

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.

Dimitrios Bousiotis et al.

r.m.harrison@bham.ac.uk

Received and published: 15 October 2020

MS No.: acp-2020-555 Title: The Effect of Meteorological Conditions and Atmospheric Composition in the Occurrence and Development of New Particle Formation (NPF) Events in Europe Author(s): Dimitrios Bousiotis et al.

RESPONSE TO REVIEWERS

The authors thank the reviewers for their insightful comments and have made many modifications in response, and to enhance the clarity of the paper.

C1

Anonymous Referee #1 Received and published: 31 August 2020 A couple of decades ago, a number of studies tried to link meteorological variables and gas phase pollutants with NPF. In some cases, the analysis was concise enough to produce evidence that a certain physical parameter played a role in NPF at a specific site. Since then, studies - mostly chamber-based - have provided evidence on the ruling mechanisms of nucleation and subsequent growth of newly formed particles. These are mainly related to the concentration of low vapor pressure compounds such as sulfuric acid, or ELVOC as well as agents that could stabilize the former (ammonia, amines, iodine) suggesting that NPF is dictated mainly by gas phase chemistry rather than meteorology. However, other parameters such as the ones investigated in this study, play a secondary yet important role. Therefore a summary of observations from European, or even better global sites, is always welcome. During the past 15 years more than 20 compilations of results related to atmospheric NPF have been published, the majority of which are summarized by Kerminen et al., 2018. Even though some of them (eg Kerminen et al., 2018, Lee et al 2019) provide insight on the parameters this study is also focusing on, none has gone been as detailed as the one presented in this work. Therefore, the compilation of results presented in this work are of interest to the community and would be worthwhile publishing if the manuscript was well written and the analysis provided informative and concise. I am afraid that this is not the case. After reading the article, I was disappointed not to find any information on seasonality for any of the parameters investigated even though multi year data were investigated. RESPONSE: It is not clear by the comment what kind of analysis is expected (whether it is the seasonality of the parameters themselves or the seasonality of their effect). The seasonality of the parameters (which was found to favour mainly summer for the growth rate, while the results for the formation rate were more variable) is separately investigated in a previously published paper for the UK sites (Bousiotis et al., 2019) and for the rest of the sites in an already submitted manuscript (Bousiotis et al., 2020) and thus was not discussed in the present study (this is noted in 2.1). The seasonality of their effect was not studied in the present manuscript as this would extend its size too much. It is

C2

worth pointing out that the variables most affected by season such as temperature and insolation are considered in this paper, and to break down the data analysis according to season would involve a great deal of repetition.

Furthermore the authors fail to deliver any error metric whatsoever (deviation, error, confidence level). The lack of the most elementary statistical analysis was striking. RESPONSE: Much of this information is included in the paper. Deviation errors are included in the SI figures for every subgroup of every variable studied (reporting these is unrealistic as they are over a thousand). R2 is reported for every slope calculated of every variable studied on the figures. p-values are reported (when significant) for every variable in every site in tables 3 and 4. We have calculated, but not included, the error of the slope for every variable calculated using the normalised gradients, but have not included this as the normalised slopes do not have any significance other than their absolute value; we include only information on their trend.

The other striking feature is the poor use of English and terminology, which I explain thoroughly below. The use of English must be improved as there are many sentences that require revision. RESPONSE: Many changes in terminology and corrections were applied throughout the manuscript to improve the level of English.

The major drawback is the generalizations and uncertain phrases used throughout the manuscript. The authors should be concise and specific instead. RESPONSE: The manuscript was updated in many cases to reduce uncertainty (whenever there was enough confidence in the statements presented)

As an example in Line 69 (76) it is advised to name the places (exceptions) were NPF is hardly observed. A nice review can be found by Lee et al., 2019 (section 4.8). RESPONSE: Exceptions where NPF events are not observed and references were added.

Example 2 Line 270: A few sites presented a strong correlation, which in all cases were background sites (either rural or urban). A few sites (which ones?) presented a strong

C3

correlation (nowhere in the manuscript strong, medium weak is defined. The reader has no idea what the author is discussing) which in all cases were background sites (either rural or urban; to the best of my knowledge rural sites are considered as background sites. What do the authors mean?). I assume that the authors are trying to point at urban kerbside sites with this sentence, yet I am not really sure what they mean. And the paragraph continues The relation (which one?) found in most cases (how many, percentage?) was positive (does this mean a positive slope? Where is it shown? In which table or graph?) apart from two roadsides (improper terminology) and GREUB, though due to the low (again low is not defined?) R2 these results cannot be used with confidence (and where do the authors draw the confidence line?). The above lines are just an example of improper phrasing used throughout the manuscript that make it very hard to follow. Similar examples can be found throughout the manuscript. A major drawback of this work is that many trends/relationships reported are not referred to any table or Figure and hence are hard to follow. RESPONSE: References to the sites mentioned in each case as well as R2 values were added throughout the manuscript to improve readability. References for the results were added in the beginning of each section (to avoid repetition). Specific references for unusual trends were also added for the figures in the SI (figure numbering in SI was overhauled). References for SI figures for simple relationships were not added as they are covered by the slopes found in tables 3 and 4. Strong, weak and other characterisations of the correlations are now accompanied with either the R2 or a range of the R2.

NPF probability sounds to me as if you are trying to predict the occurrence of nucleation events. Based on Line 218 (Equations are not numbered!) a more suited term would be NPF frequency. RESPONSE: The term NPF frequency is used within the text for the frequency of the events without taking into account any grouping of the data (into groups of condition ranges e.g NPF probability for RH in the range 60 – 65%). To separate these the term probability was used instead. Equation numbers were added.

The authors should consider adding reference formation and growth rates from other

C4

studies in their figures for comparison. I understand that this is not always possible (especially for formation rates) but is for the other two parameters in question. RESPONSE: Similar to a previous comment, such an analysis was done in other studies either already submitted or published (Bousiotis et al., 2020; 2019).

Table 2 should also include growth rates for this study. RESPONSE: The parameters of NPF are already reported in previously submitted studies (Bousiotis et al., 2020; 2019). The frequency and formation rate are reported here because they are used in the calculation of the normalised slopes, which is not done for the growth rate (see the methodology). Nevertheless, the growth rate and the number of NPF events for each site were added in Table 2.

The authors fail to summarize the seasonality of the parameters they are exploring even though they are having multi year data. This is very disappointing. RESPONSE: This comment has already been addressed.

The statement in Lines 45-46 (49 – 50) is not true. Please read Kerminen et al., 2018 for example. That work which explicitly states the opposite. RESPONSE: The sentence was rephrased into “without always following” to state that exceptions exist as pointed later in the Introduction part.

I have noted another case (Lines 98-99) (109 – 110) in the manuscript where the authors focus on the exceptions (which always exist) rather than the rule giving a very distorted view to the reader. RESPONSE: In the text it is stated that “the negative effect of CS is widely accepted” and follows mentioning the exceptions found in the literature as “cases were found”. This does not imply that the exceptions are anything more than that and it is essential that they are mentioned.

The introduction is very poor on references. RESPONSE: More references were added in the introduction and throughout the manuscript

Lines 107-111 (118 – 123) should be rephrased. I cannot make sense of it at all.

C5

RESPONSE: This sentence was included to show that the NPF events considered in our study were not driven by combustion products but by secondary formation. A parenthesis was added though which makes the sentence easier to understand.

Lines 82-84 (94 – 96). Please mention that increasing temperatures also have a negative effect as they increase the energy barrier the clusters have to overcome to become stable and grow in size. RESPONSE: The comment has been added.

Kerminen, V. M., Chen, X., Vakkari, V., Petäjä, T., Kulmala, M. and Bianchi, F.: Atmospheric new particle formation and growth: Review of field observations, *Environ. Res. Lett.*, 13(10), doi:10.1088/1748-9326, 2018.

Lee, S. H., Gordon, H., Yu, H., Lehtipalo, K., Haley, R., Li, Y. and Zhang, R.: New Particle Formation in the Atmosphere: From Molecular Clusters to Global Climate, *J. Geophys. Res. Atmos.*, 124(13), 7098–7146, doi:10.1029/2018JD029356, 2019.

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

C6

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. RESPONSE: This is addressed in a later comment

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. RESPONSE: This is explained later

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. RESPONSE: Two additional papers analysing in detail the conditions of the NPF events at all the sites were either published (for UK) or were submitted (for the rest of the sites). This is noted in section 2.1. A figure has been added to the SI to show the seasonality.

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 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"? RESPONSE: The term "relevant data" refers to all the data available at each site and are considered in each analysis (and of course when those are available). At each site the data were almost in their entirety available and the limitation was in most cases the SMPS data (its availability for each site is reported in Table 1). The data is always

C7

considered only when they are available for each variable studied (by the code used in the analysis, as they were calculated one by one), so no hours of data with missing values were included. In other words, when e.g. temperature was studied only the hours with both SMPS and temperature data were considered.

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. RESPONSE: A justification for the sites chosen is given in the Site description section. As mentioned earlier, the analysis of the events, as well as the typical conditions for all the sites were given in two separate earlier papers.

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. RESPONSE: This was already done in two earlier papers (Bousiotis et al., 2020; 2019).

Where there any limitations of the study? If yes, these should be discussed. Further, errors should be included. RESPONSE: The study is pretty straightforward, and the only limitation was the lack of data for some variables at some sites e.g. SO₂ data was not available at all sites. A comment was added for this limitation at the end of 2.1 Site Description and Data Availability section.

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. RESPONSE: As explained in the response to the first comment, these are provided in other studies (Bousiotis et al., 2020; 2019). Stronger correlations were

C8

highlighted with bold numbers.

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. RESPONSE: The statement: "This study, apart from providing insights into the effect of a number of variables on the occurrence and development of NPF events in atmospheric conditions across Europe, also shows the differences that climatic, land use and atmospheric composition variations cause to those effects. Such variations are probably the cause of the differences found among previous studies." was added in the last paragraph of the text (838).

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). RESPONSE: Many changes were made in the manuscript to improve readability.

Specific comments Line 38-62: in the abstract the authors could also mention: 85 years of data; how good these correlations are (r^2)' mention "meteorological conditions" e.g. such as ... RESPONSE: The information that a combined dataset of 85 years was used was added (52). Added the highest R^2 values found for some variables (54 – 56). Added "(such as solar radiation and relative humidity)" (58)

Line 42 (46): "except at very clean air sites" – more information is needed to this statement. Something is missing. Please review or explain. RESPONSE: This phrase has been removed.

Line 54 (60): What "higher values" means there exactly? Provide a number. RESPONSE: Added the word "average". No values can be provided as what is implied by the results is that the importance of some variables becomes less as the average values (average conditions) become higher or lower, depending on the general trend.

Line 61- 62 (67-68): you could give these values in brackets RESPONSE: No val-

C9

ues were added as the text implies that one increases with the other simultaneously, similar to the meteorological conditions. Line 97 (108): "negative effect" on? "on the occurrence of the events" was added.

Line 99 (110): "average conditions"? What does it mean here? RESPONSE: No change in the text. It means the average CS which is well covered with the term "average conditions" in this case.

Line 107-111 (118 – 123): hard to follow, please review and shorten RESPONSE: Already mentioned by referee #1 and addressed.

Line 121 (133): please add references to the studies you refer to RESPONSE: References were added

Line 122 (134): NPF probability? Or NPF occurrence? As mentioned above, probability doesn't really fit here RESPONSE: NPF probability was not changed as every time it is mentioned it implies the results from the analysis/modelling that was done in this study. An explanation for this was provided (213)

Line 124 (138): 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? RESPONSE: Section naming was not changed as it is considered sensible for a chapter named "Data and Methods" to have section 1 named "Sites and data" and section 2 as "Methods". The countries and cities included in the study are mentioned. A list of the data available in each site is found in Table 1,

C10

as mentioned in the text. A justification for the sites chosen was also added. The sites do not belong to a single network and thus such information is not provided.

Line 127 (140): 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. RESPONSE: The number of events was moved to the beginning of the Results section

Line 131 (142): it is also referring to the result. I suggest moving this sentence to the result section. RESPONSE: The reference for the Table with results was moved to the beginning of the Results section.

Line 136-143 (156 – 165): please add more details to the approach taken in this study. What “Ia” exactly refers to and which additional criteria was used (line 142). RESPONSE: The process of NPF event extraction was rewritten and more details were added for the approach taken (156 – 169).

Line 137 (158): add size range of the nucleation mode you consider in your study RESPONSE: Added “(smaller than 20 nm in diameter)”.

Line 139 (160): you could mention confidence level in the brackets RESPONSE: Changed to level of certainty to avoid misunderstandings.

Line 151 (178): add “respectively” after “particles”. You could already mention there Fuchs correction factor and keep it explained below. RESPONSE: Added the word “respectively” (178). Second was not mentioned to avoid repetition and flow distraction.

Line 149 (176): Formulas need to be numbered RESPONSE: Equation numbers were added

Line 188 (216): given group? please explain RESPONSE: The NPF probability is calculated for the range of data in a specific group (time range, range of a given variable ex. for relative humidity from 50 to 55% etc.). Text was slightly modified to reflect this better.

C11

Line 191 (219): 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. RESPONSE: For the analysis done, data was separated into smaller groups as mentioned earlier. Thus, the term probability is considered more appropriate than frequency. This is clearly stated in the text.

Line 196 (224): low significance? Please give a number RESPONSE: The results found from the analysis of raw data, due to the large spread, almost never provided with any significant result (the R2 was always very low). A single number cannot be provided as the results are numerous. The word “statistical” was added.

Line 212 (241): extreme values? Please give a number RESPONSE: As previous. The cases that extreme values that biased the results were many. For example, an extreme value of wind speed (a single very windy day with no event) would result in an NPF probability of zero for that wind speed range. This though would result in biasing the whole analysis by a very limited range of data.

Line 239 (268): 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. RESPONSE: The sentence mentioned (the number of events extracted) was moved to the Results section as suggested. The events studied were those that all the work was focused on. While other, less clear events (without the expected growth, advected, uncertain etc.) were also extracted, they were considered only as exceptions or special cases in previous works. For further information about the events for each site as well as the comparative study between them, references were added in the Site Description section.

Line 245 (277): what is relevant data? Please explain more clearly in methods section

C12

and refer to it. Diagnostic features – wouldn't these be better in methods? RESPONSE: Added "depending on the variable studied". Relevant data refer to the data available depending on the variables studied, e.g. to find the frequency of NPF events, the days with available SMPS data were only considered. These diagnostic features are used to present the results in Table 2 and thus were not moved to Methods.

Line 252 (286): "slopes and R2" please use correct terms for these or more careful description RESPONSE: Changed to the terms gradient (instead of slope) and coefficient of determination for R2.

Line 261 (296): very strong? Please provide references to the exact figures in the figures/supplemental material when discussing results RESPONSE: In this case very strong correlations were considered for $R^2 > 0.75$ as explained in the parenthesis (and a clarification is added to any characterised correlation). The correlations (whether weak or strong) are found in Tables 3 and 4 (references added). References to the figures in the SI are not needed when discussing slopes and R2 and were only referenced when variable/unusual trends were found.

Line 279 (315): 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. RESPONSE: The value of the R2 was added in a parenthesis. Also, changed the numbering scheme for the figures in SI. References to figures in the SI that present results not in the tables (i.e. variable trends) were added in the text.

Line 296 (333): reference? RESPONSE: A reference was added

Line 301-303 (339): why? Could you explain? When describing results maybe worth mentioning these for various site specifics? Anything in common? RESPONSE: A possible explanation was added "This may be due to the different seasonality of the events found for the Greek sites (being more balanced within a year), as there was

C13

increased probability of NPF events for the seasons with higher RH compared to other sites, making it a less important factor for their occurrence."

Line 369 (414): which factors remain constant? RESPONSE: Added the word "meteorological"

Line 377 (420): reference? RESPONSE: References for this are found in the introduction

Line: 398 (441 - 443): maximum? Low? RESPONSE: Maximum changed to greatest. Low wind speeds changed to "close to zero wind speeds".

Line 420 (464): Ethesian: add few words what these are could be added RESPONSE: A brief description has been added ("a pressure system that develops in the region every summer").

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? RESPONSE: Added the definition of p-value (line 286). Used bold text for all correlation $r > 0.50$

Figures: no need to mention in each caption "of the present studies" RESPONSE: The phrase was removed

Line 433 (479): you could already mention here which pollutants (such as. . .) are studied and described in the upcoming sections. Added the chemical compounds studied in a parenthesis Line 752 (806): "at higher values"? Changed to "at sites with higher average values"

Line 755 (810): "meteorological conditions" such as? "such as temperature or relative humidity" was added

Line 756-757 (812): is that the only explanation? How about chemistry/composition at such type of site? Anything else that might play a role? RESPONSE: Added "compared

C14

to the urban environments and the more complex chemical interactions found there”
Line 782-783 (840): seems out of place here; it would be more suitable at the beginning
of conclusion section or removed. RESPONSE: Moved the sentence to the start of the
Conclusions section (line 796).

Interactive comment on Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2020-555>, 2020

#####

RESPONSE: Additionally, four authors were added in the list of authors

Table 1 was updated

Bousiotis, D., Pope, F. D., Beddows, D. C., Dall’Osto, M., Massling, A., Nøjgaard, J. K.,
Nørdrstrom, C., Niemi, J. V., Portin, H., Petäjä, T., Perez, N., Alastuey, A., Querol, X.,
Kouvarakis, G., Vratolis, S., Eleftheriadis, K., Wiedensohler, A., Weinhold, K., Merkel,
M., Tuch, T., and Harrison, R. M.: An Analysis of New Particle Formation (NPF) at
Thirteen European Sites, Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2020-414>, in review, 2020.

Bousiotis, D., Dall’Osto, M., Beddows, D. C. S., Pope, F. D., and Harrison, R. M.: Analy-
sis of new particle formation (NPF) events at nearby rural, urban background and urban
roadside sites, Atmos. Chem. Phys., 19, 5679–5694, <https://doi.org/10.5194/acp-19-5679-2019>, 2019.

Interactive comment on Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2020-555>, 2020.