Kulmala, M.: Experimental observation of strongly bound dimers of sulfuric acid: Application to nucleation in the atmosphere, Physical review letters, 106, 228302, 10.1103/PhysRevLett.106.228302, 2011.

Pöschl, U.: Atmospheric aerosols: composition, transformation, climate and health effects, Angewandte Chemie International Edition, 44, 7520-7540, 10.1002/anie.200501122, 2005.

741 Riccobono, F., Schobesberger, S., Scott, C. E., Dommen, J., Ortega, I. K., Rondo, L., Almeida, J., Amorim, A., Bianchi,
742 F., Breitenlechner, M., David, A., Downard, A., Dunne, E. M., Duplissy, J., Ehrhart, S., Flagan, R. C., Franchin,
743 A., Hansel, A., Junninen, H., Kajos, M., Keskinen, H., Kupc, A., Kürten, A., Kvashin, A. N., Laaksonen, A.,
744 Lehtipalo, K., Makhmutov, V., Mathot, S., Nieminen, T., Onnela, A., Petäjä, T., Praplan, A. P., Santos, F. D.,
745 Schallhart, S., Seinfeld, J. H., Sipilä, M., Spracklen, D. V., Stozhkov, Y., Stratmann, F., Tomé, A.,
746 Tsagkogeorgas, G., Vaattovaara, P., Viisanen, Y., Vrtala, A., Wagner, P. E., Weingartner, E., Wex, H., Wimmer,
747 D., Carslaw, K. S., Curtius, J., Donahue, N. M., Kirkby, J., Kulmala, M., Worsnop, D. R., and Baltensperger,
748 U.: Oxidation products of biogenic emissions contribute to nucleation of atmospheric particles, Science, 344,
749 717-721, 10.1126/science.1243527, 2014.
750
751 Salma, I., Németh, Z., Kerminen, V.-M., Aalto, P., Nieminen, T., Weidinger, T., Molnár, Á., Imre, K., and Kulmala, M.:
752 Regional effect on urban atmospheric nucleation, Atmos. Chem. Phys., 16, 8715-8728, 10.5194/acp-16-8715-
753 2016, 2016.
754
755 Sánchez, G., Serrano, A., and Cancillo, M.: Effect of cloudiness on solar global, solar diffuse and terrestrial downward
756 radiation at Badajoz (Southwestern Spain), Optica pura y aplicada, 45, 33-38, 2012.
757
758 Seinfeld, J. H., and Pandis, S. N.: Atmospheric chemistry and physics: from air pollution to climate change, John Wiley
759 & Sons, 2012.
760
761 Smith, J. N., Moore, K. F., Eisele, F. L., Voisin, D., Ghimire, A. K., Sakurai, H., and McMurry, P. H.: Chemical
762 composition of atmospheric nanoparticles during nucleation events in Atlanta, J. Geophys. Res. Atmos, 110,
763 10.1029/2005JD005912, 2005.
764
765 Sogacheva, L., Dal Maso, M., Kerminen, V.-M., and Kulmala, M.: Probability of nucleation events and aerosol particle
766 concentration in different air mass types arriving at Hyytiälä, southern Finland, based on back trajectories
767 analysis, Boreal Env. Res., 10, 2005.
768
769 Sogacheva, L., Saukkonen, L., Nilsson, E., Dal Maso, M., Schultz, D. M., De Leeuw, G., and Kulmala, M.: New aerosol
770 particle formation in different synoptic situations at Hyytiälä, southern Finland, Tellus B, 60, 485-494,
771 10.1111/j.1600-0889.2008.00364.x, 2008.
772
773 Vakkari, V., Tiitta, P., Jaars, K., Croteau, P., Beukes, J. P., Josipovic, M., Kerminen, V. M., Kulmala, M., Venter, A. D.,
774 and Zyl, P. G.: Reevaluating the contribution of sulfuric acid and the origin of organic compounds in atmospheric
775 nanoparticle growth, Geophys. Res. Lett., 42, 10.1002/2015GL066459, 2015.
776
777 Vehkamäki, H., Kulmala, M., Napari, I., Lehtinen, K. E., Timmreck, C., Noppel, M., and Laaksonen, A.: An improved
778 parameterization for sulfuric acid–water nucleation rates for tropospheric and stratospheric conditions, J.
779 Geophys. Res. Atmos, 107, 10.1029/2002JD002184, 2002.
780
781 Yu, F., Luo, G., Bates, T. S., Anderson, B., Clarke, A., Kapustin, V., Yantosca, R. M., Wang, Y., and Wu, S.: Spatial
782 distributions of particle number concentrations in the global troposphere: Simulations, observations, and
783 implications for nucleation mechanisms, J. Geophys. Res. Atmos, 115, 10.1029/2009JD013473, 2010.
784
785 Zhang, R., Khalizov, A., Wang, L., Hu, M., and Xu, W.: Nucleation and growth of nanoparticles in the atmosphere, Chem.
786 Rev., 112, 1957-2011, 10.1021/cr2001756, 2011.
787
788

Table 1 Correlation ~~coefficient~~coefficients between different meteorological parameters, gas concentrations and condensation sink (CS) during clear-sky events and non-events during spring (Mar-May, 1996-2015) and time window 9:00 – 12:00. The light blue refers to medium positive correlation (>0.45) and the dark blue to high positive correlation (>0.~~45), Red~~7), light orange refers to ~~high~~medium negative correlation (<-0.45) and dark red to high negative ones (<-0.7).

	CS	T	RH	CO	NO_X	SO_2	O_3
	Events						
CS	1						
T	0.28	1					
RH	-0.06	**-0.64**	1				
CO	0.33	-0.37	0.26	1			
NO_X	**0.53**	-0.19	0.21	**0.47**	1		
SO_2	0.4	-0.29	0.14	0.36	**0.58**	1	
O_3	0.23	**0.52**	**-0.51**	-0.06	-0.08	-0.08	1
	Non-Events						
CS	1						
T	0.15	1					
RH	-0.12	**-0.81**	1				
CO	**0.53**	**-0.68**	**0.5**	1			
NO_X	0.34	**-0.51**	**0.45**	**0.7**	1		
SO_2	0.23	**-0.55**	0.42	**0.56**	0.41	1	
O_3	0.43	**0.62**	**-0.64**	-4E-04	-0.07	-0.13	1

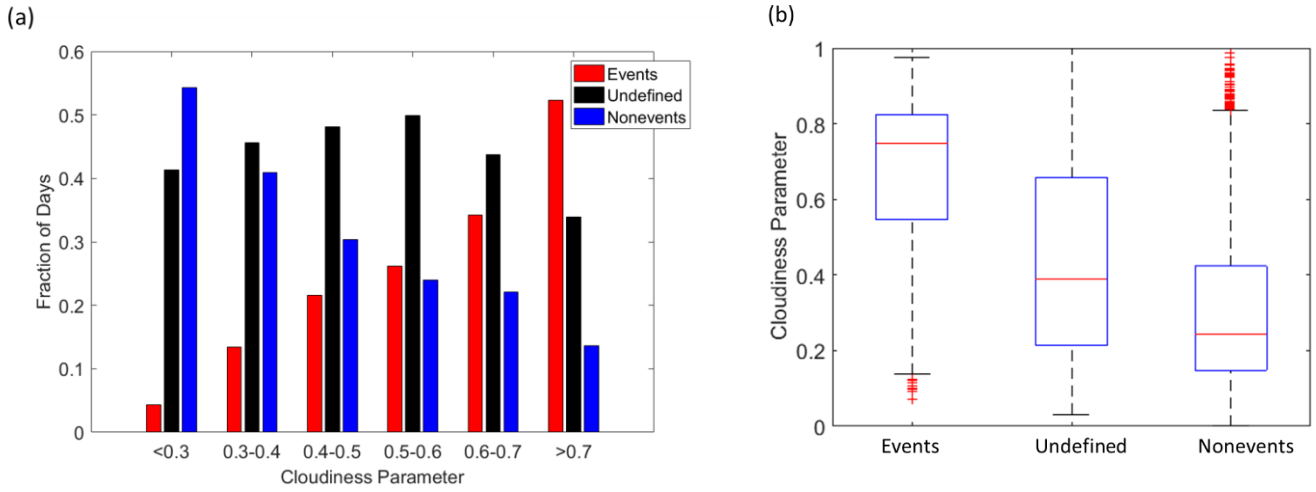
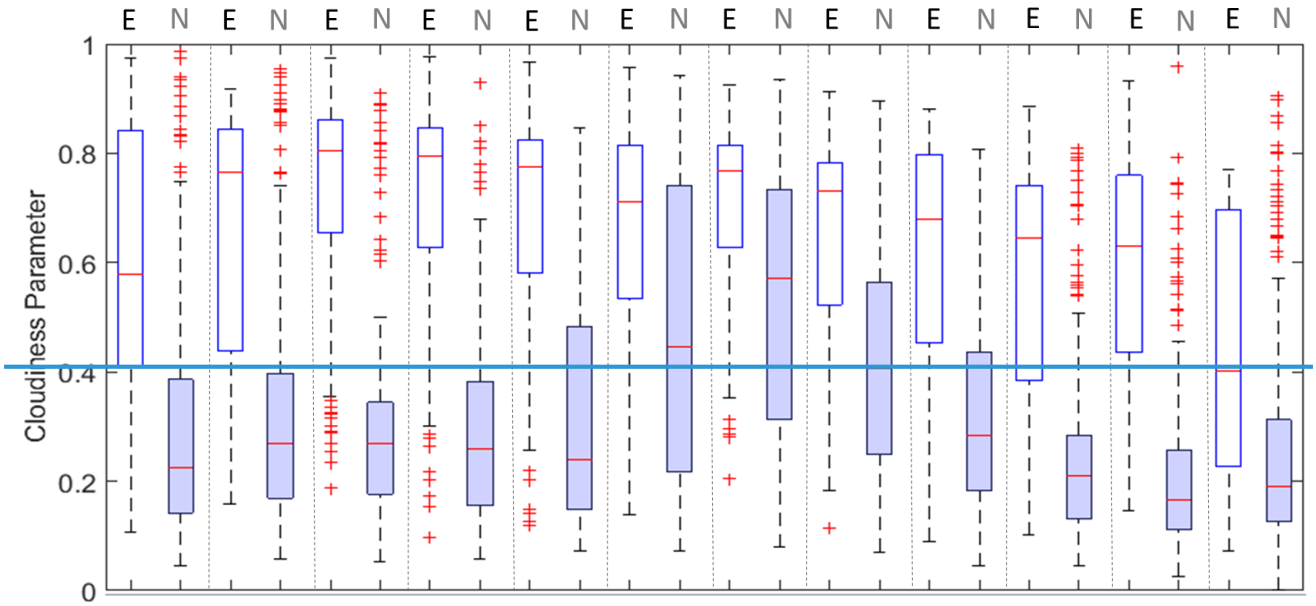
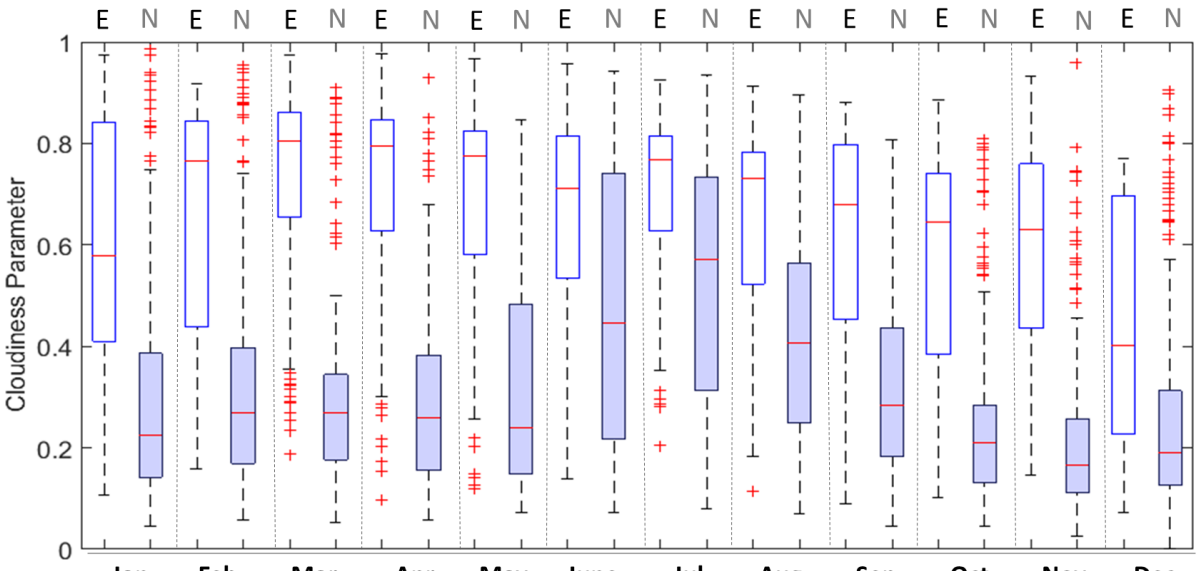


Figure 1: (a) Figure showing the fraction of days which are classified as NPF events, non-events, and undefined days during different sky cloudiness conditions. (b) Cloudiness daily (9:00 – 12:00) medians and percentiles recorded during NPF event, undefined and non-event days. The red line represents the median of the data and the lower and upper edges of the box represent 25th and 75th percentiles of the data respectively. The length of the whiskers represent 1.5 x interquartile range which includes 99.3% of the data. Data outside the whiskers are considered outliers and are marked with red crosses.



	Jan	Feb	Mar	Apr	May	June	Jul	Aug	Sep	Oct	Nov	Dec
E	27	68	277	296	267	162	94	110	188	115	30	17
N	316	218	143	125	110	138	192	162	158	229	308	318



	Jan	Feb	Mar	Apr	May	June	Jul	Aug	Sep	Oct	Nov	Dec
E	27	68	277	296	267	162	94	110	188	115	30	17
N	316	218	143	125	110	138	192	162	158	229	308	318
E (*P* >0.7)	9	30	165	180	151	83	59	60	82	43	10	5
N (*P* >0.7)	18	18	16	10	11	31	68	21	9	6	4	17

Figure 2: Monthly variation of cloudiness daily (9:00 – 12:00) medians and percentiles recorded during NPF events (E; white) and non-events (N; shaded). Numbers below the plot correspond to the number of data points included in each boxplot. ~~**The red line represents the median of the data and the lower and upper edges of the box represent 25th and 75th percentiles of the data respectively. The lines extending 1.5 times from the central box represent the remaining of the data yet still within the relevant statistical limit. The outliers are represented by the red crosses**~~**Number of clear-sky events (E (P>0.7)) and clear-sky non-events (N (P>0.7)) accompany the plot. See Figure 1 for explanation of symbols.**

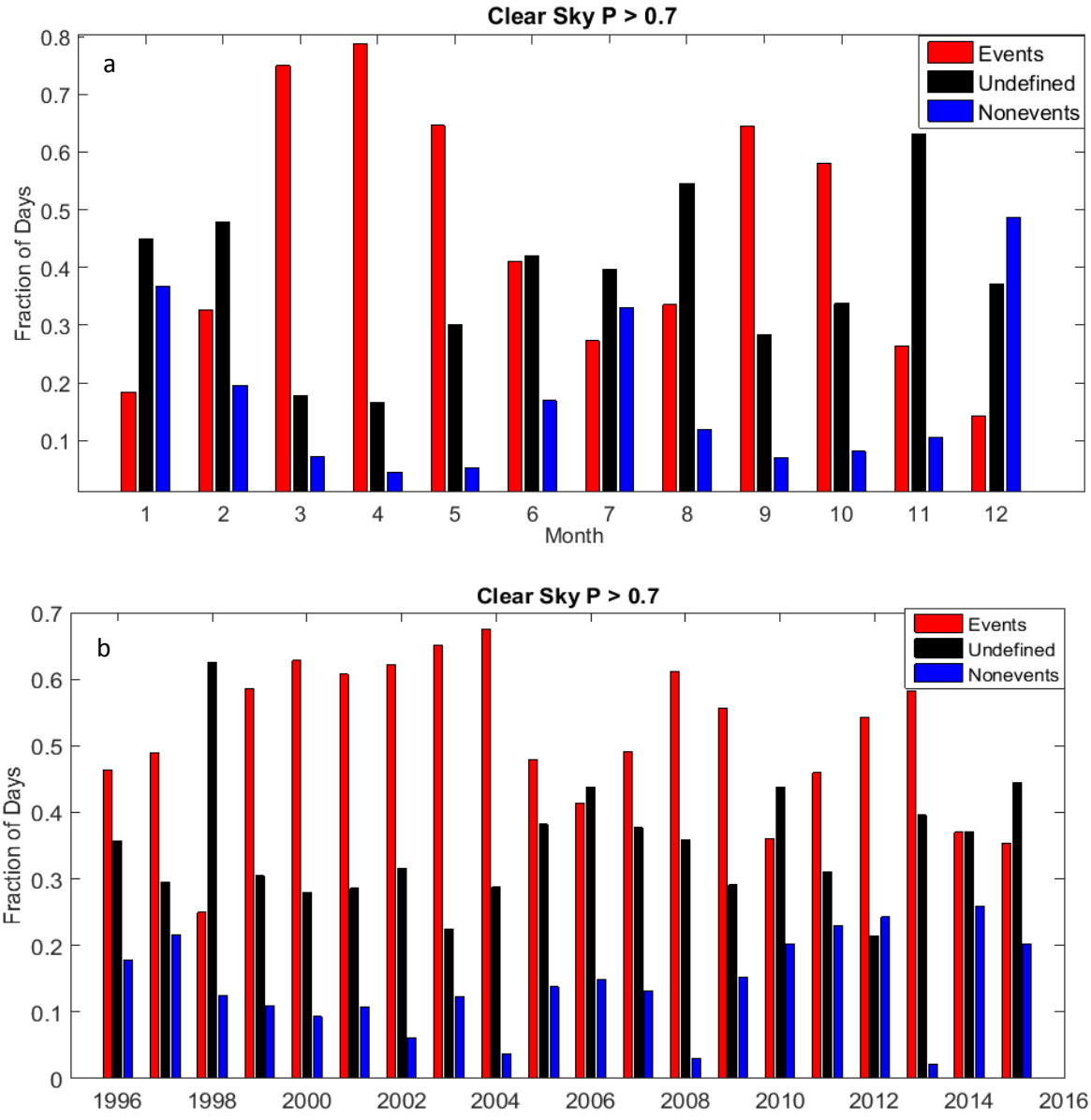
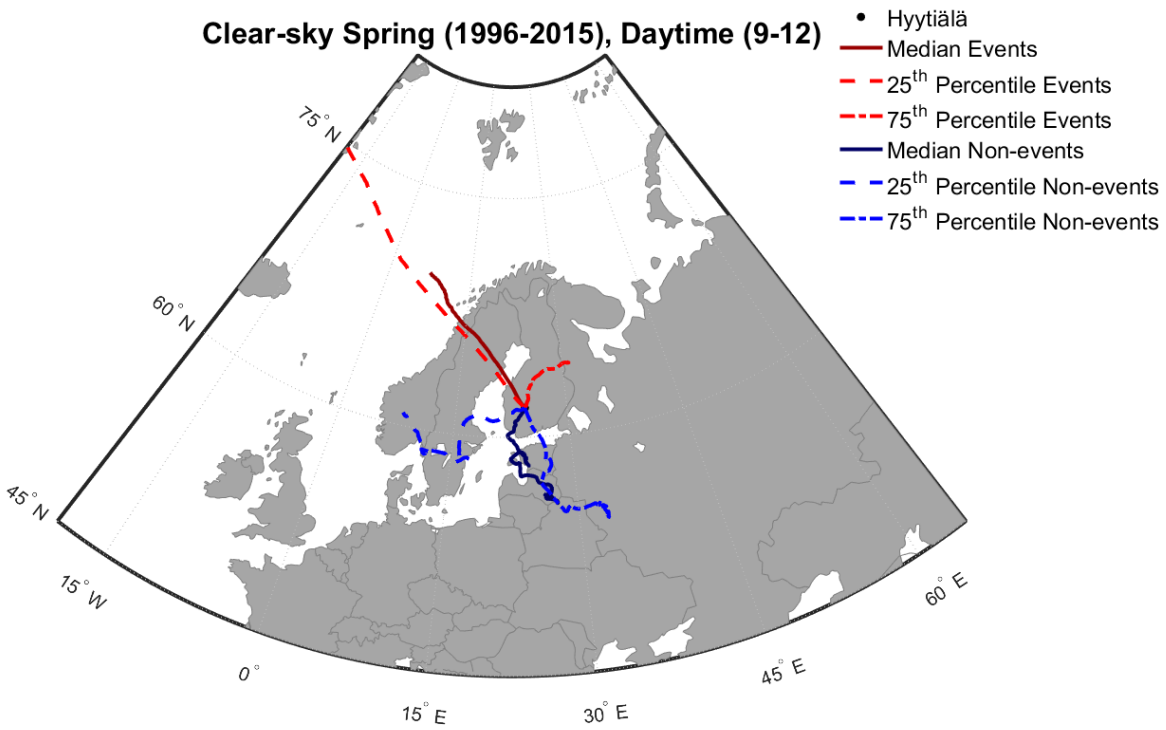


Figure 3: (a) Monthly and (b) yearly fraction of clear-sky days classified as NPF Events, undefined and non-events. In year 1998, global radiation data is limited to 5.4%, making the classification bias.

806

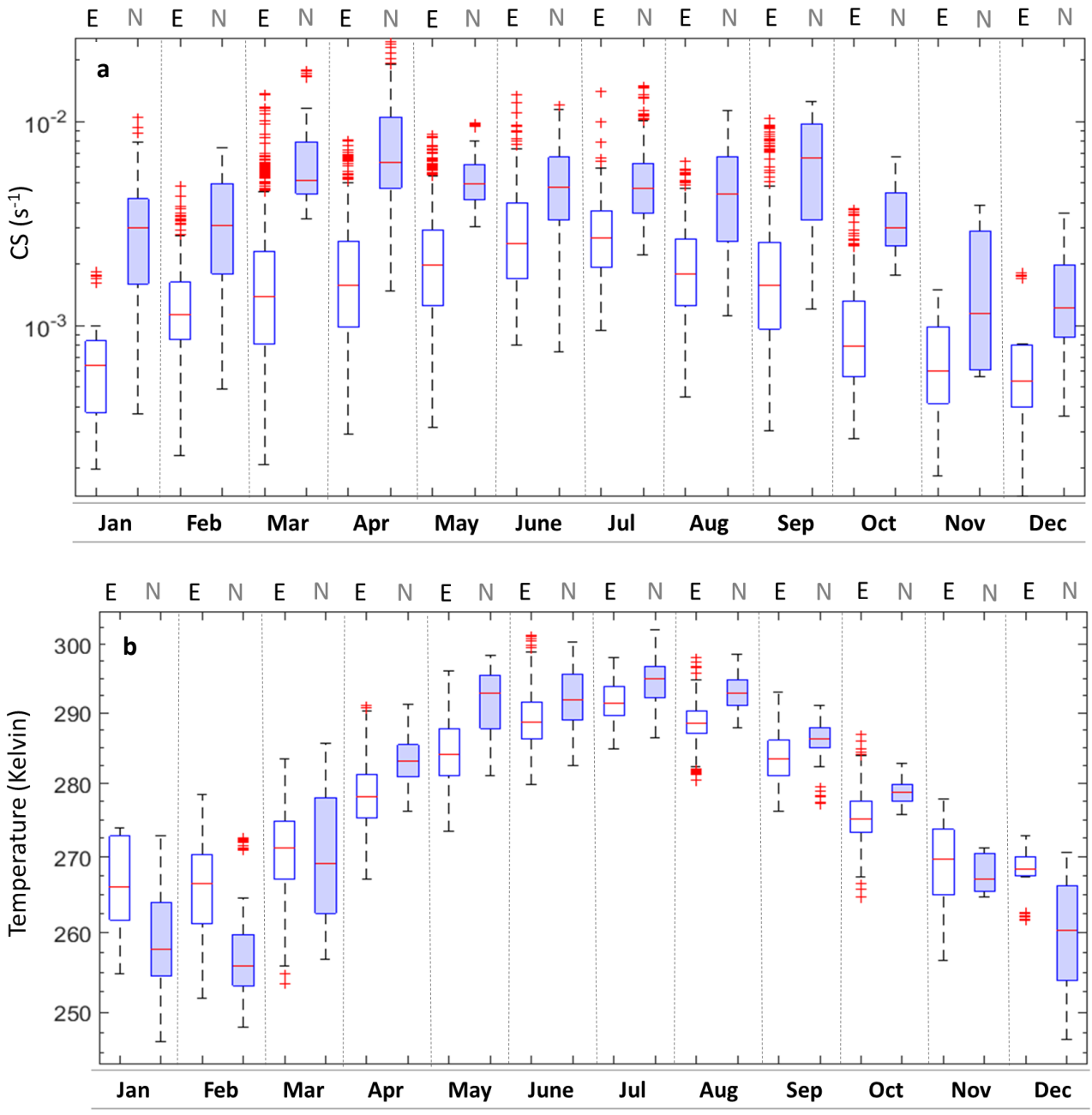


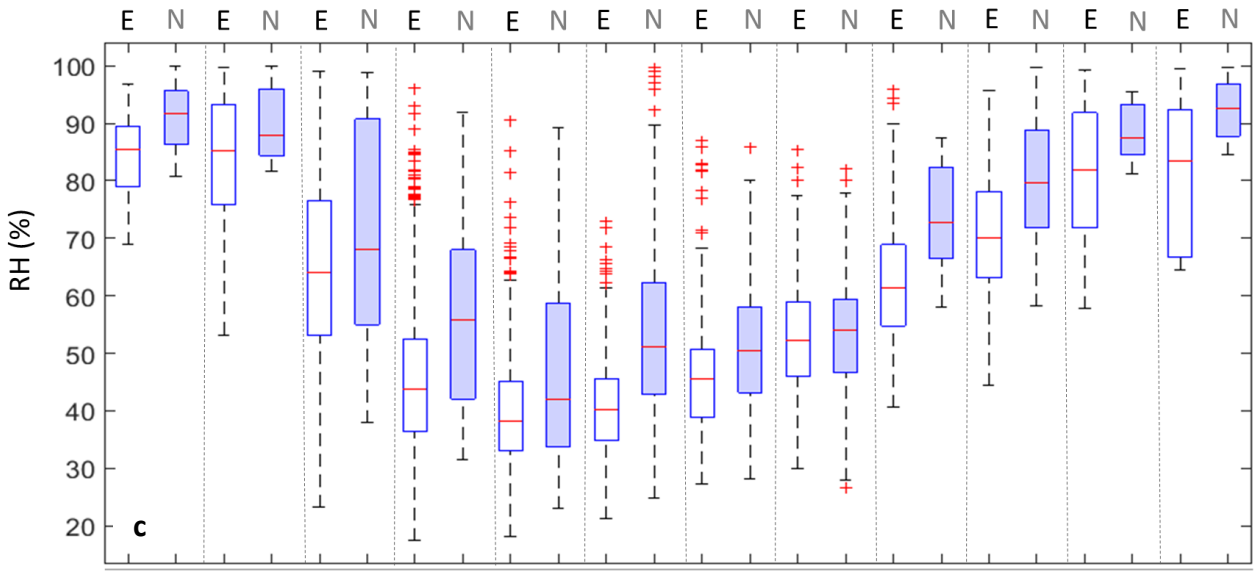
807

808 **Figure 4: Median and percentiles of 96 hours backward air-mass trajectories arriving to Hyytiälä during spring time (9:00-**
809 **12:00).**

810

811


a
b
E N E N E N E N E N E N E N E N E N E N E N E N
CS (s-1)
10-2
10-3
Temperature (Kelvin)
300
290
280
270
260
250
Jan Feb Mar Apr May June Jul Aug Sep Oct Nov Dec



	Jan	Feb	Mar	Apr	May	June	Jul	Aug	Sep	Oct	Nov	Dec
E	9	30	165	175	146	80	57	59	82	43	10	5
N	18	18	16	10	12	33	69	21	9	6	4	17

Figure 5: Median and percentiles of monthly variation (9:00 – 12:00) at P>0.7 of (a) CS (b) Temperature and (c) RH during NPF events (E, white) and non-events (N, shaded). See Figure 1 for explanation of symbols.

812

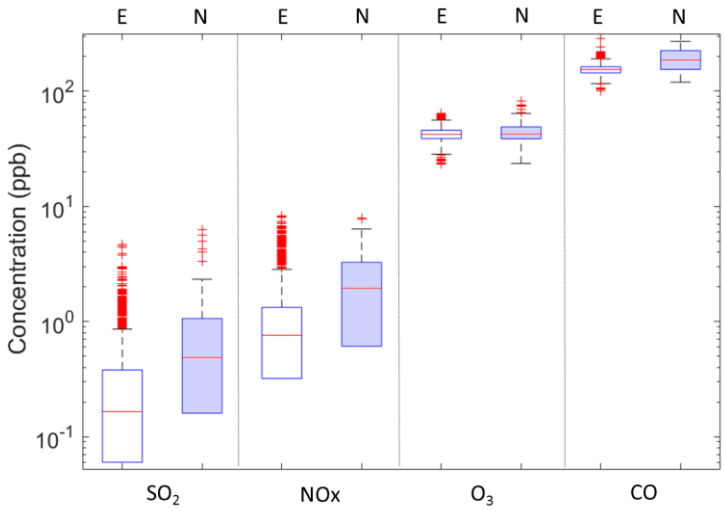


Figure 6: Spring time (months 3,4,5) medians and percentiles of trace gases during clear-sky events (E, white) and non-events (n, shaded) during daytime (9:00 – 12:00). See Figure 1 for explanation of symbols.

813

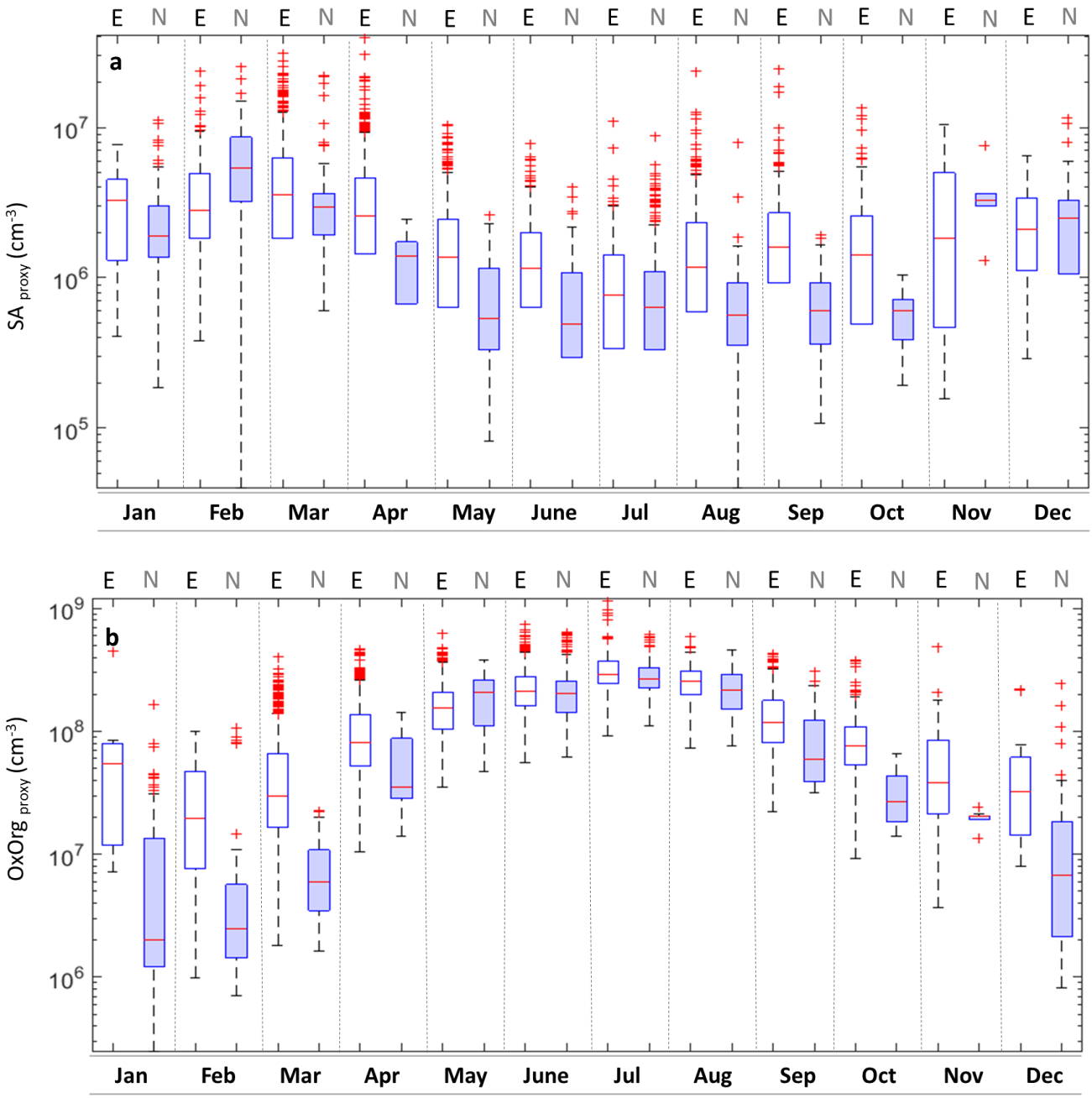


Figure 7: Monthly variation of medians and percentiles of (a) SA proxy and (b) OxOrg proxy at P>0.7 during the time window 9:00 – 12:00 of NPF events (E, white) and non-events (N, shaded). See Figure 1 for explanation of symbols.

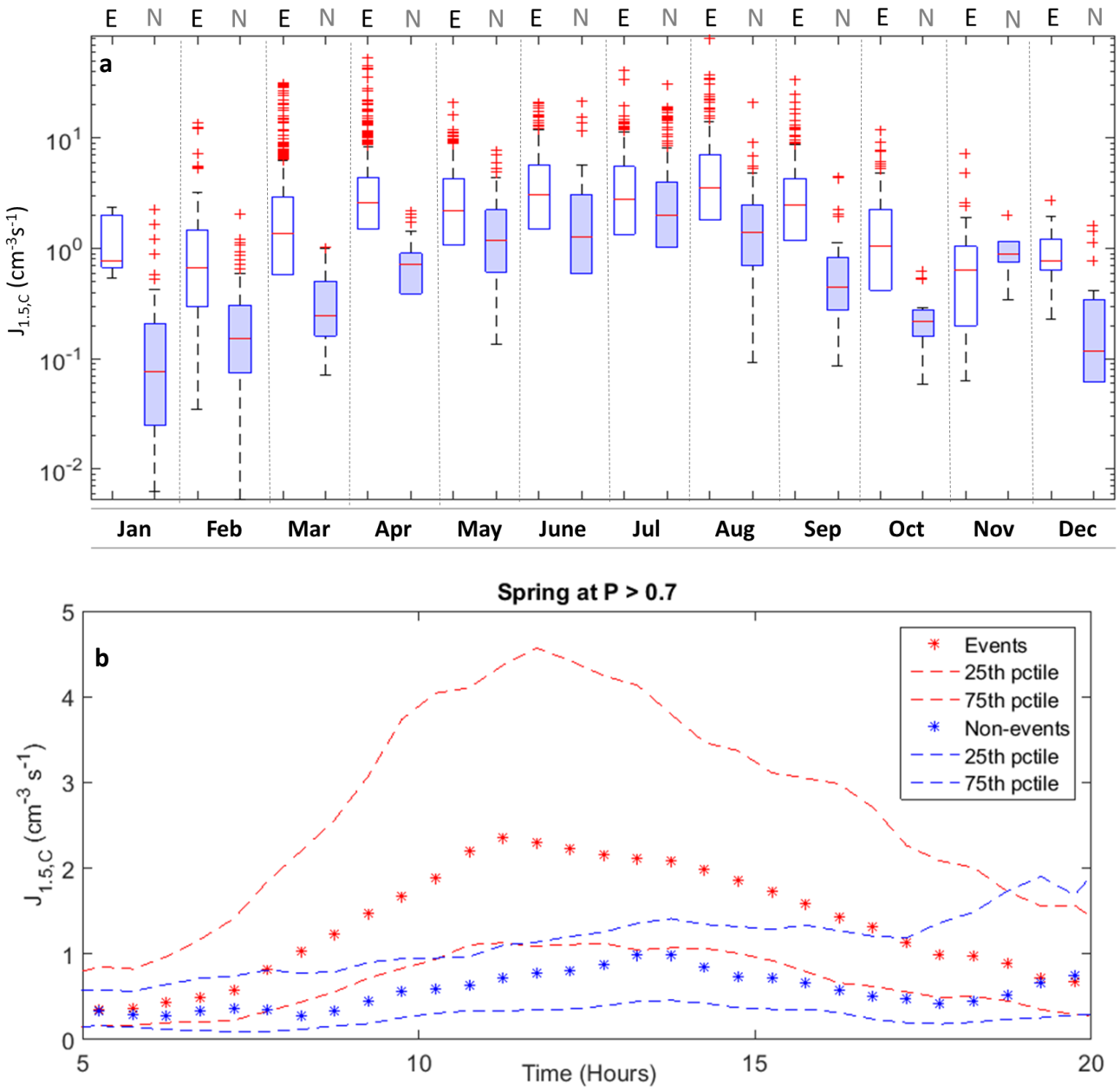
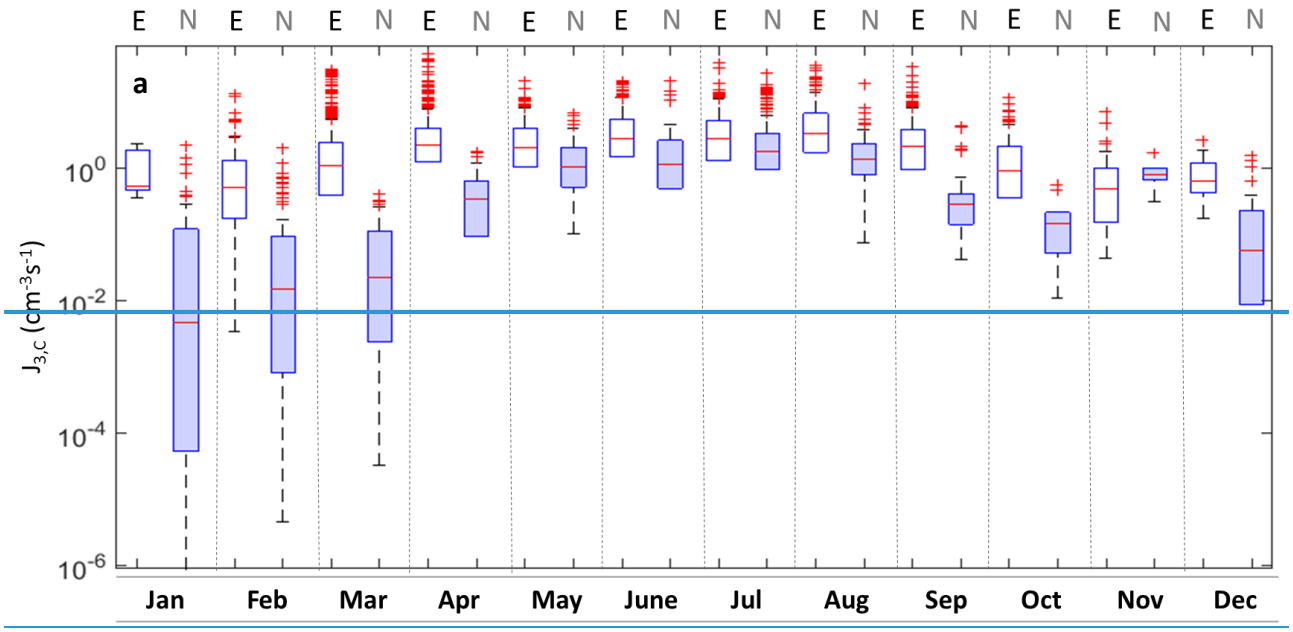


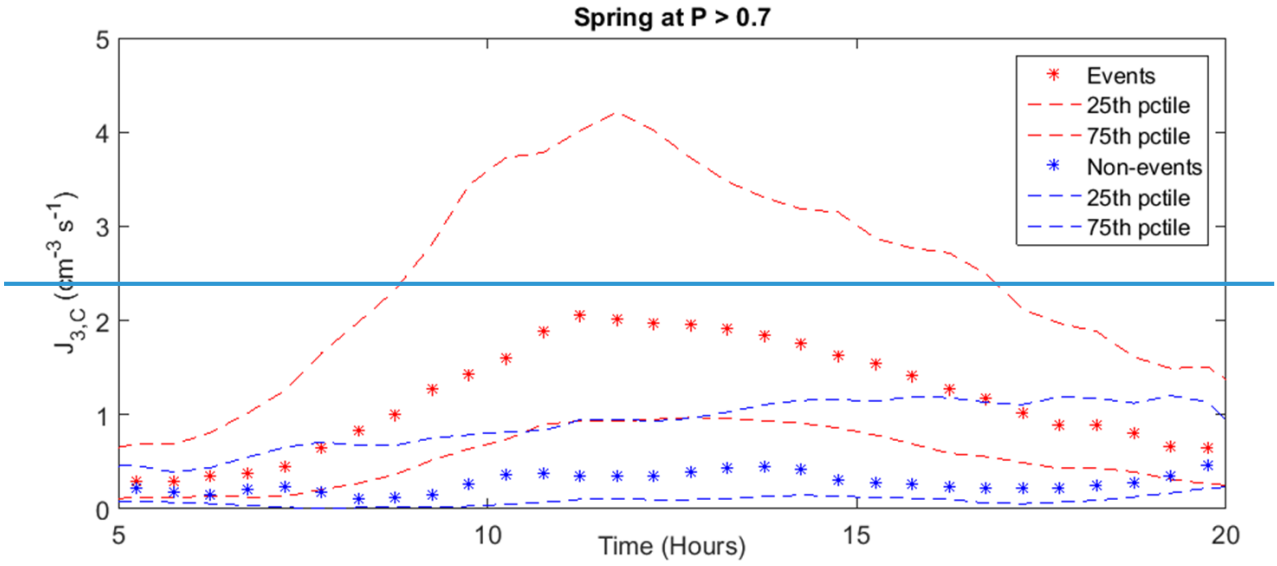
Figure 8: (a) Monthly variation of medians and percentiles of $J_{1.5,C}$ during the time window 9:00 – 12:00 of NPF events (E, white) and non-events (N, shaded). See Figure 1 for explanation of symbols. (b) The diurnal cycle of $J_{1.5,C}$ during Spring. The nighttime is missing in this plot due to unavailable SA proxy which uses UVB to be calculated.

816

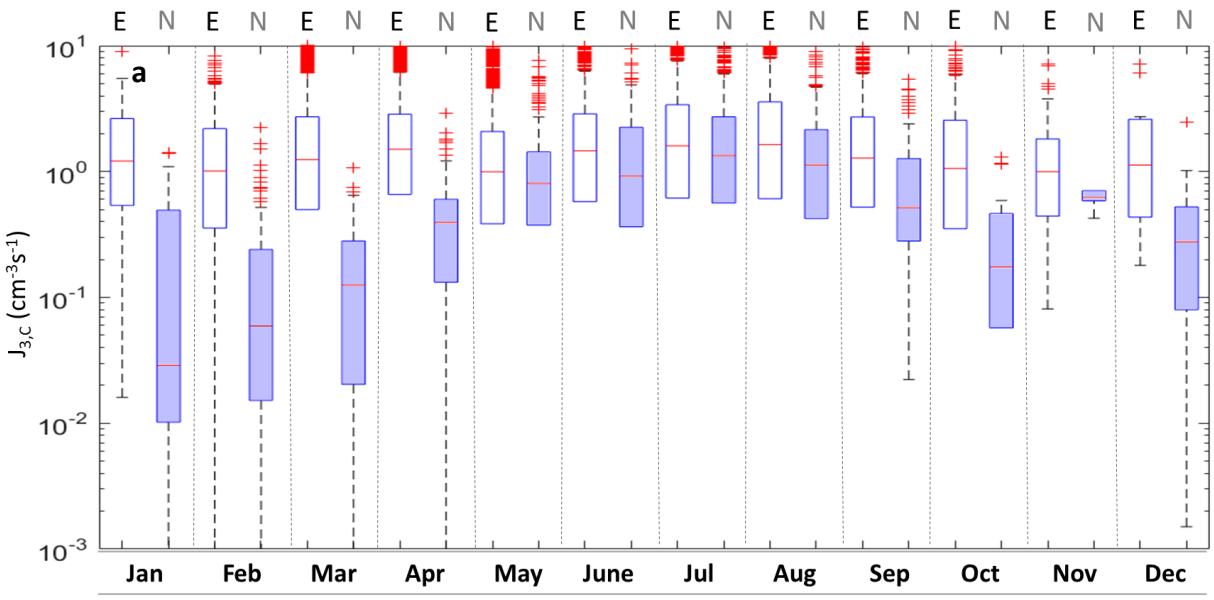

E N E N E N E N E N E N E N E N E N E N E N E N
a
10^{0}
10^{-2}
10^{-4}
10^{-6}
$J_{3,C}$ ($cm^{-3}s^{-1}$)
Jan Feb Mar Apr May June Jul Aug Sep Oct Nov Dec

817

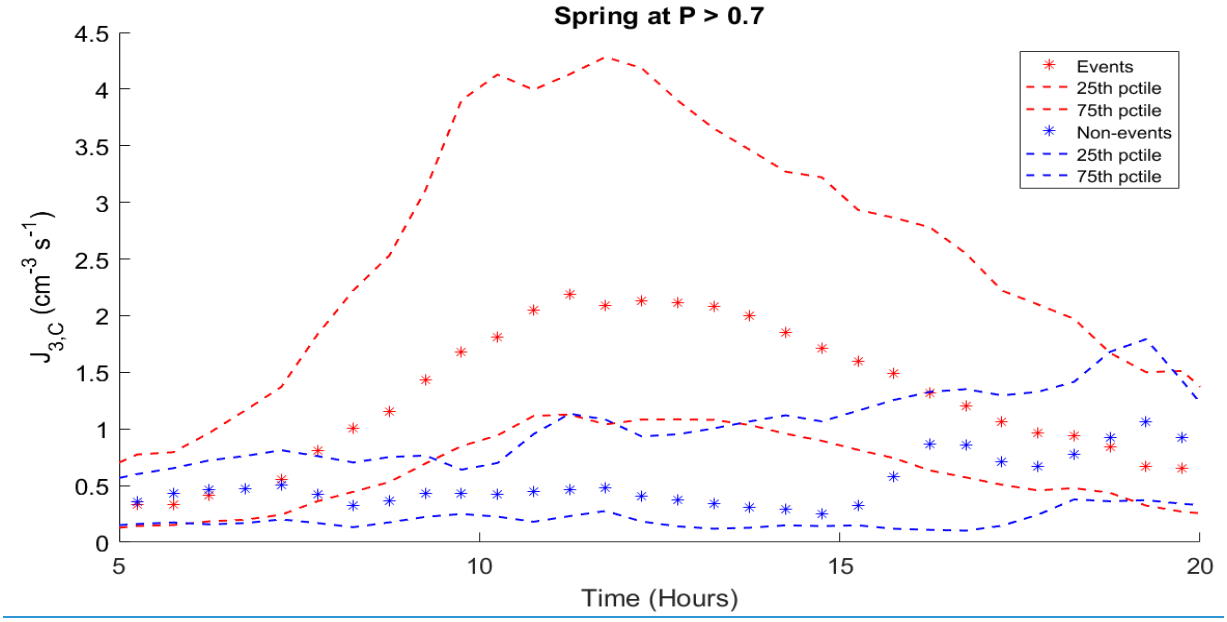
818


Spring at P > 0.7
Events
25th pctile
75th pctile
Non-events
25th pctile
75th pctile
$J_{3,C}$ (cm^{-3} s^{-1})
Time (Hours)
0 1 2 3 4 5
5 10 15 20

819



820



821

822

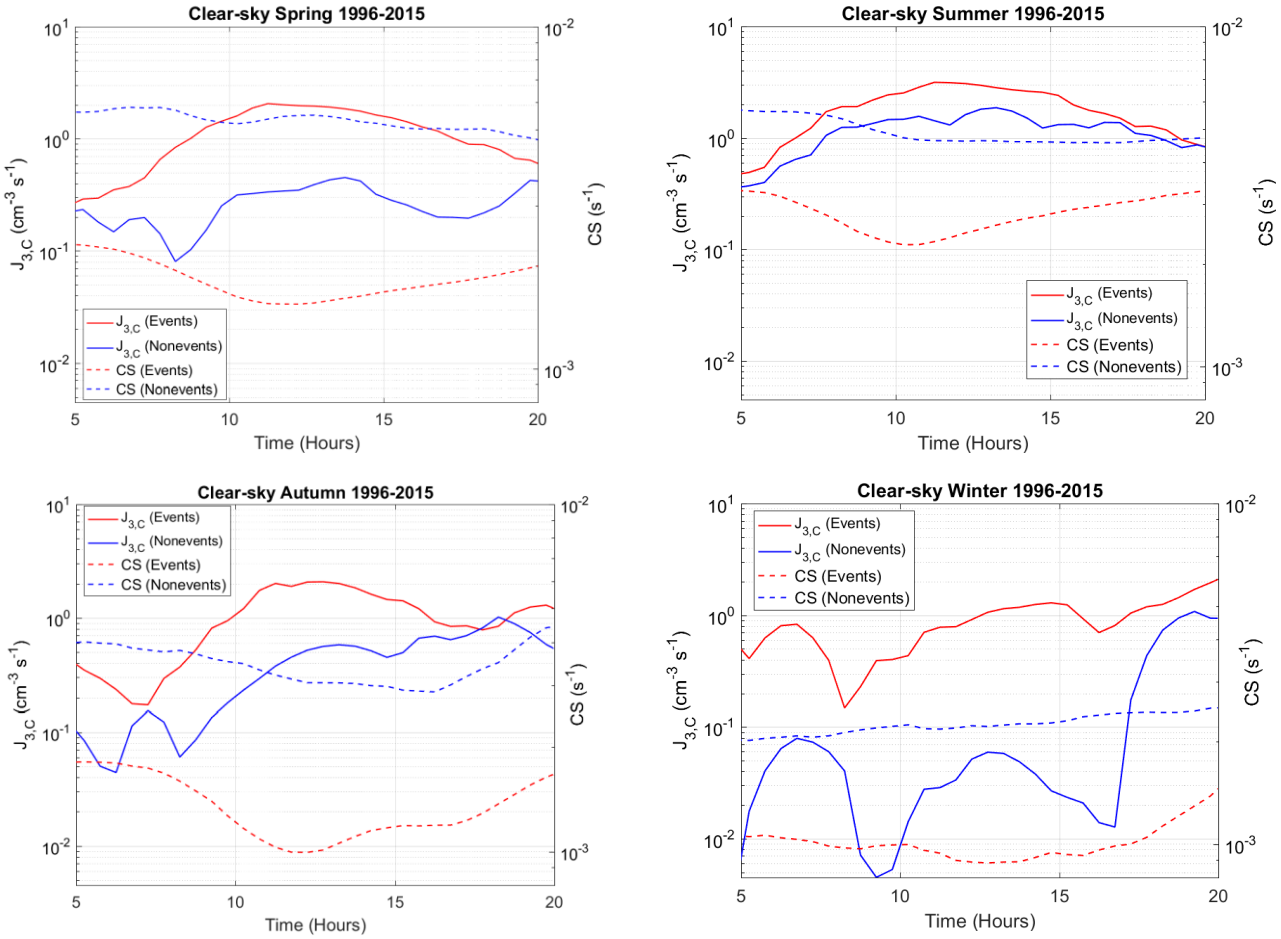
823

b

824 **Figure 9: (a) $J_{3,C}$ medians and percentiles during different months separated classified NPF events (E, white) and non-events**
825 **(N, shaded) (9:00 -12:00).** ~~The red line represents the median of the data and the lower and upper edges of the box represent~~
826 ~~25th and 75th percentiles of the data respectively. The lines extending 1.5 times from the central box represent the remaining~~
827 ~~of the data yet still within the relevant statistical limit. The outliers are represented by the red crosses~~See Figure 1 for
828 explanation of symbols. **(b) The diurnal cycle of $J_{3,C}$ during spring. The nighttime is missing in this plot due to unavailable SA**
829 **proxy which uses UVB to be calculated.**

830

831



832

833 **Figure 10: ~~Relationship between seasonal~~Diurnal cycle of median values of calculated formation rate of 3 nm particles ($J_{3,C}$)**
834 **and condensation sink (CS) during different ~~times of the day~~seasons on clear-sky events and non-events.**

835

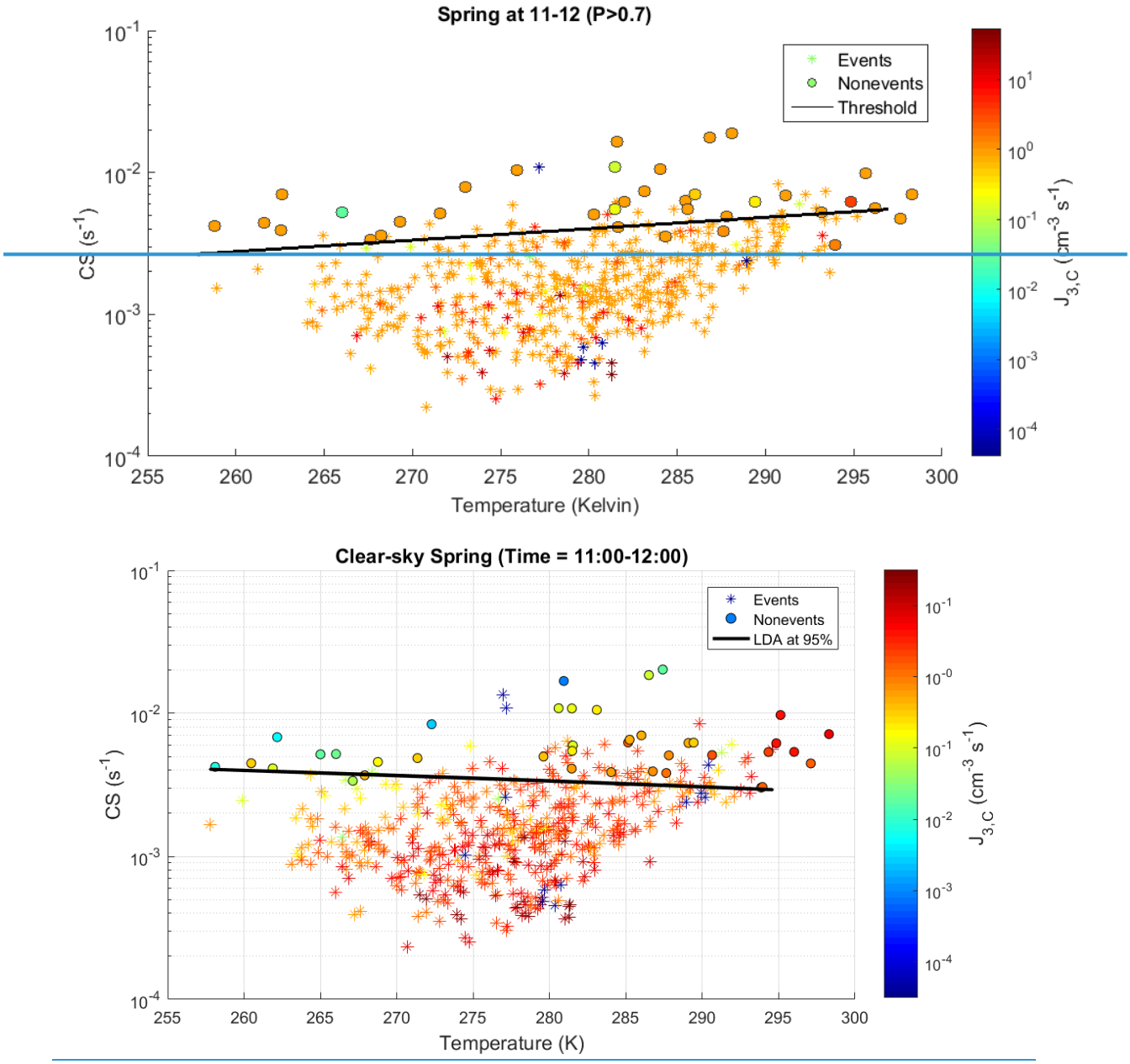
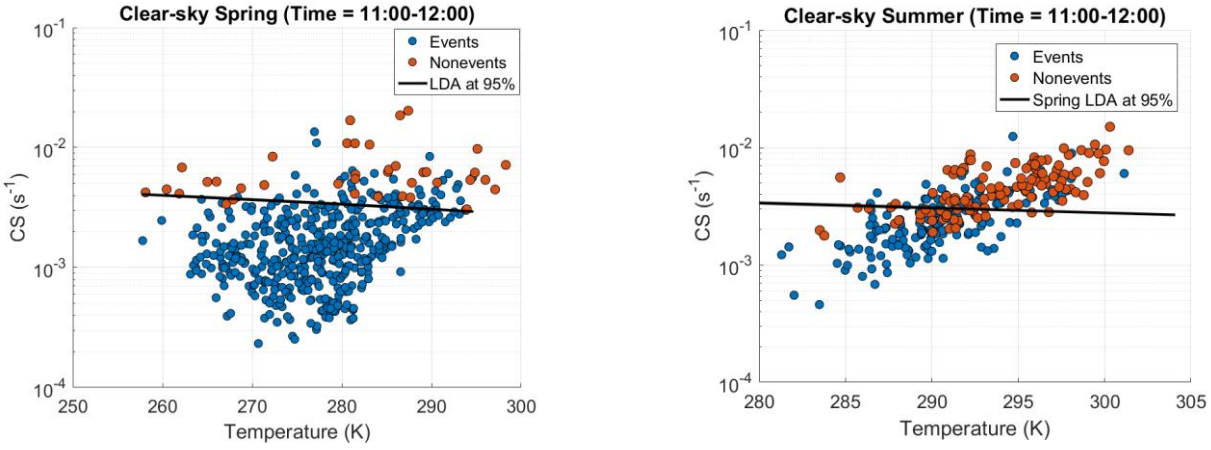


Figure 11: Relationship between temperature and CS during spring time (11:00 – 12:00) NPF clear-sky (P > 0.7) event days and non-event days color-coded with $J_{3,C}$. Horizontal line is calculated from LDA at 95% confidence relative to nonevents and is demonstrated by equation (6).



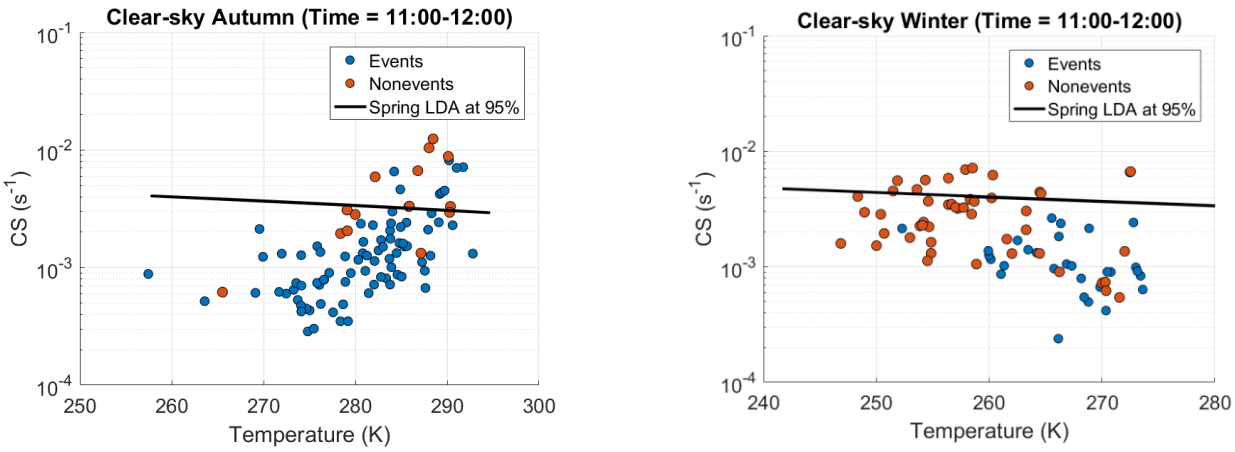


Figure 12: Relationship between CS and Temperature (time window: 11:00 – 12:00) NPF clear-sky event days and non-event days. Horizontal line is calculated from spring LDA at 95% confidence relative to non-events and is demonstrated by equation (6).

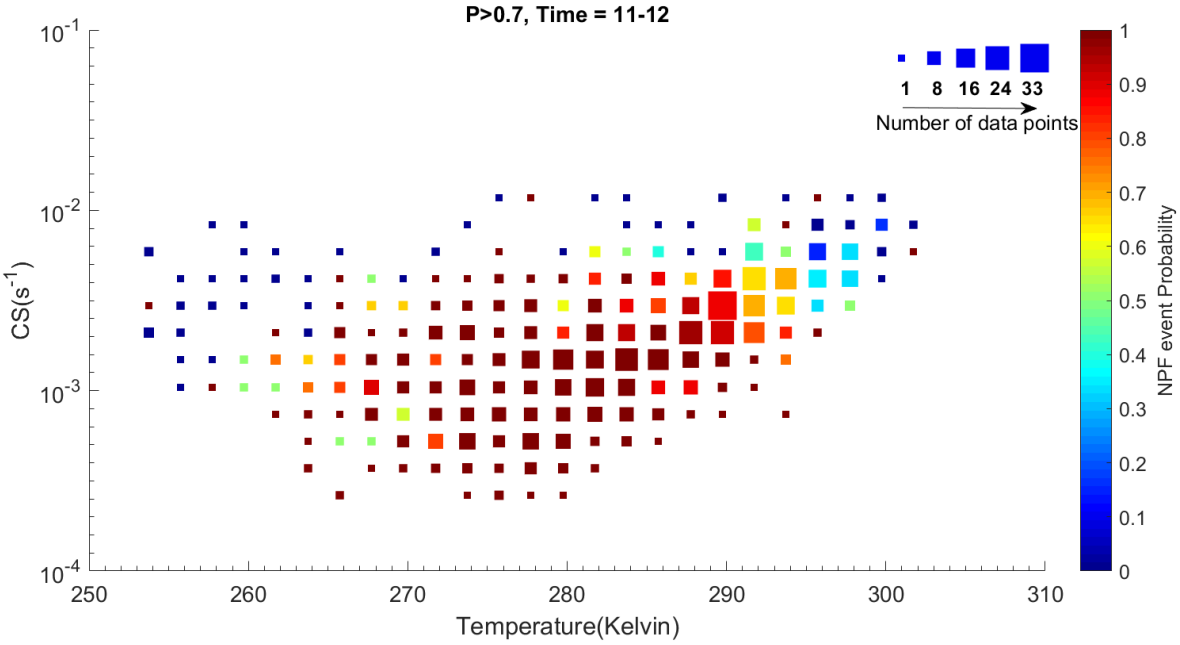


Figure 13: NPF probability distribution based on the CS and temperature conditions during clear-sky days (11:00 -12:00). Marker size indicates number of days included in the probability calculation within every cell.