

## ***Interactive comment on “Long-term analysis of clear-sky new particle formation events and non-events in Hyytiälä” by Lubna Dada et al.***

**Anonymous Referee #2**

Received and published: 6 December 2016

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).

C1

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.

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 probability 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. 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.

Detailed notes:

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:

- \* Were only events that started during this three-hour window included in the analysis?
- \* 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? \* what is the basis of using the value

C2

0.7?

The reasoning between this central point 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?

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.

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...')

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 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.

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

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

lines 331-350: I don't really understand what is shown in figure 10, and therefore also  
C3

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.

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 Ålowest possible CS which still fit the linear separation"; was some kind of data selection applied to produce the figure?â€

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.

---

Interactive comment on Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-859, 2016.