Atmos. Chem. Phys. Discuss., 6, S5057–S5067, 2006
www.atmos-chem-phys-discuss.net/6/S5057/2006/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Cluster analysis of the organic peaks in bulk mass spectra obtainedduring the 2002 New England Air Quality Study with an Aerodyne aerosol massspectrometer" *by* C. Marcolli et al.

C. Marcolli et al.

Received and published: 5 December 2006

We thank the referee for his/her thoughtful comments.

Response to general comments:

The paper presents the first use of a data analysis tool based on hierarchical clustering, first developed for use with the PALMS laser ablation single particle mass spectrometer, to deliver information on the mass spectral fingerprint of organic material as measured by the Aerodyne Aerosol Mass Spectrometer. In my opinion this is a very worthwhile goal as there is a considerable amount of information in the organic mass spectra of

the AMS that has to my knowledge not yet been mined effectively. The paper also tries to identify on the basis of these clusters the sources and processing of the measured organic aerosol mass. It goes some way to convincing the reader that several of the categories are likely to be biogenic nature, but this is largely by comparison with gas phase precursor data that from the derived MS.

Response: in section 3.2, the biogenic nature of categories 2-5 is also established based on mass spectral correlations with reference spectra.

It also tried to convince the reader in several places that the biogenic signatures seen in several classes age towards the primary oxygenated class. There is no sound basis for this argument and the authors need to speculate once and leave it there. The paper is well written and offers some new insights and makes a valuable contribution to the body of work associated with the development of the AMS. Several referees have already commented extensively on this paper in considerable detail. In particular, referee 4 makes a number of very important detailed points that need to be considered. The authors' responses to these points are largely well considered and I will not dwell on those in this contribution. There are, however, one or two points that remain outstanding that I feel should be considered by the authors. The main problem area in the paper, which is identified by all the reviewers, surrounds the validity using a method developed for clustering different particle types together based on a measurement of the chemical characteristics of single particles and probing the number frequencies with which these clusters are observed, and applying it to an data stream that delivers data on the composition of the ensemble of particle in the atmosphere in a given time interval. There is nothing incorrect in such an approach in itself and it can, as the authors show deliver significant insight into the organic mass fraction of the aerosol. However, what is missing from the paper is a real discussion of what information can be retrieved from such a method and what the pitfalls are. Hierarchical Cluster Analysis essentially groups similar mass spectra together over a given time interval to form a small number of clusters that are of distinct character. However, when applied to PALMS and

ACPD
6, S5057–S5067, 2006

EGU

AMS data this means two very different things. In the case of PALMS, the individual mass spectra represent chemical signatures of different particles and the clustering represents a way of grouping chemically similar particles together and assessing the relative frequencies of occurrence of these groupings i.e. how many particles of a given type are sampled in a given period. When applied to AMS data the analysis means something quite different. This message is not really stated explicitly and in my opinion it should be. The criticisms made by the reviewers are fair and could be addressed with a separate section in the paper entitled something similar to "Differences between using Hierarchical Cluster Analysis for analysis of single particle and ensemble averaged mass spectrometric data". This should lay out the general applicability of the method for single particle and ensemble averaged data sets and highlight what is delivered and what cannot be inferred from either method. The remainder of the paper should then be an assessment of the power of HCA on ensemble averaged data and its shortcomings, as it is currently written, with an extension of the discussion section as indicated by reviewer 4. I will not go into detail of what is being done as several of the reviewers have laid this out previously. I will though offer a simple example that illustrates the points made by referees 2 and 4. Suppose a high mass loading is experienced for a small fraction of time in a given period and the rest of the time low loadings are observed. The current analysis will provide a fixed number of MS distributed evenly in time, each of which are normalised. So the period of high loading is neither proportionately represented on either a mass or number basis but on a temporal basis. When the MS are combined based on their similarity to obtain different classes, the information on mass or number population will not be immediately available. Rather, what is represented is a comparison of the relative dominance of a cluster type during a given time period. However, by combining the results with other forms of time series and correlation analysis the types of aerosol contributing to the AMS data record in certain periods can be identified and some conclusions drawn. This is what the paper is about and it can have an important contribution to unpicking different organic aerosol classes. I have no doubt, given by the authors' responses to the previous reviewers, that they

ACPD

6, S5057–S5067, 2006

are aware of the benefits and limitations of the method. However, the authors do not convey the essence of these complexities in the current version. To save the reader or subsequent user misconstruing what the paper is trying to achieve and to make it plain what is being tested I would like to see that at the end of section 2 Experimental Methods, the authors spend a paragraph or two discussing these issues in a separate sub-section 2-2. I am sure this will leave the reader in no doubt what can and cannot be done and how the results should be interpreted. The following sections then serve as a test of the power of the method.

Response: We understand that we have not been explicit enough about these issues in the current manuscript and will improve this in the revised version. We add a section on the comparison of cluster analysis with single particle or ensemble data. Note that single particle applications also do not include mass information and have their own pitfalls with laser ionization and particle transmission biases.

The second area I wish to discuss is the use of the method on this type of dataset. As far as I can see the clustering works by finding the minimum dot product of two MS from evaluating every pair of MS in the whole dataset, averaging these two MS and then repeating the process on the reduced umber of MS until a certain number of MS groups remain. At each stage the clustered MS are treated identically to a signal MS and with the same weighting, at least I see no comment to the contrary. In effect this means that a final cluster is an average of two mass spectra, both of which may be groups of a large number of MS but they may be simply an average of a large group and a single MS. In the latter case the final MS will be significantly different from the bulk of the MS in that class or for that matter from either the mass averaged MS or the number averaged MS (the latter can be gained by normalised each MS in the class and averaging them). It may be useful to compare these and illustrate the differences. The use of HCA is very useful for identifying aerosol types and atmospheric conditions and is therefore a key weapon in the data analysts arsenal but great care needs to be taken when using the MS retrieved from this type of analysis as it can be greatly biased by

ACPD
6, S5057–S5067, 2006

EGU

the way the classes are formed. As in this paper the final classes derived from the field data are then compared with single MS taken from the laboratory, the authors do need to discuss this and take great care to convince the reader that such effect are small.

Response: the clustering works by finding the maximum dot product. The clusters are treated the same way as individual mass spectra to find the highest dot product, but not to calculate the new average spectrum. Here, the weight of the clusters is proportional to the number of spectra they consist of. Otherwise, the influence of the last spectra that are added to a cluster would be unduly high. We kept the explanation of the method short because it was the same as in Murphy et al. 2003. In the revised version, we add to the experimental section the following sentence: The number of spectra in each category is retained to obtain a weighted average when subsequent spectra are included.

The last main area I wish to pick up on is the area of counting statistics and the use of zeroing the negative mass peaks in the individual MS prior to averaging. This is something picked up on by reviewer 4. The response by the authors is not satisfactory in my opinion, nor is it in the mind of reviewer 4. I will again illustrate the problem I have with an example from the paper. Suppose that the particle population is chemically identical in two separate periods but the mass concentrations are different such that in the first period there is considerable mass and many peaks show signal well above the noise level. In the second period, the signal is much lower and many more peaks are not present above their detection limit. In the latter these peaks would be set to zero and the relative importance of the remaining peaks would dominate. Hence two distinct clusters would be present for two populations that are chemically identical. This will not happen when applied to the PALMS data where the ion yields from single particles can be directly used. The authors really need to consider this aspect it can seriously compromise the results if one is not very careful. I suspect that the clusters 7 and 13 suffer from this. The relevance of cluster 13 in these circumstances must be questioned. I would strongly urge the authors to apply some criteria to ensure that an

ACPD

6, S5057–S5067, 2006

artefact of this type does not occur. As reviewer 4 states, I fail to see the need anyway.

Response: We certainly will consider the arguments of reviewers 4 and 1 when applying cluster analysis to a new dataset. As is done for the PALMS data, we normalized the spectra before calculating the dot products and when the spectra are normalized, any variation in total signal is removed, e.g., the normalized spectra are the same for two identical spectra with different total intensities. For the present dataset, the most successful way to perform the cluster analysis was to use the linear signal. Like this, the five to ten most intense masses were responsible for the clustering (note that the dot product was used as the similarity criterion). These masses are the least affected ones by noise. A too high noise level will certainly influence the clusters that are formed and reduce their significance. We agree with the reviewer, that in this case, the outcome will be different whether the average error is subtracted or not. Case in point: categories 7 and 13. The dot product between categories 7 and 13 is 0.76 and the average organic mass for category 7 spectra was 3.3 +/- 1.8 microgram/m3 whereas for category 13 spectra it was 1.2 +/- 0.5 microgram/m3. To investigate the effect of zeroing, we generated a spectrum identical to the category 7 spectrum but with zeros instead of the lowest signals that represented approximately 50% of the total signal. The dot product of this artificial spectrum with the full signal category 7 spectrum was 0.94, which indicates a slight effect of zeroing low-signal linear data since most of the signal is in a few peaks. However, the main difference between category 7 and category 13 was the relatively stronger signal at m/z 44 in category 13, leading to the relatively low dot product between categories 7 and 13. Indeed, when the signal at m/z 44 was excluded the dot product of category 7 with category 13 was 0.82 and category 7 with the artificial spectrum was 0.85, demonstrating some similarity between the two categories. This reinforces our conclusion that the principle difference between the two categories is the relative intensity of m/z 44 which we have described in the manuscript.

Specific Comments:

ACPD

6, S5057–S5067, 2006

Page 4607 line 17 insert ‘a’.

Response: we do this.

Page 4607 line 20-25 The dataset was divided into 4 parts and the analysis was run separately to start. This implies that after some period all the data were combined to complete the analysis. Was this the case, it should be said explicitly? Though this eases the computing requirement does this compromise the data analysis in any way?

Response: For the first three passes, we used a strict stopping criterion, so that only very similar spectra were clustered. The influence of dividing the spectra in four subsets was thus minimized. We change the sentence to: the clustering algorithm was run separately for each subset with strict criteria until the number of categories was low enough to be easily handled in one set for the last pass. We also add the sentence: At the end of clustering, a final pass is made to ensure all of the spectra are placed in the appropriate category, which reduces the bias of the sequence of combining).

Page 4608 line 8: “..quite small molecules or strongly ˇ E” reword to “..quite small molecules or was strongly ˇ E”

Response: we do this.

Pg 4608 line 25: m/z 30 can also be due to non nitrated organic, this needs to be said.

Response: We have added several sentences about other sources of m/z 30 (and 46) to the experimental section as well as the section discussing the categories with a large m/z 30 peak.

Pg 4611 line 2, but only a few spectra of simple anthropogenic systems are available and it is widely acknowledged that these may not be representative of anthropogenic SOA as a whole.

Response: We fully agree with the reviewer. We mention this caveat further down in the manuscript by stating: However it must be noted that the delta patterns for the aerosol

ACPD

6, S5057–S5067, 2006

products of only one aromatic anthropogenic precursor compound (meta-xylene) have been investigated.

Pg 4612 line 1: “Although the contribution to any wind direction ˇ E” surely you mean “Although the overall contribution across all wind directions..” as the first statement is not true for some particular directions.

Response: the sentence is true for any wind direction. Note the different scales for categories 1 and 2 compared with 3-5.

Pg 4613 lines 8-11: The argument made here that the maximum organic mass being observed in the afternoon and correlating with the solar radiation is consistent with SOA formation and gas to particle partitioning only holds if the SOA formation is in situ. Is this likely to be the case? If true this implies a formation rate of around 0.6 ugm-3hr-1. This is pretty substantial and implies significant photochemistry. Category 1 dominates the aerosol mass and shows the same feature, yet the argument is made elsewhere that cat 1 is aged aerosol. It is therefore unlikely that such an appreciably amount of SOA will form in such well aged aerosol, essentially the organic aerosol mass changes by 10% per hour a significant production rate implying a youthful and very actively photochemical airmass. The authors need to discuss this in more detail.

Response: the diurnal averages shown in Figure 9 include pollution plumes as well as pristine air masses. The increase in organic mass is especially strong during plume events, and these periods therefore contribute above average to the observed trend. During the high organic mass event on July 22 and 23, when sampled air masses had been impacted by major urban areas such as New York City and Boston, the organic mass increased by about 1 ug/m3h during the first few hours of daylight. We agree that this is a significant growth rate. This air mass indeed seems to have been highly photochemically active although category 1 was dominant. Interestingly, on July 23, while the organic mass grew during day, the occurrence of category 2 increased (from 7% to 23%) at the expense of mostly category 1 (from 93% to 73%), indicating that

ACPD

6, S5057–S5067, 2006

SOA precursors were still available maybe by mixing in of fresh air masses. Indeed, the photochemical age during this time period was less than 10 hours. We therefore think that the SOA formation was in situ.

Page 4613 lines 20-24 This is speculation as far as I can see there is no justification for arguing that processing is responsible for transforming cat 2 to cat 1. The arguments that lead to a source of cat 2 have some merit but cat 1 can arise from either anthropogenic or biogenic sources from the results presented. In fact cat 1 arises from air masses impacted by pollution by and large and not from other regions implying pollution plays a role, either as the source of the SOA or by creating the photochemical environment in which the SOA are formed to give the fragmentation pattern observed.

Response: In the manuscript, we declare the transformation of category 2 to category 1 spectra as one of several possible reasons for the observed trend. A transformation of category 2 to category 1 spectra gains additional plausibility from the fact that category 2 is most abundant when winds are from the direction of the major urban areas. The time evolution of category 2 therefore reflects the aging of the air masses as derived from the anthropogenic tracers. Also, the average photochemical age of category 2 spectra was less than that of category 1 spectra.

Pg 4614 lines 1-5 as stated by another reviewer I believe, the photochemical marker is only of use for anthropogenic air mass tracing and aging. It is meaningless for biogenics.

Response: We totally agree that the photochemical age determined by anthropogenic tracers only applies when the aerosol and the tracer have the same history. This seems to be the case for category 2 which is most abundant when the winds are from the southwest since this direction also hosts the source region of the urban plumes. We think that the low correlation of categories 3-5 shows their independence of anthropogenic VOCs.

Pg 4614 section 3.5 I rather agree with reviewer 4. This discussion is rather spec-

ACPD

6, S5057–S5067, 2006

ulative. Some back trajectories, estimates of time from source and possibly source modelling is required to investigate whether the timescales postulated are reasonable.

Response: Unfortunately, the HYSPLIT back trajectories are inconsistent with the gas phase, aerosol, and wind field measurements for this time period and location. For example, the data from the night shows high monoterpene concentrations and the measured surface winds were from the north (see Figure 13 in de Gouw et al., 2005). Hence, the data indicates that the monoterpenes left the coast at night and were not exposed to sunlight prior to sampling. In contrast, the back trajectories for the same time period indicate that the air mass came from the east and spent some time in sunlight from the previous afternoon prior to sampling. If this were true, the monoterpene concentrations should be relatively low. As the day progressed, the measured surface winds were northeasterly and the monoterpene concentrations decreased after sunrise (around 1000 UTC). The isoprene source term remained high for a few hours (from 1000 to 1500 UTC), but the isoprene concentrations decreased like the monoterpenes. This is consistent if these species react with sunlight while over the water. Yet the back trajectories indicate that the air mass was above the continental nocturnal boundary layer and should not contain any photochemically-active biogenic species. We added this to the manuscript: Shipboard measurements from 1000 to 1500 UTC indicate that the surface winds were from the same direction after sunrise, the wind speeds were approximately 5 knots, the ship moved much more slowly, and its track roughly paralleled the coast where the emissions were relatively constant. Hence, the time since emission for the air mass sampled from 1000 to 1500 UTC was approximately 6 hours. Therefore, the air sampled had left the coast during the night and was exposed to increasing levels of sunlight as the day progressed.

Pg 4615 bottom again this is pure speculation, there is nothing definitive here at all, statements such as this should be removed. The accumulation of statements like this leads to the reader's overview that it has indeed been shown that cats 2-5 are transformed into cat 1, it has not.

ACPD

6, S5057–S5067, 2006

Response: we delete this sentence in the revised manuscript

Page 4616 line 12: The products do not need to be the same, simply the chemical functionalities of anthropogenic and biogenic products, it is these that give a characteristic signature to the AMS.

Response: We will change the manuscript accordingly.

Section 3.7 See my general point, to what extent are 7 and 13 different, can artifacts due to zeroing sub noise data be excluded?

Response: Although the effect of zeroing sub-noise data may cause spectra with lower overall signals to be placed into a different category, the main difference between these two categories is due to the m/z 44 peak. We now mention the somewhat higher similarities in the remaining peaks (see answer to the general comment)

Page 4619 I agree with referee 4, this section needs to really explore what the technique is telling us and what its shortcomings are.

Response: We do this in the revised manuscript.

Figure 1 Use degree symbols for latitude and longitude

Response: We do this in the revised manuscript.

Figure 9: top panel left axis label should read microgramme not gramme

Response: Thank you for pointing this out.

Interactive comment on Atmos. Chem. Phys. Discuss., 6, 4601, 2006.