

Interactive comment on “Interpretation of organic components from positive matrix factorization of aerosol mass spectrometric data” by I. M. Ulbrich et al.

V. Lanz

lanzv@ethz.ch

Received and published: 30 April 2008

As mentioned along with my first comments, this manuscript represents an important contribution to the puzzle how organic Q-AMS spectral data can be related to OA sources and secondary OA components. I have a few general and some more specific comments with a focus on the real data case.

General comments:

A) Verification of factor interpretations

Abstract, p. 6731, lines 4-5: “It is critical to use correlations between factor time series

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



and external measurement time series to support factor interpretations“.

According to my experience, this correlation might not always be enough to ensure an accurate interpretation of PMF retrieved factors: the correlation between factor scores and the auxilliary data series can be due to several reasons (e.g., as both reflect human activities, meteorology etc.) More options can be considered, such as the comparison measured vs. modelled emission ratios (e.g., $POA_{modelled}/NO_{x,meas.}$ vs. $POA_{meas.}/NO_{x,meas.}$), diurnal, weekly or seasonal cycles (boxplots of the scores) etc. as advocated earlier (Lanz et al., 2007, 2008a, 2008b.). This could be mentioned.

B) Three-factorial solution of the real Pittsburgh data: OOA-II vs. HOA

From an atmospheric science view, I believe an important section of this manuscript is the “3-factorial solution“ (best solution as selected by the authors) within chapter 3.1 (Real Pittsburgh data). The interpretation of these 3 three factors (OOA-I, OOA-II, and HOA) is very well-founded and could additionally be supported, e.g. the interpretation of OOA-II and HOA:

OOA-II - as reported in this study - is dominated by m/z's (18), 43, 44, and 57 (Fig. 5, panel a). While m/z 43 may represent both oxidized ($C_2H_3O^+$) or hydrocarbon-like ($C_3H_7^+$) aerosol components originating from different sources, mass fragment 57 has been suggested as a marker for primary combustive (diesel) sources ($C_4H_9^+$) in many previous studies. Mass fragment 44 represents mostly non-gaseous CO_2^+ , a signal that is likely due to highly oxidized substances such as di- and polycarboxylic acids. Primary vehicle exhaust can be the principal source of the di-carboxylic acids in urban areas (Yao et al., 2004), potentially causing the enhanced m/z 44 signal here as well. At first glance, the time series of OOA-II further seems to be somewhat correlated with the TS of HOA (given the more or less simultaneous peaks for HOA and OOA-II, e.g. on 9/11, 9/13, 9/14, 9/19 etc.). From Fig. 5, Panel b) I would expect that OOA-II is also correlated to NO_x and CO. On the other hand, m/z 57 might also represent $C_3H_5O^+$ and the correlation of HOA and OOA-II may perfectly reflect the rapid partitioning of semivolatile

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



OVOCs (down-mixed during the early morning hours?) into the freshly emitted HOA particles. No doubt, the results reported here will stimulate further discussions about the chemistry and temporal behaviour of differently aged SOA.

It might be fruitful in this context to calculate the diurnal boxplots (as absolute and relative contributions) of the three factors' scores (OOA-I, OOA-II, and HOA) and show or describe them. This could also give an additional answer to the gasoline vs. diesel particle question (two birds with one stone). Factor interpretations could be corroborated by diurnal boxplots: e.g., for the Zurich data the traffic factor (extracted from organic aerosol as well as from gas-phase data) showed a different weekend vs. weekday pattern with respect to its daily cycle (Lanz et al., 2007 and 2008b).

C) Split vs. coerced sources/spectra

One drawback of using the chi square-metric underlying PMF2 is that it places no restriction on the complexity of the model, making it potentially prone to overfitting. The authors input 2 and 3 profiles, respectively, in order to generate synthetic data. Within the applied PMF model, they increased the number of assumed source profiles to higher values, $p > 2$ and $p > 3$, respectively, and describe what happens. Then, the authors rightly found that the split factors may correlate well with real profiles.

On the other hand, it can be expected that in reality the situation is vice versa (and even more frequent): Much more sources (with distinct spectral fingerprints) actually have an influence on the OA at the receptor-site than can be/are specified within receptor models; widely different organics from various source types with varying profiles are present in ambient air (wood burning and biomass combustion particles generated at different conditions, diesel and gasoline exhaust, primary aged particles, secondary and differently aged particles etc.). E.g. for wood/biomass burning, Schneider et al. (2006) reported widely different AMS spectral signatures. Passant (2002) has collected the profiles of more than hundred volatile hydrocarbon sources . . . - there is no doubt: this certainly yields composite source profiles and various artefacts in bilinear receptor-

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

model outputs due to coercing real sources (with distinct spectra). Further, it is possible that, e.g., at $p=3$ factors the profile of an additional 4th (real) source is (mathematically) approximated by linear combinations of the first three factors extracted by PMF (not to be confused with the “mixing artefact”).

Future studies may detail this latter situation. In the current abstract, discussion and conclusion part, this situation could possibly be mentioned as well.

D) Untypical vs. outlying spectra

It is possible that a very small fraction of samples in the real Pittsburgh dataset represents pure/absent sources (or distinct spectra) (geometrically spoken – see papers by R.C. Henry – these samples would represent the vertices/edge-points of a solution simplex spanned by the source vectors). These samples in question would then represent unusual but crucial data points (in order to recover real source profiles) and should, therefore, rather be up-weighted than down-weighted. By using the robust mode in PMF2, this source information may be lost (depending on the data structure) . . . possibly causing the loss of retrievable/interpretable source profiles and/or influencing their shape. Did the authors consider this possibility? In other words: are there any differences in the results of the robust and non-robust mode pointing to this possibility?

Specific comments (continued, comp. ‘Comments (Part I)‘)

4) p. 6731, lines 25-27: at this instance, it could be instructive to mention that ¹⁴C analyses and receptor modelling of organic AMS spectral data were combined (Lanz et al., 2008a)

5) p. 6731, line 27: the time-resolution of radiocarbon (¹⁴C) analysis is in the range of several hours rather than several days.

7) p. 6733, lines 25-29: It would be more accurate to state: “Lanz et al. (2008) *applied* a hybrid receptor model [. . .] *specified* by the Multilinear Engine“.

8) p. 6734, line 1, OOA-nomenclature: the ratio of m/z 44-to-m/z 43 in OOA found in

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

winter (Lanz et al., 2008a) was 2:1, which is in between OOA-I (m/z 44: m/z 43 ~ 3:1) and OOA-II (m/z 44: m/z 43 ~ 0:1) found in Zurich summer (Lanz et al., 2007). Therefore, we labelled it OOA rather than OOA-I.

9) p. 6734, line 2: “more advanced“. I did not fully comprehend what is “more advanced“ than what? With respect to what criterion?

10) p. 6735, line 14: what is meant by “internal correlation“ here? The correlation of two species within the PMF data matrix? This would also be the case for certain chromatographic data (along with the high precision etc.).

11) p. 6739, lines 7-8: at this instance, it might be instructive to refer to already published ME-2 based work on organic AMS spectra (Lanz et al., 2008a) or work including organic spectral tracers (Buset et al., 2006).

12) p. 6740, line 26: what are “ ‘solid body‘ geometric rotations“? Please explain or rewrite.

13) p. 6741, line 15: I do not understand what is meant by “is still satisfied will little additional error“

14) p. 6742, section 2.2.2 (“Singular value decomposition“): At this point in the text, it is not clear why this subsubchapter was introduced in the manuscript. What about principal component analysis (PCA) etc.? Skip this chapter.

15) p. 6743, line 18: “the inverse“ has also a mathematical meaning and might be somewhat misleading here.

16) p. 6746, line 19 (25): I would replace “statistics of correlation“ by “measures of correlation“ as a matter of taste.

17) p. 6749, line 1: “reduced aerosols such as meat cooking“; seed oil cooking and meat charbroiling aerosols or meat smoke are not fully reduced, but contain oxidized species as well (Schauer et al., 1999a and 1999b, Nolte et al., 1999). It would therefore

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

rather surprising if meat cooking (as AMS determined by Mohr et al., in prep.) had the same spectral fingerprint as diesel combustion particles at all cooking conditions. This could be discussed a little more.

18) p. 6749, lines 14-16: “the . . . spectrum lies 28 degrees out of the . . . plane“. I am not sure what is meant at this instance? Is the authors' argument a geometrical one (such as: the solution space is three-dimensional)? If this is the case, they could illustrate this (e.g., by a set of suited projections of the samples onto planes - in analogy to Henry, 2003). In any case, it should be specified what is meant by “28 degrees“ at this instance.

19) p. 6751, lines 14-15 (and the corresponding passage in the conclusions): I do not understand how the second part of the sentence is related to the first one and our study (Lanz et al., 2007) (please consider specific comment 3). Please explain or write this sentences anew.

20) p. 6766, line 1: Concerning the identification of OOA-II, it might be instructive to refer to your AAAR poster and/or to Lanz et al., 2007.

21) p. 6766, line 6: one could add something like “. . . or it simply means that distinct profiles of additional sources (not retrieved by PMF) can be approximated by linear combinations of the PMF resolved profiles“ (not to be confused by the “mixing artefact“).

22) p. 6766, lines 10-12: structure in the residuals may arise from the ageing, decay etc. of all components (OOA-I, OOA-II, HOA), may arise from minor sources etc.

References

Buset et al., Atmos. Environ, 40, S482-S499, Suppl. 2, 2006.

Henry, Chemom. Intell. Lab. Syst. 65, 179–189, 2003.

Lanz et al., Atmos. Chem. Phys., 7, 1503–1522, 2007.

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

Lanz et al., Environ. Sci. Technol. 42, 214–220, 2008a.

Lanz et al., Atmos. Chem. Phys., in press, 2008b (ACPD: 8, 907–955)

Nolte et al., Environ. Sci. Technol., 33, 3313–3316, 1999.

Passant, Speciation of UK emissions of non-methane volatile organic compounds
AEAT/ENV/R/0545 Issue 1, 2002.

Schauer et al., Environ. Sci. Technol., 33, 1566–1577, 1999a.

Schauer et al., Environ. Sci. Technol., 36, 567-575, 1999b.

Schneider et al., Int. J. Mass. Spec., 258, 37–49, 2006.

Wittig and Allen, Atmos. Environ., 42, 1319-1337, 2008.

Yao et al., Atmos. Environ., 38, 7, 963-970, 2004.

Interactive comment on Atmos. Chem. Phys. Discuss., 8, 6729, 2008.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

