Atmospheric
Chemistry
and Physics

Open Access

Discussions

EGU

ACPD

**Interactive
comment**

Interactive comment on "**Chemometric analysis of aerosol mass spectra: exploratory methods to extract and classify anthropogenic aerosol chemotypes**" *by* **Mikko Äijälä et al.**

Anonymous Referee #2

Received and published: 11 October 2016

This manuscript describes an interesting application of cluster analysis for analysis of ambient aerosol data obtained with an Aerosol Mass Spectrometer. In this method, short pollution time periods are analyzed with positive matrix factorization. The factorization yields background and pollution factor mass spectra that are then analyzed with cluster analysis to classify the distinct types of pollution factors that are obtained. As currently written, this manuscript introduces technical details of an analysis method and would seem to be more appropriate for AMT than ACP. It is important for the authors to highlight how this technique provides improved or new insight into study of atmospheric aerosols so that inclusion in ACP is better justified. I recommend publication in ACP after this change and changes suggested below are made:

Main Comments

1) In general, the paper is a little longer than it needs to be because it includes a lot of detailed background information about some topics while not enough information is given about necessary details. For example, page 8, Lines 10-13 only offer no quantitative information about how air pollution events are selected. Can some of the words such as "temporary", "distinct rise", and "unambiguous separation of pollution plume from background" be quantified?. 2) In figure 1, pollution events of varying time scales and multiple apparent pollution peaks are seen. What exactly is the process used to make these selections? What controls the length of the time period that is used as a pollution event? Is wind direction data used for selection? What is the sensitivity of the PMF pollution event solutions to the exact time period range selected around the pollution event? Did you consider as an alternative to this manual plume method to run a traditional PMF analysis on the entire dataset and identify plumes as time periods where the residuals of the PMF analysis are high? 3) It is not clear to me why the pollution event PMF analysis used in this manuscript necessarily provides a more unambiguous separation of "pollution" and "background" than PMF analyses that are performed on the whole dataset. In fact, if the pollution event is simply a result of changes in wind direction that mixes in a different well mixed airmasses, then the PMF factor that is extracted would necessarily be just an average mass spectrum of all the sources present in the polluting airmass. No advantage would have been gained by this method to allow separation of individual sources and this would seem to be a weakness of this method. This aspect is not discussed in the manuscript. 4) The manuscript refers to ambiguities in PMF analysis as a weakness and implies that this analysis somehow solves or provides a better solution to this problem of ambiguity. In fact, the manuscript clearly states the difficulty of separating the various primary aerosol sources. One of the advantages of the traditional method of doing PMF or ME-2 on the entire dataset in this context could be the fact that it can exploit differences in temporal profiles of primary sources (i.e. different diurnal cycles) and also exploit the fact that source mass spectra are similar to allow for separation of multiple primary

sources within a well mixed pollution event (i.e. a event such as that mentioned in comment 3 above). A comparison between the classification results and a traditional PMF of the entire dataset would have been a good way to address this and to highlight similarities and differences in results. The manuscript should more clearly state discuss the advantages/disadvantages of using this method compared to PMF. 5) The strongest part of this manuscript is the application and interpretation of the various clustering metrics to understand similarity and differences between the cluster spectra. It may be useful to highlight more strongly how these metrics could be applied to spectra obtained with typical PMF/ME-2 analysis. Would use of the cluster analysis metrics to reference spectra and PMF solutions provide a means of automating classification in PMF analysis? Also an intriguing part of this that could be discussed in more detail is the possibility to use the cluster analysis to define a-values and reference spectra for ME-2.

Interactive comment on Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-632, 2016.