Atmospheric
Chemistry
and Physics
Discussions

Open Access

Interactive comment on "The Effect of Meteorological Conditions and Atmospheric Composition in the Occurrence and Development of New Particle Formation (NPF) Events in Europe" *by* Dimitrios Bousiotis et al.

Anonymous Referee #2

Received and published: 26 September 2020

General comments

The focus of this study is to investigate the effect of meteorological conditions and atmospheric composition on the occurrence of new particle formation (NPF) events at 16 sites (rural, urban background and roadside) located in 6 European countries. The results are based on more than 85 years of meteorological and atmospheric composition data. The authors are using a binned linear regression to find correlations between parameters such as windspeed, temperature, pressure, or solar radiation intensity, ozone or volatile organic compounds mixing ratios to name a few, and the occurrence of NPF,

particle growth and the formation rates. This is an interesting study and of interest to the community, however the following comments should be addressed before publishing.

On many occasions the authors claim that certain variables are weakly or strongly correlated but do not provide any numbers or figure to support these statements. Please provide references to the exact figures in the manuscript or in the supplemental material (SM). This is an issue reappearing throughout the manuscript. Following on that, Figure S1 in SM contains many figures and only one caption. These figures are not marked with a number/letter. Please consider adding individual numbering (or introduce letters) so the respective figures corresponding to individual sites when being discussed in the manuscript could be easily found in SM.

It looks like the authors use terms frequency of NPF occurrence and NPF probability interchangeably. NPF probability doesn't really fit here since you do not predict NPF events. However, the NPF probability term is explained in the text and in the equation (line 191; also please number equations). In results, however, the authors are using term frequency of NPF occurrence (line 245). Please clarify, review the explanation in text and use the correct term throughout the manuscript. I assume what you want to use is NPF frequency.

I understand you identify the number of days with NPF according to the method by Dal Maso et al. (2005) with additional certain criteria. It would be good to report the numbers of NPF events for each site (and season?). Please explain what days with "relevant data" are when calculating the frequency.

If I understood correctly to calculate the frequency, you divide the number of days identified as NPF event-days by all days that you have data available or "relevant(?)" data available? I am curious how does the frequency changes when when you use the number of all days with all data and not only with "relevant data available"? It would be good to mention this number somewhere in the manuscript or in the SM? Following on

the above, please explain what is in e.g. line 191 "available data" and "given group"?

More detail on the site selection criteria would be helpful. Do these sites belong to a network? How were these sites selected? In Line 120: the authors mention "geographical region and type of environment". I suggest adding more description on the sites (e.g. in SM), their characteristic and typical meteorological conditions, e.g. features they have in common/differences, number of NPF studied and identified.

Having this extensive dataset, I encourage the authors to discuss variability e.g. seasonal, site to site/regional. A figure where you plot frequency of NPF occurrence or number of NPF events for each station in each season (e.g. bar plot?) on (y axis), for each site type (x axis) would be helpful. Exploring e.g. seasonal variability would add value to the paper.

Where there any limitations of the study? If yes, these should be discussed. Further, errors should be included.

Are there general trends for these three site types? Maybe you could discuss these more. It would be helpful to highlight (e.g. text in bold) data in Table S1 e.g. significant correlations.

What is the importance of the result of this study? The authors could discuss it more e.g. in conclusion. I feel that is missing in the current version.

Please improve the language. It is critical to make the text more concise and clearer. It is hard to follow the line of thoughts at points. There are some repetitions and long sentences that could be shortened (e.g. lines: 76-82, 107-111, 325-330).

Specific comments

Line 38-62: in the abstract the authors could also mention: 85 years of data; how good these correlations are (r2)' mention "meteorological conditions" e.g. such as . . .

Line 42: "except at very clean air sites" – more information is needed to this statement.

Something is missing. Please review or explain.

Line 54: What "higher values" means there exactly? Provide a number.

Line 61-62: you could give these values in brackets

Line 97: "negative effect" on?

Line 99: "average conditions"? What does it mean here?

Line 107-111: hard to follow, please review and shorten

Line 121: please add references to the studies you refer to

Line 122: NPF probability? Or NPF occurrence? As mentioned above, probability doesn't really fit here

Line 124: I suggest calling this section: "2. Methods", 2.1 as is. 2.2 as is or similar. This way you can remove 2.2 Methods so it does not appear twice. In 2.1 the authors could mention which cities/countries/sites were used; which meteorological and atmospheric composition variables did you use in this study already at this point. Which stations had a full set of data and which only some etc. Maybe also mention which are dependent and which independent variables. And what do you consider relevant data days, what do you mean by available data: e.g. in line 189. Please be more specific upfront. You could also add information on how these sites were chosen? Any criteria you applied to select these? Are they belong to a network? Are they similar or different in any respect?

Line 127: I feel that the number of events (1950) is already a result so it should go to the result section and not methods. Also, it is mentioned before the description of the NPF selection method itself.

Line 131: it is also referring to the result. I suggest moving this sentence to the result section.

Line 136-143: please add more details to the approach taken in this study. What "Ia" exactly refers to and which additional criteria was used (line 142).

Line 137: add size range of the nucleation mode you consider in your study

Line 139: you could mention confidence level in the brackets

Line 151: add "respectively" after "particles". You could already mention there Fuchs correction factor and keep it explained below.

Line 149: Formulas need to be numbered

Line 188: given group? please explain

Line 191: Again: I am not sure I follow this equation: what are these groups? Is it just a number of days with NPF that was accompany by all relevant data? From explanation you seem only to take days with NPF that were accompanied by relevant data.

Line 196: low significance? Please give a number

Line 212: extreme values? Please give a number

Line 239: Results and Discussion? You could include here sentence in line 126: You mention 1950 events studied, could you provide information on how many were identified? It would be helpful if authors mention that in the paper a summary of data can be found in the manuscript and in the SM data/results for individual sites is presented.

Line 245: what is relevant data? Please explain more clearly in methods section and refer to it. Diagnostic features – wouldn't these be better in methods?

Line 252: "slopes and R2" please use correct terms for these or more careful description

Line 261: very strong? Please provide references to the exact figures in the figures/supplemental material when discussing results

Line 279: low? Please give a number and refer to the figure. Also, you placed all

figures in the SM under Figure 1. Maybe it would be better for the reader to have them split into different figure numbers or a,b,c,d? This way it would be easier to find the one you describe at the very moment.

Line 296: reference?

Line 301-303: why? Could you explain? When describing results maybe worth mentioning these for various site specifics? Anything in common?

Line 369: which factors remain constant?

Line 377: reference?

Line: 398: maximum? Low?

Line 420: Ethesian: add few words what these are could be added

Table 3: what is a "p value"? has it been defined somewhere? In tables: the authors could use bold text to highlight significant correlations? So these patterns/trends could be clearly seen?

Figures: no need to mention in each caption "of the present studies"

Line 433: you could already mention here which pollutants (such as...) are studied and described in the upcoming sections.

Line 752: "at higher values"?

Line 755: "meteorological conditions" such as?

Line 756-757: is that the only explanation? How about chemistry/composition at such type of site? Anything else that might play a role?

Line 782-783: seems out of place here; it would be more suitable at the beginning of conclusion section or removed.

2020.