

Interactive comment on “Using non-negative matrix factorization for the identification of daily patterns of particulate air pollution in Beijing during 2004–2008” by A. Thiem et al.

P. Pakbin (Referee)

pakbin@usc.edu

Received and published: 16 August 2012

The manuscript tries to use non-negative matrix factorization in order to interpret the data taken in one monitoring site in Beijing, China during 2004 – 2008. The main objective of the manuscript as stated in the title and the abstract is to “extract characteristic diurnal air pollution patterns of particle number and volume size distributions for the study period”. There is useful information in the study and it is worthy of publication after major revision. However, the authors emphasize the methodology (i.e., NMF) over a more in depth analysis of the data. It is not especially useful to “discover” well-documented associations between number and volume size distributions and sources,

C5808

as determined by actual source tests, several of which are cited in the paper.

This reviewer believes that different data processing methodologies could have been more helpful for interpreting this high temporal resolution data set which covers number and volume size distributions over a 4 year period at only one monitoring site. Is it possible that sources in a given sampling site have changed from year to year within these 4 years? The weekly and seasonal variations in PM sources are identifiable, but unless there were unusual circumstances, I don't see how source in general can change drastically from year to year in a 4 year period. Basically other than episodic events like wildfires, etc., the yearly variations in PM sources at a given site must be negligible. Thus analysis of the data from 1 year of this study could yield similar results. On a separate note, authors claim in the abstract (P. 12017, L. 8) that the number concentrations have been measured in high time- and space-resolution; while the data is obtained at a single site, so why do they claim that this is also a high space resolution data set? NMF would have been most useful if there were more than one sampling site in a one year period of time. Ideally these sampling sites must have been chosen to represent different areas of the city with different PM sources (i.e. one site affect by local traffic, one relatively out of city to reflect background PM, one in agricultural areas to reflect increasing effect of windblown dust, etc.).

The data could have been interpreted without using NMF, and by comparing the diurnal variations of particle number and volume size distributions in different seasons, and on days with unusual events. Using NMF, the authors' interpretation of the factors is purely speculative. The absence of meteorology data is one of the major caveats of this study. The authors talk about variations in traffic density and changing car fleet, without using this information in interpretation of NMF results. If such data was available, it would have been significantly useful for identifying vehicular sources, based on diurnal variations in traffic density (and changing car fleet as the authors mention in the abstract) and comparing it to the diurnal variations in number and volume size distributions. The discussions are too general, and more emphasize is on methodology than the quan-

C5809

titative results. The authors should at least present some specific PM concentration averages, and highlight some of seasonal variations in number and volume concentrations. For example, more discussion of the highest PM concentrations (for each size fraction) would be useful. The dates on which these occurred, wind speeds/directions, fractions of size fractioned concentrations to total PM concentrations, etc. would provide far more insight into the causes of these high levels that is currently available from NMF.

Moreover using NMF, has resulted in 'mixed factors', also pointed out by the other reviewer, where the first factor, NMF-N1, is affected by three different sources. If the authors want to use this method, it is advisable to do the analyses for each season separately. This way these mixed factors may appear differently in each season, enabling authors to gain more insight in the possible sources of their observations. Most of the discussion focused on the methodology can be moved to supplemental material in order to include more comprehensive discussion on the single events, presenting quantitative data and discuss in more detail the seasonal variations in number and volume size distributions.

Interactive comment on Atmos. Chem. Phys. Discuss., 12, 13015, 2012.