**Atmospheric
Chemistry
and Physics
Discussions**

Interactive comment on "A statistical analysis of North East Atlantic (submicron) aerosol size distributions" *by* M. Dall'Osto et al.

Anonymous Referee #3

Received and published: 26 September 2011

In this paper, the authors use a relatively new statistical technique to categorise aerosol number-size distributions from 2008 collected at Mace Head at the GAW site. Size distributions with similar characteristics are binned into 1 of 12 types known as clusters. These clusters are then grouped into 4 more general classifications, namely coastal nucleation, open ocean nucleation, anthropogenic and background clean marine. Although the results do not present any major new insights, they do provide a means for statistically analysing large datasets in a routine way and attempt to relate these to other parameters such as air mass history.

With some re-working the paper will be acceptable for ACP.

Major Concerns.

The authors switch between describing individual clusters and the 4 average categories and I feel there needs to be some additions to bring all the analysis together. For example, the back trajectories starting with 'm' produce or contain all 12 clusters, concluding that air originating from those sectors are less sensitive to long range effects rather local or regional processes. The coastal nucleation event clusters are slightly mis-leading as they are presented as a separate class when in fact they are size distributions which could be background clean marine or open ocean nucleation which have had ultra-fine particles added to them. It would be interesting to know or to speculate what cluster they would fall under if only data from $Dp > 50nm$, say, was used. To me, the bigger picture analysis shows that in the cold winter months, the classical bimodal distributions of the background marine can dominate from the clean sector due to the meteorology. As the weather warms, then contributions from open ocean nucleation can add to the background and/or as the air approaches the coast then ultrafine particles can contribute to the size distribution depending on the solar radiation and tides. This is all perturbed if the wind in anthropogenically influenced by the local wind direction (clusters 6,9 and 10). This is touched upon in parts, but not really summarised. Also, how will this help the global models discussed in the introduction? How would a modeller use this analysis?

I think the manuscript would benefit from a short section, maybe a very small appendix, on the cluster analysis itself. The paragraph in the text does not really explain the technique and relies on the reader to research the Beddows et al reference. Some details on the basic process and defining all the terms, how the clusters are grouped, why there are 3 clusters in each of the 4 general categories (or is that just luck?) and a plot or table of the diagnostics used to validate the results of the analysis. This provides a reference for other people wanting to use this technique whilst not clogging up the body of the text for those just interested in the results.

There are a lot of grammatical errors and the manuscript needs a final proof read. For example bi-modality and bimodality are both used in the abstract. There are errors in

the results section when referring to numbers in tables. For example, Page 21686, Ln 22 states cluster 7 has the lowest Nd>3 of 892. Clusters 11 and 12 have ND>3 of 764 and 773. This happens quite often and made the review much harder. The authors need to check all numbers in the text against those in the figures/tables and make sure these errors do not compromise the interpretation of the results.

Specific concerns:

Why is 'cluster' capitalised in the text?

In the introduction, the author states that 60% of the air arriving at the station comes from the clean sector and that clean air is defined as at BC concentration less than 50 ng m-3. Yet in table 3, there is not a single cluster with a BC loading less than 50ng m-3. Can the authors clarify this please? Was 2008 a dirty year?

Please refer to tables as 'Table 3' for example,and not table 3e. If the authors feel they need to identify a column specifically, state it.

21679 Line 25 onwards. Please be consistent with units. nm in one line, then switching to um in another for the same aerosol mode.

21680 Ln 27-29, "Some examples of particle size distributions Cluster analysis for substantial SMPS datasets can be found in the literature. Similar approaches have previously been used:" Please can you clarify these sentences. Are the similar approaches the same as the analysis for the substantial SMPS datasets? If not, please give examples. Should the sentence read: "....found in the literature, where similar approaches etc"?

21682 The section on the instrument description needs a little bit of tidying up. Firstly, please be consistent with the instrument name. In this section the 3025A is referred to as TSI 3025, 3025A CPC3025. Which one is it? Both a 3025 and 3025A exist as model numbers. Also, technically, 3025's are Ultrafine Condensation Particle Counters (UCPC). The description also first introduces the CPC's then the SMPS and back to

the CPCs. The section would read much better if all the CPC details were in the first paragraph and then the SMPS' described. Finally, please clarify how you can have 88% data coverage but some hours are not available so there is only 75% data coverage? Has the data been filtered for contamination or below detection concentrations?

TSI Inc no longer stands for Thermo-Systems Inc. It is just TSI Inc

21683 ln 15. Define WS (first use of) Ln 19 WD define. Ln 15, Please explain what a meteorological discontinuum is. Do you mean the pressure was on average lower in August then July/Sept? This is poorly written for a journal ".....somehow creating...."

21684 Ln 11, suggest replacing "spring was associated with" by "spring experienced more". Ln20/21 the author is referencing figure 4 twice. It is repartition.

21685 line 5. Difference is the second largest, not the largest.

21686, Ln 7 there is no table 3d, Leave as table 3 or table 3, column d. Same for table 3e, Ln 21 why is it peculiar that a background marine number concentration should be low? Also, it is not the lowest, clusters 11 and 12 have ND>3nm <892

21687 Table 3e should just be table 3. Ln 12 the TEOM loading for that cluster is 16.5. Ln 18. The ND>3 is in fact he lowest at 764. The difference is the second lowest. Ln 21. Maybe justify that statement about the coarse mode and sea salt with reference to the higher wind speeds and the O'Dowd paper 1997 (I think). Ln 26. The ND>3 at 773 is the second lowest. The TEOM concentration is 13.3

21689 Ln 2, Sect. Should be section. Ln 5 replace among with during.Ln 13 they showed the highest ND>3 concentrations, not ND>10. Please be specific. Ln 15, 2840 is not lower than 1905. Ln 16. The mode is not at about 9 nm, that is where a SMPS point is. The mode itself, if a mode was fitted, would be around 10-11nm. I am being pedantic, but the authors claim Aitken mode aerosol are for Dp>10nm. This really strikes at the heart of distinguishing what class of aerosol cluster 3 is and how the classes are defined. I would argue there is good cause to have cluster 3 at the

open ocean nucleation cluster. Ln 20 - page 21690. This needs some re-wording. I can see the shift in the peak of the occurrences, but the explanation needs elaborating more for readers unfamiliar with coastal nucleation.

21690 Ln 20 Cluster 8 had the 6th highest BC loading.......Ln 23, should it be 4,5 and 8?

21691 Ln 12. Table 2 not table 1. Also, I do not agree with the statements in this section. Firstly, table 2 does not show correlations. It lists occurrences. Secondly, cluster 9 cmP and cP have occurrences of 15 and 16, the two lowest. It is more likely to find cluster 9 in mP air masses. Furthermore, how does 6 fit into this picture of air mass origin? Is this basically telling the reader that the source of the air does not matter if the local wind direction is outside the clean sector? That is my conclusion. Also, the last statement needs re-wording and clarifying. The sub micro particles are dominated by the nucleation events in number. What I believe the authors to mean is that the scattering data confirms the large number of accumulation mode aerosol, as seen in figure 4c. Ln 24 cluster 1 is more westerly than cluster 11 and close to cluster 12.

21692 Ln 2 wrong table referenced again. Also, the argument about the high scattering and high PM does not hold up against the data from other clusters. There are other clusters with PM loadings higher than cluster 7 but with lower scattering eg cluster 1 + 3. Cluster 9 has a lower PM loading but a similar scatter to cluster 7. Can the authors discuss this? The nepholometers will only detect particles down to about 100nm, so it is not the nucleation particles having an effect.

21693 Sect should be section. Ln 4 - 10. This needs to re-writing. Comments like Cluster 3 did not present a clear seasonality, somehow in between Cluster 1 and Cluster 3. It should be cluster 2 at the end and if there is no clear trend how can it 'somehow' be between the other two? A trend exists or it does not. Ln 23 cluster 5 peaked in September, not the summer months. Ln 29 cluster 9 does not spike, it is a clear

seasonal trend. Also, where is figure 11?

21694 Ln10 - 19. This is complete repartition. There is nothing in table 4 that cannot be seen in figure 7 and has not already been stated. Remove this section. Ln 25 from Yoon et al onwards. How does this discussion contribute to the paper?

21695 Ln 13. The authors results do not support the statement that cluster 11 dominates summer months. It is clear there are cluster 11 events in winter/spring.

Table 2. Given that the authors have assigned clusters a type, this table would be better organised into these types, rather than in ascending order of cluster number. Eg, 1,2,3 4,5,8, 6,9,10, 7,11 and 12. Same for Table 3 and 4.

Fig2. Why are the dots joined up? This is a frequency plot, not a time series. Markers or bars only please.

Fig 4. Legend incorrect. I believe the last sentence should be (e) shows average. The legend also needs explaining. Why if e shows the average of the 4 cluster types are there 5 traces?

Fig 7 and table 4 are showing the same data in different formats. Suggest using just one and organising them into the cluster types.

Interactive comment on Atmos. Chem. Phys. Discuss., 11, 21677, 2011.