Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2016-916-AC1, 2017

Interactive comment on "Estimation of atmospheric particle formation rates through an analytical formula: Validation and application in Hyytiälä and Puijo, Finland" *by* Elham Baranizadeh et al.

Elham Baranizadeh et al.

tuomo.nieminen@uef.fi

Received and published: 12 July 2017

We thank referee #1 for his/her thorough review of our manuscript. The comments were extremely valuable and we have redone most of the analysis based on them. The main weakness of the previous submitted version was the poor performance in estimating the growth rate with the mode fitting method, which further meant poorly predicted time lag and poor performance in the time resolved formation rate comparison. We have now reanalyzed the growth rates with the so-called maximum concentration method and the results are overall muck better. All figures and table 1 are modified accordingly,

and we also removed the old Figure 3 comparing time-lags, as we believe that it is unnecessary in the new version. We also removed the standard error color coding, related to the uncertainties when determining GR, from Figure 1.

Below we give our detailed responses to the referee's comments.

General Comments:
A thorough, data-based evaluation of whether particle formation rates can be extrapolated from measurements at larger sizes, as attempted by this paper, is vital for the aerosol community as so many data exist with only larger size information available. While the method used to tackle this problem is valid and useful, the evaluation requires development and more nuanced analysis before the substantial conclusions stated in the paper can fairly be reached (see below for specific comments).

Specific Comments:
A major assumption, that the two measurement sites are directly comparable with the method used for extrapolating nucleation rates, is made in the paper. Kurten et al showed that the method used, while valid in many circumstances, may not be valid for situations where pre-existing populations of aerosols do not dominate the coagulation sink and newly formed particles play a larger role in this sink. The differences in background aerosols at the two sites should be discussed in relation to this. Differences between the two site may also influence the magnitude of growth rates and coagulation sinks, which may affect the accuracy of the J extrapolation, which should be addressed.
The referee has a valid point that, generally, differences in aerosol dynamics (i.e which processes are dominant for the growth and loss of the newly formed particles) between two sites could potentially lead to erroneous conclusions when comparing the scaled J values. However, here this is not the case. The background distributions in our two sites are quite similar, both in total concentration and mode location. The mean values of CoagS of 7 nm particles are 5.3077$\cdot 10^{-5}$ s^{-1} and 5.3272$\cdot 10^{-5}$ s^{-1} in Hyytiälä (mean value of all analyzed NPF event days during 2002-2012) and Puijo (mean value of all

analyzed NPF event days during 2007-2015), respectively. For both sites the nucleation mode concentrations are so small that both the contribution of self-coagulation on growth as well as the contribution of newly formed particles on the sink are negligible. The contribution of particles of different sizes on the sink has been investigated by Lehtinen et al. (Boreal Environment Research 8, p. 405-411, 2003 – see fig. 3) for Hyytiälä size distributions. Particles below ca. 50 nm in diameter have typically negligible effect on condensation/coagulation sinks. As Puijo size distributions are very similar, this conclusion holds also there.
We added after Eq. 1 on line 110: "Lehtinen et al. (2003) studied the contribution of particles of different sizes to the condensation sink at Hyytiälä and found that particles below 50 nm in diameter have typically negligible contribution. This is a reasonable assumption at Puijo also as the concentrations and size distributions are similar to those at Hyytiälä. The mean values of CoagS of 7 nm particles are 5.31e-5 s-1 and 5.33e-5 s-1 in Hyytiälä (event days during 2002-2012) and Puijo (event days during 2007-2015), respectively."

Line 59: The assumptions that the coagulation sink is time independent and the growth rate size independent should be more fully investigates. Kurten et al. highlights the possibility and affect of time dependent coagulation sinks. If this is not a problem for these two sites it should be explained more explicitly.
This is true and in our analysis we do not take time dependence of CoagS and GR into account. This is, however, intentional from our part since we wish to follow the procedure of Kulmala et al. (Nature Protocols) in order to analyze formation rates consistently with most other studies previously analyzed.
Below in Figure 1 we show the median diurnal variation for CoagS (3nm). We also added in the revised manuscript Fig. 2 the CoagS time evolution for each of the three example events. It is clear that there may be significant time evolution in the CoagS/GR term of Equation 1, which is of course one of the key reasons why the simple approximation equation is not perfect.
The mentioned assumptions are mentioned in the text after Eq. 1, but to clarify this, we

added to the conclusions (lines 253-256 of the revised manuscript): "when considering detailed daily time evolution, the agreement is not as good. This is caused by three main things: 1. there are significant fluctuations in experimental size distribution data, 2. the extrapolation method assumes a constant value for CoagS/GR, and 3. there is a time lag between J3 and J7 and a poor estimation of the growth rate GR results in comparing values at different times."

Line 75: some discussion of how the different environments of Hyytiala and Kuopio affect the average size distributions and patterns of nucleation would be helpful here. This can affect how accurately equation 1 can be applied. Equation 1 assumes that the coagulation sink is dominated by larger pre-existing populations, which is less applicable in cleaner environments. If the Hyytiala environment is much cleaner, for example, than Kuopio, then the two situations are less comparable for this method of calculating formation rates. This assumption is mentioned on line 114, but it's validity for both situations requires further discussion.
See our reply to the first comment above: nucleation mode has negligible contribution to CoagS both in Hyytiälä and in Puijo. The coagulation sink levels in Hyytiälä and Puijo are very similar.

Line 145: averaging of m and CoagS(d1) between t and t' may be inaccurate, especially for high Js and low GRs – is the any indication of this in the data? How much do m and CoagS(d1) differ between t and t'?
The median variation of the CoagS over all the analyzed NPF event days is shown in Figure 1 below (not included in the revised manuscript). To illustrate the temporal variation we added into Fig. 2 of the revised manuscript also the time evolution of CoagS for the selected three NPF events. This variation naturally limits the validity of the constant CoagS assumption when applying Eq. 1.
We also added at the discussion of Fig. 1 (lines 196-199 of the revised manuscript): "This is caused by three main things: 1. there are significant fluctuations in experimental size distribution data, 2. the extrapolation method assumes a constant value for

CoagS/GR, and 3 there is a time lag between J3 and J7 and a poor estimation of the growth rate GR results in comparing values at different times."
The effect of the CoagS variation on m, and further on γ is, however, minor. We now mention this when discussing the result in Figure 1 (lines 199-200 of the revised manuscript): "The variation of CoagS with time also affects m and γ in equation 1. This is, however, negligible as CoagS(7 nm)/CoagS(3 nm) is a very weak function of time."

Line 172: This discussion of how well J3,est and J3,obs agree needs further development. Suggest removing qualitative judgement of 'reasonably well', and leaving only quantitative measurements of this. While the 0.78 correlation coefficient is helpful, the (linear?) fit result that this relates to would give a better measurement of the systematic difference between the two, this needs to be given here and on figure 1. This would then also quantify the following assertion that equation 4 overestimates the formation rate.
We have now removed 'reasonably well' and added the linear regression line to Fig. 1 of the revised manuscript, as the referee suggested. We also now show the results for both growth rate ranges studied: 3-10 nm and 7-20 nm. Note that, as mentioned at the beginning, we have redone all calculations – now using the maximum concentration method to determine the growth rate. Now, especially the time resolved comparison shows a much better result than previously. The slopes and correlation coefficients for the regression lines are 0.90 and 0.90 for the mean J3 values and 0.87 and 0.83 for the time resolved ones, respectively. There is a slight overestimation bias for small and underestimation for large J3 values. We have added this to the discussion of Fig. 1 in the revised manuscript (beginning of Section 3.1): "Figure 1 shows the comparison of estimated formation rates J3,est (Eq. (4)) with the observed ones J3,obs, as calculated directly from the measured size distribution evolution according to Eq. (2) in Hyytiälä. In the top figures, the size range 3-10 nm is used to evaluate the growth rate, in the bottom ones 7-20 nm. We analyzed 65 NPF event days for which the formation and growth rates could be quantified. Each data point in Figures 1-b and 1-d represents the arithmetic mean of the 3-nm particle formation rates (J3,est and J3,obs) for a single

NPF day during the time window from 07:00 to 19:00 local time. The mean is also a measure of the total particle production strength of each event. The results show that, when using GR in the range 3-10 nm, the estimated mean J3,est values correlate with J3,obs with a correlation coefficient of 0.90 and a slope of 0.90 using bilinear fitting. Furthermore, 91% of estimated J3,est are within a factor of two of the observed J3,obs. The corresponding numbers when using GR in the range 7-20 nm are 0.92 (correlation coefficient), 0.87 (slope) and 93% (J3,est within factor of 2 from J3,obs). Equation (4) seems to have a tendency of slightly overestimating the formation rate of 3-nm particles. There is not much difference in the results with different GR size ranges. The total means of J3,obs and J3,est (not shown in the figure) calculated using GR3-10 are 0.57 and 0.61 cm-3 s-1, respectively, confirming the tendency of Eq. (4) in slightly overestimating the 3-nm particle formation rates."

Line 175: Standard deviation should be given along with the daily means. These means are taken over a long period of time, during which I suspect J varies quite a bit. If J does vary a lot of this time period, then taking a daily mean is not very meaningful.
We chose not to add the standard deviations to the plot with means as the time resolved all-data-plot reveals the variation in J-values. The daily mean is meaningful in the sense that it is a measure of the overall strength of a nucleation event. Another variable choice would be the total number of particles produced at some size, but as most of the existing literature reports rates, we chose this approach.

Line 184: The quoted daily median values of J are very small. Where in the day do they actually occur? Is it actually before a significant nucleation event occurs? If so these values are not very meaningful – suggest either cutting data to only encompass the nucleation event or finding a more meaningful statistic here.
The median values were calculated over the same time window (07-19) as the mean values given in the text. We, however, agree with the referee that median is not necessarily the best statistic to use here, and decided to remove the median values from the text altogether.

Line 188: I would argue that reduction of percentage of points within a factor 2 of Jobs from 85% to 78% still reasonably significant and could indeed indicate that GR3-10 different and more accurate than GR7-20, which has strong implications for the conclusion that it's ok to use this GR in extrapolating results from Puijo. It would be more meaningful here to look at the fit equation again rather than simply correlation and percentage within factor 2 to understand this difference better.
We have now fitted the scatter plots of J3,est vs. J3,obs as suggested by the referee, and explained before, and show the fits in Figure 1 of the revised manuscript. Using the maximum-concentration GR for calculating J3,est, the effect of using GR7-20 instead of GR3-10 is much smaller and overall the results are better (91% and 93% of the daily mean J3,est are within factor of 2 from J3,obs).

Lines 191-195: Would prefer to see a full comparison of difference between observed and calculated Js here using GR3-10, GR3-7 and GR7-20, as well as a developed discussion of the degree of agreement and implications of this for using this method for J extrapolation. "Did not affect the results . . .by much" is too qualitative and glosses over what could be an important result here.
This is an excellent suggestion, and based on it we expanded Figure 1 in the revised manuscript to include results using both size ranges for the growth rate GR calculation, 3-10 nm and 7-20 nm. Figure 2 below shows the new Fig. 1 of the revised manuscript, and we also made the discussion related to it more quantitative.

Lines 196-203: The lack of correlation on temporarily resolved data here may indicate that the growth rates are wrong – this should be discussed here. It could also be because, by taking averages over long events where J varies significantly, the correlation seen early was simply an artifact of such heavy 'smoothing'. Is there another, meaningful measure of J (e.g. peak J of an event) that could be compared to asses this?
The reanalysis of growth rates with the maximum concentration method has improved the time resolved results significantly as explained before (see figure above). We also choose to stick with looking at mean formation rates as they are a good measure of

overall event strength.

Lines 205-208: Given the lack of correlation for time resolved Js, testing the affect of different GRs here does not have much meaning – suggest leaving this out completely.
Regarding the original version of the manuscript, we agree with the referee. However, now with much improved performance with respect to time resolved formation rates, this comparison is meaningful, we think.

Lines 211-212: "For some NPF days, the estimated time dependence and values of Jest are in fairly good agreement with those of observed Jobs." This statement needs better quantification to be of value. What proportion of days (since we're looking at a relatively small number of event, suggest quoting both total number of events examined and number of those with time dependence and value agreement here instead of just a percentage). How 'fairly good agreement' was judged needs explanation Also, are there distinguishing features of this sub-group where agreement is good? E.g. slow growth, classic 'banana' nucleation pattern?
With the new analysis for GR these figures (see Fig. 2 of the revised manuscript) have now also changed – and the results are generally much better. Still, the motivation behind Fig. 2 is the same: we show why for some events (and estimated GR) the analysis works better and for some worse. Thus we also chose not to give quantified information on the comparisons of Fig. 2. We explain this now clearly with the discussion of Fig. 2.

Line 213: quantify 'most of those days'
As the performance related to time resolved data is now much better we have modified this part of the text: "However, the time-dependency of J3,est is not consistent with J3,obs for some most of the days and, instead, typically the J3,est peak occurs earlier than the J3,obs peak (see e.g. Figure 2-e), indicating that our method of estimating GR is not always satisfactory perfect and typically underestimates the GR values."

Line 221: why does this burst of particles of 3-7nm occur and not then grow? Is this

indicated by the calculated GR and coagulation sink? Or is it perhaps a transport artifact? If it is the later it should be removed from the analysis as it is not nucleation. If it's the former then the equation used to calculate J3 should be able to handle it. Therefore this needs full investigation and explanation.
We investigated the event in more detail and found that it is a transport artifact. This is, of course, one of the general problems when analyzing events measured at one fixed location. In the figures we do not see the same aerosol growing, but particles formed at various location appearing at the measurement site at various stages of their growth. If we have a large enough homogeneous region of similar formation and growth, there is no problem. However, if there are inhomogeneities and the air mass transport direction changes during an event, we see dynamics as in fig. 2c and f. As this day was still classified as an event according to the protocol by Kulmala et al., we chose to include it – also to show what kind of challenges there can be.
We added at the end of section 3.1: "This is one of the general problems when analyzing events measured at one fixed location. We do not observe the same aerosol growing, but particles formed at various location appear at the measurement site at various stages of their growth. If we have a large enough homogeneous region of similar formation and growth, there is no problem. However, if there are inhomogenities and the air mass transport direction changes during an event, we see dynamics as in fig. 2c and f."

Lines 225-226: Estimated time-lag longer than observed time-lag indicates that the GR used is too low, which has implications for the calculated J and the time-dependence of the nucleation event. Can this explain the poor ability of this method to reproduce the time-evolution of nucleation events? This should be discussed.
Yes, true. As mentioned before, now the growth rate analysis has been redone using the maximum-concentration method, and this approach gives much better results.

Line 227: 15 days out of how many in total?
We removed the old Fig. 3 from the revised version of the manuscript, as we think it is

now not necessary and the text related to that figure was also removed.

Line 227: 1.5 hours difference: what percentage of the total time lag is this?
With our new results, having improved growth rates as well as a much better match between the observed and estimated formation rates, we decided to remove (the old) Fig. 3 as well as related text.

Line 229: quantify 'reasonably good accuracy'
For the $J_{3,est}$ calculated using GR_{3-10} from maximum-concentration method, the fraction of data points $J_{3,est}$ vs. $J_{3,obs}$ which are within factor of 2 is now 91% (67 out of 74 events) for the daily mean values. For all the 10-min data points it is 58% within factor of 2 (77% within factor of 3, and 84% within factor of 4). We now focus on the numbers (when discussing our results) and leave out these more vague statements.

Line 244: This monotonic increase in number of event days per year with time is indeed worth noting. Is this because of improvements in instrumentation/data quality? Change of activity or climate in the local area? Some discussion warranted. Do other things, such as total number of nucleation mode particles, size of coagulation sink, or anything else also monotonically change over this time period that might indicate why this is happening?
We took a look at this once again and now feel that this trend is far too short to be considered a significant trend. In Hyytiälä this time period shows a decrease both in SO_2 and CS, which have opposing effects on nucleation event probability. As we cannot quantify/justify such a trend with our supporting measurements we decided to remove the sentence.

Line 245: Given the lack of correlation shown early between median J3 est and obs, using J3 est here for analysis does not seem justified. Surely mean J, where some correlation between estimated and observed values was calculated is the value to use in figure 6?
This was a typo. The presented values are means, which makes much more sense.

Line 249: How does lower average GRs in Puijo affect the analysis? Lower GR gives larger time difference between J7 and J3, mean that inaccuracies in coagulation sinks and neglecting of time dependence of some quantities plays a larger role. Discuss.
The referee has a good point here. As we reanalyzed the growth rates using the maximum-concentration method, the average GR in Puijo is actually slightly higher than in Hyytiälä. We added to the revised manuscript (lines 236-237): "As the growth rates in Puijo are on average higher than in Hyytiälä, there is less time needed for the particles to grow from 3 to 7 nm. This means that our assumption of time independent growth rate and coagulation sink during growth should hold in Puijo as good as in Hyytiälä."

Technical Corrections:
Line 35: commas needed around 'at several locations'
Corrected.

Line 134: 'used a parabolic differentiation method ON the measured number concentration' instead of TO
Corrected.

Line 276: new paragraph needed for "the ultimate aim of this work"
Corrected.

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2016-916, 2017.

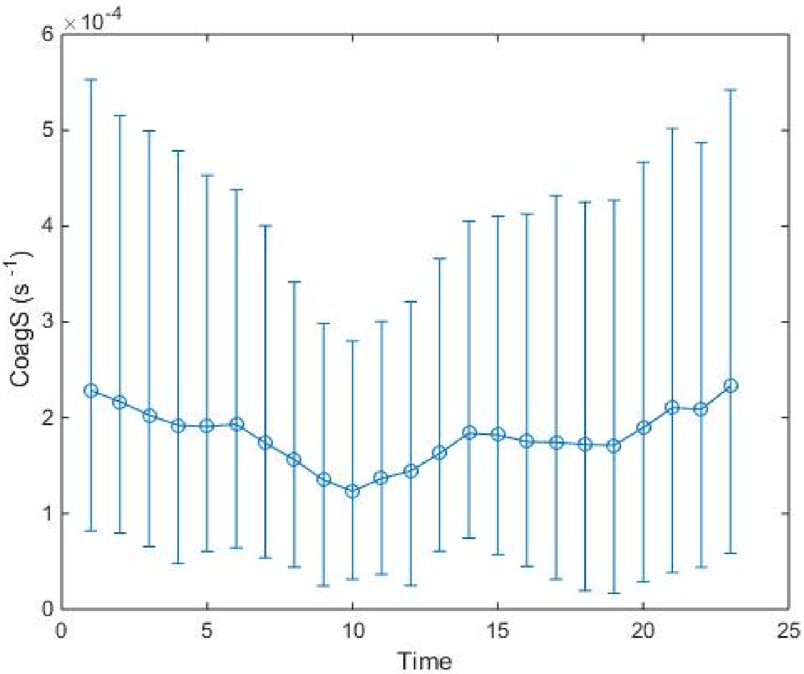

Fig. 1. The median diurnal variation of coagulation sink of 3nm particles for all the NPF events analyzed in this study in Hyytiälä. The error bars indicate the 25th and 75th percentiles of the CoagS data.

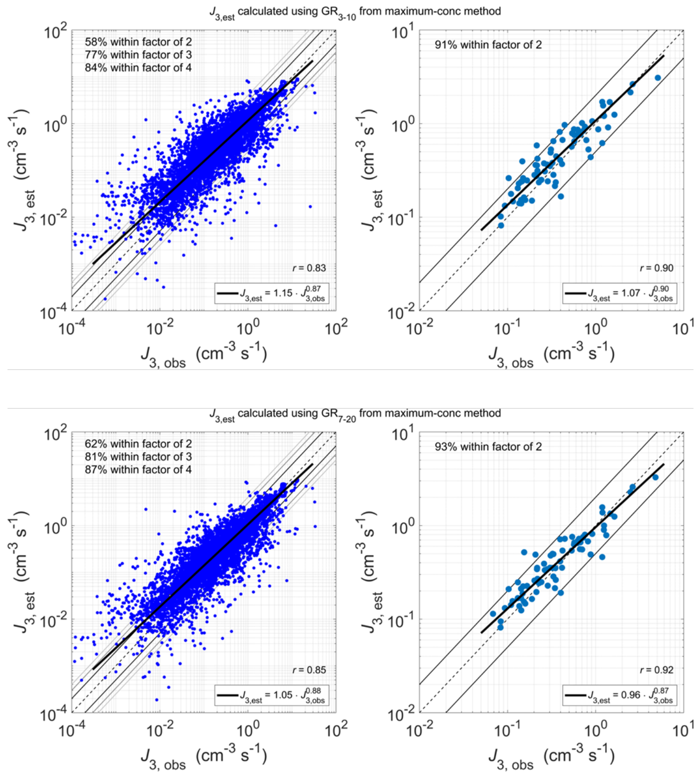

Fig. 2. Estimated J3,est and observed J3,obs formation rates of 3 nm particles in Hyytiälä, calculated using GR by maximum-concentration method. This figure is included in the revised manuscript as Figure 1.