

## ***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.***

**H. Jorquera (Referee)**

zorquera@ing.puc.cl

Received and published: 1 August 2012

### General Comments.

The manuscript intends to find new insights on the sources and processes that explain ambient levels of PM<sub>10</sub> in a monitoring site at Beijing, China, by looking at PSD data taken with high temporal resolution over a 4 year period. The topic is certainly within ACP scope and new results would be expected out of such analysis. However the manuscript lacks a thorough discussion of results so it needs a major revision before it can be judged suitable for publication on ACP.

I have found the following deficiencies that must be addressed:

C5264

1) Given the large number of observations (864 daily profiles), I would have expected that the factors extracted out of the NMF methodology should have had a better physical meaning than the ones described in section 3.1 and 3.2. For instance NMF-N1 is a mixture of three sources. . . this is a poor result for a data-rich-driven factor analysis where usually separate traffic, industrial and residential sources are resolved when PSD are used (Gu et al., 2011, using PMF methodology were able to resolve local and long range dust, fresh and aged traffic sources in a medium size city in Germany). I see no ‘soil dust’ source identified despite Beijing being subject to PM loads from dust storms. The authors should strive to identify those sources in the context of analysis of the NMF-V factors which is not carried out in depth, despite evidence like Figure 5. For instance, since NMF-V factors are reflecting the sources’ mass contributions they could be compared with results from receptor modeling studies carried out at Beijing. I am also concerned with the apparent discontinuity of PSD temporal profiles at midnight which has not been observed in similar studies (Gu et al, 2011, cited in the MS). Is it possible that such a gap is due to some artifact in NMF implementation? Is it possible to add a constraint to NMF to enforce an agreement at midnight?

2) The analysis of time-varying source contributions is superficial. I would recommend looking at workday-weekend and seasonal variation to confirm source identification and discriminate among traffic, industrial, long range and residential sources (Yue et al, 2008, cited in the MS). Since in the Abstract section’s first sentence: “increasing traffic density and a changing car fleet” authors are suggesting potential trends in traffic impacts at Beijing, they should either remove that sentence from the Abstract or conduct a trend analysis for the traffic sources, after they are better resolved within the NMF methodology.

3) The discussion on wind trajectories of potential sources reaching the monitor site should be expanded with more quantitative meteorological data to confirm source identification and potential location. I understand that Beijing and adjacent cities is part of a cluster with ~ 180 million inhabitants so a proper source attribution is a challenging

C5265

issue and if the manuscript approach can provide new insights this would add in more value for regulatory purposes, which is a key goal stated in the Abstract section.

4) The manuscript is poorly edited. The introduction should be focused on how different processes and emission sources lead to PSD with different size modes (in number and volume) because those are the kind of results that will be discussed. The health effects literature review should be reduced to a minimum. There are many vague, incomplete sentences scattered across the manuscript; I'd recommend hiring an English editor to improve on this.

Specific Comments.

Abstract

P13017, first sentence: if trends in traffic emissions are hypothesized, they should be discussed in the respective section. Otherwise rephrase the sentence. (See also comments of the Conclusions section below).

P13017, second sentence (lines 4-7): if one goal is to identify sources and their locations for a better regulatory policy, then this should be answered in the manuscript conclusions.

P13017 third paragraph (lines 17-28): the sources (factors) obtained with NMF should be clearly identified and their mean contribution to number (or volume) quoted. The NMF-V factors should be compared with chemical composition receptor model results for Beijing.

Introduction

P13018-13019: reduce discussion on PM-associated health effects and focus on the sources and processes that lead to distinctive PSD in ambient particulate matter. Summarize previous findings for Beijing focusing on source apportionment and PSD measurements, to help following discussion of results.

C5266

P13020: be more succinct on discussing factor analysis approaches, leaving details for section 2.3.

P13022-23: A map of Beijing showing the typical wind patterns would be helpful for source identification (by season for instance) of city and regional scale sources.

P13024: Explain why you chose a maximum number of iterations set to 1000. What were the typical numbers of iterations required for convergence in NMF? More explanation is needed to understand the stopping criteria used.

P13026 lines 18-28. The choice of number of factors ( $r$ ) was made on inspection of resulting PSD profiles to avoid combination of sources (factors). However the results for NMF-N1 are identified as a mixture of three sources. Reconcile this.

Page 13027 lines 1-11. It is unclear which was the final weighing of data for PSD (if any), especially when considering volume data. This should be stated for the results quoted in the manuscript.

Page 13027 lines 13-16. It seems curious that NMF-V factors could not be physically interpreted. What about correlating them with available pollutant measurements? What about the time dependency of pattern amplitude? Workday - weekend differences?

2.3.3 Relevance of NMF factors.

P13027 lines 26-27. Why was variance assumed to be constant for all factors? Was this assumption verified ex-post? Please justify.

P13028 I don't understand the scaling using equations (4) and (7) after NMF has converged. It seems more convenient to normalize  $W$  to 1.0 (like any probability density function) and leave the amplitudes  $H$  pick up the time-varying source contribution in units of particle number (or volume) concentration. In this way both factors can be analyzed on their own. The use of equations (4) and (7) is non-standard in factor analysis so comparisons with other studies (like traditional receptor modeling versus the NMF-V sources) is not direct.

C5267

### 3. Results and Discussion

P13029 lines 9 –12. If the physical interpretation of the W factors depends on the time-varying H factors, why these were not extensively used to identify the sources/processes that correspond to each W found? What is the meaning of the “particle burden, measured in terms of the median value of the corresponding time series”? What are the units of H factors?

P13029 lines 15 – 19. If NMF-N1 is a “background pattern” that provides “a kind of basic load” then why is the time-varying contribution H zero at so many days? (Figure 4b) Could Beijing ever “shut down” their basic PM sources/processes for a whole day? I doubt it.

P13030 Provide further support for the identification of factors NMF-N2 and NMF-N3. NMF-N2 should be analyzed using the time-varying contribution H (comparing seasonal behavior for instance) and NMF-N3 with the help of wind trajectories to identify likely source locations. All time-varying contributions should be presented.

P13032 lines 17 –19 authors state that they could not identify NMF-V factors for they only change in the coarse mode. Have they tried looking at the temporal amplitudes H's for seasonal, weekly patterns? How about looking for correlations of the temporal amplitudes H's with the ones computed for the NMF-N factors? What about looking at correlations of NMF-V temporal amplitudes with gaseous or particulate pollutants at nearby monitoring stations? Last, but not the least, perhaps the results do not support so many NMF-V factors after all. I do think more analysis is required here to achieve a credible explanation.

#### 3.4 Categorization of NMF patterns

P13035 lines 1-5. The identification of NMF-N4 and NMF-N5 as secondary aerosols requires further support. Are inorganic (sulfates, nitrates) and organic (SOA) aerosols included in those two factors? If so, are they mixed in both factors or separately re-

C5268

solved? Is it possible to correlate those factor's time amplitudes H with ambient measurements of nitrates, sulfates, OC?

### 4. Conclusions

P13036 lines 18-19. This sentence should be repeated in the abstract as the main goal of this work, being part of the first paragraph in the Abstract.

P13036 lines 25-26 and beginning of P13037. I do not think that the authors have made the case that they can fully combine NMF-N and NMF-V factors to identify all PM sources/processes at Beijing, especially for coarse mode particles. Certainly dust storms alone cannot explain all coarse mode data.

P13037 lines 17-18 Time varying contributions H were used only in their mean values per season and long term means. Correlations among time-varying contributions and local pollutant concentrations and among NMF-N/V factors were not studied at all, yet they could provide support for further source identification.

---

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

C5269