Interactive comment on "Aerosol size distribution seasonal characteristics measured in Tiksi, Russian Arctic" *by* E. Asmi et al.

Anonymous Referee #1

Received and published: 19 August 2015

Asmi et al. present a valuable set of data from a poorly-researched region. The material is suitable for publication in ACP but the manuscript in its current form leaves quite a few things to be desired.

MAJOR COMMENTS

I. First of all, the clustering of DMPS data seems counter-intuitive. First data are clustered, then a combination of factors is proposed why the clustering produces the observed results. Why not use those factors (season, air mass origin, wildfires, etc) as parameters to sort the size distribution data? While clustering can be a useful tool (eg for trajectories), its advantages over a more intuitive approach seem unclear in this case. Why group data around a mathematical concept when actual, physical parame-

ters are readily available? Unfortunately, the clustering itself is hardly described at all in the manuscript; it is thus hard to assess the (presumably existing) deeper motivations for this approach.

Also the clustering results produce more questions than answers. While data points inside a cluster should be rather similar, the numbers in tables 1 and 2 suggest that inner-cluster variability with regards to N is as large or at times even larger than inter-cluster variability. Of course this is a result of normalizing the data, BUT: the reader cannot assess variability with regards to size distribution shape as this information is not given (figure 3). And this would be the point of all the normalizing, right? Especially terrible is the number of cases in some clusters in tables 4 and 5. When some clusters have 8000 cases and others merely 8 the discussion of differences between clusters doesn't seem all that relevant. I find it highly doubtful if those are clusters in the actual meaning of the word at all. They seem merely singular cases which don't represent much else than themselves.

Finally, four (4) different clusterings of the same data that are not based on actual physical characteristics seem a not-too-great idea when it comes to manuscript readability. Just compare "the normalized number cluster 6" to "the spring-time continental aerosol under cloudy conditions"; only one of those has some tangible meaning. In any case, the manuscript attempts to characterize the clusters in those terms - why then not start with them in the first place?

I would very much prefer the manuscript without the whole cluster business. I know this would mean substantial changes but I'm really not a fan. And considering the above issues it might be the easiest solution.

II. Given that the manuscript is not terribly huge, I wonder why new particle formation is kinda completely ignored. The abstract promises "valuable information on [...] sources of Arctic aerosols." Surely NPF is a source?

III. Wildfires. Given that they raise pollution levels to almost Chinese levels (page

18125) I wonder why the manuscript doesn't include a proper investigation dedicated to them.

IV. The supplementary information is technically extremely poor. Unzipping produces (among other things) an eps file of gargantuan proportions (100+ MB). And the captions are somewhere else. Just make it a nice and easy pdf file with all the information in one place.

DETAILED COMMENTS

1. The abstract has all those numbers, it thus also needs information on the DMPS size range.

2. Is arctic haze an aerosol source? I would rather call it a phenomenon.

3. 18110/11: "biogenic emissions in summer which affects the smaller, nucleation and Aitken mode particles." - should maybe be: "biogenic emissions in summer which affect the nucleation and Aitken mode particle concentrations." or so

4. 18110/14: "significant" - how significant? 50%? 100?

5. 18110/17: "nearby" -> "local"

6. 18113/16: "theoretical diffusional losses during the particle transport are taken into account in the data inversion." - while i agree that this is somewhat trivial and doesn't need equations and such, at least a reference would be nice.

7. 18114/2: "charge of the regular maintenance done weekly" - what is this regular maintenance?

8. 18114/4: "quality checked weekly" - what is this quality check?

9. section 2.2.1: the nominal cut-off of the 3772 is 10 nm if memory serves correctly. how do you measure down to 7 nm? sure, most readers can guess. but it should still be mentioned.

10. 18115/1-5: so how do the data differ? i suppose you have some overlap where data from both locations are available.

11. 18115/14: "additional 36%" - 36% of 74% or 36 %-points?

12. 18117/about the APS: you focus on a comparison of counting efficiency. how do you know that APS sizing is reliable? and if it isn't, what sort of error does this introduce?

13. 18119/4: "median values extending from 3 to 5ms-1." - what exactly does this mean? daily medians? weekly? monthly? i also wonder if the std.dev wouldn't be a better metric to describe how stable wind speeds have been.

14. 18119/5: "sudden onset" - i don't see anything sudden in the figure. and really, why would there be?

15. 18119/6: "RH decreases" - this all refers to RH_water, right? or is the picture the same for RH_ice?

16. 18119/12: "median wind speeds" - median what? daily? weekly? monthly?

17. 18119/26: "each size distribution were normalized with its maximum concentration" - what does that mean? i.e. what is the max. concentration? the bin with the highest concentration? is that really the best way to bring out differences in size distribution shape? why not use total N instead?

18. 18120/3: what's the "peak mode"? i mean, it is quite intuitive but i don't think it's actually a thing. if it is, as i assume, the mode with the highest concentration, than one should probably say that the highest concentrations are found in the aitken mode which has a diameter of xx nm. or whatever.

19. 18120/first paragraph: the paragraph really doesn't say anything about nucleation? simply looking at the numbers, one would think that those clusters represent different times after nucleation with 3 being the freshest and 1 the oldest. or is there a reason

not to think that way?

20. 18120/2nd paragraph: these lines just state the minimum information from the figure. some interpretation would be nice.

21. all 18120: in addition to medians, i always find it nice to know something about the range within a data set. which means that quartiles would make a suitable piece of information. probably also in other sections.

22. 18121/first paragraph: lacks a statement of what is typical for continental and marine clusters. of course, referring to the major remarks in the beginning, this is a backwards approach. to me, it would be much more intuitive to start by clustering the trajectories and then separate the aerosol data by trajectory cluster.

23. 18121/last paragraph: these four modes are standard textbook fare. i am not quite sure what you mean by saying you "identified" them.

24. 18122/2: "Total particle number concentrations (Fig. 4a and Table 3) show two annual maxima in March and in July." - really? i don't see it. the attribute i would use for the data in figure 4a is non-descript or something along those lines. to say there is an annual cycle in Ntot seems like an over-statement

25. 18122/3: "A similar interannual cycle is seen for each particle mode separately" - while the previous comment may be a matter of taste, this isn't. Nnuc has neither regular march nor july maximum. one is shifted to february half of the time, the other one sometimes happens in june. and to call these tiny differences cycle is too optimistic. similar things could be said about the other modes. the thing of course is that the figure is an unfortunate choice if cycles are to be presented. log-scale diminishes the differences one would like to see (and which the table says exist). also, if it's a cycle we want to see, why not make a figure that actually shows a 1-yr-cycle instead of the full time series? i would ditch figure 4 AND the table and make a new figure with the data that is now in the table.

C6053

26. 18123/8ff: once again, i fail to see the benefits of this cluster business. what information do we gain from this detour?

27. 18125/1ff: i have said this before but i want to point it out again: mass concentrations of 65 just show how silly clustering can become. "cluster" 4 is not a cluster but a single event that produces shanghai-level pollution in siberia. it is no doubt a very interesting event (that would deserve a bit more attention in the manuscript) but it's not a cluster in very much the same way that one person is not a group.

28. 18126/17: "y = 0.014x +0.12" - units?

29. 18127/25: "In addition to natural vegetation sources, another temperature dependent aerosol source in summer Siberia are the intense fire events, as already discussed in Sect. 3.3." - yes! but where is the proper analysis? why not identify all trajectories that came through wildfire and then have a closer look? it would also be beneficial to exclude them from the current analysis to determine how much of the temperature dependence is actually related to BVOCs.

30. 18128/1: "the maximum average CCN100 concentration" - average over what?

31. 18128/3: "only 13 cases were available" - which would make it all the more important to determine which of those are wildfire-related.

32. 18128/section 3.4.2: it would be nice to see a little investigation (a case study if nothing else) how these pollution events develop.

33. 18129/7: "interannual" - i guess it should be iNNerannual. iNTerannual variability (fig. 4) is not that large.

34. figure 1: the map is terribly small. the photos, on the other hand, are not that essential.

35. figure 3: as said before, it is unclear how the data for the figure were normalized.

36. figure 4: too messy, too small, labels hardly readable.

37. figure 6: the temperature information should also be in the figure, not only in the caption.

38. figure 7: fit quality information for 7a is missing. also: how did you determine that linear and quadratic fit are the only options? why not try an exponential fit? this figure and the fit would probably benefit from an exclusion of wildfire-related episodes. the same is true for figure 8

39. figure 9: this is not terribly important, could move to supplement. a figure that describes a haze event could be nice though (as said above).

40. supplement: make a single pdf.