Atmos. Chem. Phys. Discuss., 6, S1947–S1953, 2006 www.atmos-chem-phys-discuss.net/6/S1947/2006/ © Author(s) 2006. This work is licensed under a Creative Commons License.



ACPD

6, S1947–S1953, 2006

Interactive Comment

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: 20 July 2006

We thank the referee for his/her thoughtful comments, which we have addressed in detail below.

General comments

The paper "Cluster analysis of the organic peaks in bulk mass spectra obtained during the 2002 New England Air Quality Study with an Aerodyne aerosol mass spectrometer" by C. Marcolli et al. is an interesting paper that presents a new method for the analysis of the organics mass spectra measured with the widely used Aerodyne Aerosol Mass



Full Screen / Esc

Interactive Discussion

Discussion Paper

EGU

Spectrometer AMS. This topic is well suited for ACP since the general understanding of organic aerosol is still too low but is needed for detailed modelling of aerosol chemistry. While this is an interesting and important paper that deserves to be published, I have some general critical remarks on the method, and I would like to see the authors discussing their method, their findings, and the benefits of their method more critically before publishing in ACP. The authors use a hierarchical cluster algorithm that was developed for clustering mass spectra from single particles, measured with a laser ablation time-of-flight mass spectrometer (PALMS). The algorithm is based on the fact that one has full mass spectrometric information for every single particle that is included in the analysis. This given, it can be compared if different particles look similar in terms of composition or not. The quadrupole AMS used in this study is an aerosol ensemble instrument, which can not analyse the particles one by one (because a quadrupole is scanning ion filter), but needs to average over a certain number of particles in order to obtain a representative chemical composition of the averaged particle ensemble. Thus, I would conclude that comparison of mass spectra can only be done if the spectra have been averaged over the same number of particle or, even better, over the same organic aerosol mass concentration. The authors mention this issue on page 4611, line 20-23, when they find that the diversity is higher when organic mass concentration is low. However, their discussion remains vague: Knowing the absolute particle concentration (very likely a CPC was on board, maybe even an SMPS?), the authors can calculate how many particles were averaged. This result can be related to the measured organic mass concentration and the size distribution, and by that it can be estimated if the averaged aerosol ensemble is representative or not. In general, I would recommend using an averaging algorithm that averages over an aerosol-massweighted time period instead of over a constant time period of 2 minutes.

Response: The success of a cluster analysis does not depend on whether the same mass of aerosol particles has been sampled in every mass spectrum. Particles sampled at high mass loading will have relatively less weight in the cluster analysis than particles sampled at low mass loading, but this does not necessarily influence the rel-

ACPD

6, S1947–S1953, 2006

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

evance of the clusters. It is always possible to calculate how much organic mass is represented by each cluster (as we did in section 3.6). We agree that below a critical value of sampled particles, the diversity depends on counting statistics. Based on the AMS particle counts from the particle time-of-flight (p-ToF) mode in this dataset, at high mass loadings (>10 microgram/m3) more than 500 particles and at low mass loadings (<2 microgram/m3) around 100 particles were sampled per mass spectrum. Note that the actual number of particles sampled for each spectrum is higher than the counts in p-ToF mode because small particles do not have enough mass to be counted individually but signals from several small particles are detected in both the total (counted plus uncounted) p-ToF signal and in the ensemble mass spectra. We do not think that the high diversity at low mass loadings was altogether a statistical effect since there were also time periods where the mass loading and the diversity were low. It is not possible to average over an aerosol-mass-weighted time period without reducing considerably the time resolution since the smallest available time unit is the 1 min mass spectrum collected every two minutes. Instead, it would be possible to weight the mass spectrum representing the averaged cluster by the organic mass that the mass spectra represent. We will address these issues in more detail in the revised manuscript.

Secondly, if one samples with the AMS within one "air mass" (the definition of which is not always clear in atmospheric science), then it is not surprising that the mass spectra, if averaged over a representative amount of particles, look similar. Thus, clustering averaged aerosol composition data within one air mass doesn't make sense to me.

Response: We agree that air masses that change with time are a prerequisite for a successful cluster analysis. We address this requirement more explicitly in section 3.8 of the revised manuscript.

I would like to see if the result is the same if one independently classifies the air masses by other tracers, (e.g., O3 levels, isoprene levels, or backtrajectories), and subsequently averages the mass spectra from each air mass and compares this result with the dominating clusters. From the result given in Figure 11 (comparison with

ACPD

6, S1947–S1953, 2006

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

isoprene and ozone), one would expect to find similar "clusters" or "averaged MS over one air mass".

Response: The proposed type of analysis is somewhat the reverse of what is shown here. We mention in section 3.8 that the formation of clusters and their interpretation could be improved with pre-classification of the air masses. Here we show that without prior knowledge of the air mass the algorithm has uncovered clusters of spectra that are present in similar air masses. However, 75 % of the spectra were placed into category 1, making it likely to be a significant component of different classified air masses. Indeed, we observed this in average spectra of air classified with back trajectories as northwesterly/southwesterly according to Bates et al. (2005). Hence, clustering spectra without prior knowledge of the air mass appears to uncover more distinct groups of spectra.

At the end of section 3.2 the authors make use of the delta analysis (pages 4610/4611 and Figure 6): Does the finding that there are no pronounced differences in the delta patterns imply that categories 1 through 5 are so similar in the MS signatures? Or does is imply that the delta-analysis is not sensitive enough? Is it really important to show Figure 6 or can the finding be described in the text?

Response: The delta analysis has previously been shown to distinguish spectra from laboratory particles generated with various biogenic, alkene, and aromatic precursors (Bahreini et al., 2005). We show here that in this field dataset there were no obvious differences in the delta patterns. This is because categories 1-5 are very similar in their high mass signals and perhaps the delta analysis is not sensitive or not specific enough to use with ambient data. We would like to show this Figure to the community because it might help to assess the potential of the delta analysis method.

Figure 5: Why do the authors use in the cluster algorithm the scalar product to compare mass spectra and now the correlation coefficient? What is the advantage of either of these methods and why is each of them used? Would they yield the same result? What

6, S1947–S1953, 2006

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

are the advantages of using r2 instead of r?

Response: We consulted with Dr. Daniel Murphy (who developed the clustering algorithm that we used) about these questions. Many different metrics can be used in the application of hierarchical clustering analysis. The dot product was used because it is a more natural metric for determining the similarity between vectors. The correlation coefficient, on the other hand, is more useful for determining how well two sets of data are related in a functional form (in this case, a line). Although we did not compare the cluster analysis with the two metrics, we expect the results of the cluster analysis would be similar but not necessarily identical if correlation coefficients were used as the metric instead of dot products. The correlation coefficient, r2, has been previously used to compare AMS spectra (e. g., Zhang et al., 2005), hence we felt it would be useful to have the same metric applied to our comparisons of the cluster spectra with reference spectra. We have now included a table with the dot products of the cluster and reference spectra for comparison to our clustering criteria.

Specific comments:

Page 4606, lines 3 to 15: The dot product (maybe better called "scalar product") of two identical vectors is unity if the vectors are normalized to unit length. When a mass spectrum is normalized to unit intensity, every peak intensity is divided by the total intensity (the sum of the individual intensities). In order to normalize the mass spectrum like a vector to its absolute value (the length of a vector), the peak intensities have to be divided by the square root of the sum of the squares of the individual peak intensities. Only in the latter normalization, the dot product of two identical MS yields unity.

Response: The terms "dot product" and "scalar product" are interchangeable. Since "dot product" has been used in the previous description of the hierarchical clustering algorithm (Murphy et al., 2003), we continued that terminology here. We used the latter type of normalization and make this clearer in the revised manuscript.

Page 4606, line 13: The scalar product ("dot product") should be written as A "dot" B,

6, S1947–S1953, 2006

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

EGU

and not as A x B, since the latter is the vector product.

Response: we correct this in the revised manuscript

Page 4607, line 2, and p. 4608, line 24: Was the nitrate signal at m/z 30 subtracted? Isn't it possible to infer the NO+-signal on m/z 30 due to NH4NO3 from the ratio 46:30?

Response: The nitrate signal at m/z 30 was not subtracted for the analysis. This might be possible if nothing else contributed to m/z 46, however, other organic nitrates and nitro aromatics may have a peak at m/z 46. This has been clarified in the manuscript.

Frequently the term "Pittsburgh Aerosol" is used. I would recommend using "oxidized urban aerosol" since most likely this type of aerosol is not specific for Pittsburgh, e.g.: Page 4610, line 16 Replace "the Pittsburgh oxidized organic aerosol" with "the organics aerosol measured in Pittsburgh in (season, year) by Zhang et al.". Page 4612. line 9: Replace "Pittburgh OOA" with "the oxygenated organic aerosol measured in Pittsburgh by Zhang et al.". Figure 4: Replace "Pittsburgh OOA" by "oxidized urban aerosol"

Response: The Pittsburgh oxidized organic aerosol (OOA) is not the total organic aerosol measured in Pittsburgh but just the oxidized component derived by Zhang's multivariate linear regression (2005a). We replaced this term or reworded the text where appropriate.

Page 4611, line 18-20: "The low diversity...". This sentence is pure speculation and is disproved by the following statement. I would suggest skipping it.

Response: While it is to a certain extent speculative, we still would like to make this point, because the diversity does not only depend on the aerosol mass loading (see response above). We rewrote the text on this point.

p. 4616, line 17: replace "Klinedienst" by "Klinedinst"

Thank you for pointing this out.

ACPD

6, S1947–S1953, 2006

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

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

ACPD

6, S1947–S1953, 2006

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion