

**Review of Comparison of dimension reduction techniques in the analysis of mass spectrometry data
Atmospheric measurement techniques, AMT-2019-404
Isokääntä et al.**

Summary:

This work accomplishes a cross-comparison of several data-reduction analysis techniques. The analysis is performed on a chamber experiment. Car exhaust was directly sampled into an environmental chamber. A-pinene was also added to the chamber. The mixed car exhaust/pinene was then aged via OH-initiated photooxidation. Two instruments were used, a PTR-ToF MS and an AMS. Data were then analyzed using principal component analysis, positive matrix factorization, exploratory factor analysis, clustering, and non-negative matrix factorization. The resulting simplifications are compared in terms of the ability to reconstruct the original data set (residual), number of factors/groups required to explain data variability, time-series behavior of the factors, and chemical composition of the factors. The authors find that the preferred number of factors is roughly similar regardless of technique. Some factors (or groupings) are generally consistent (in terms of time-series behavior and composition) regardless of technique. Some techniques, particularly PCA and EFA were found to be difficult to interpret and not as useful. The clustering method was not useful for AMS data.

Major comments:

The manuscript is generally well-written and organized. The sampling techniques and experimental conditions are appropriate.

This work presents a clear, helpful, and timely cross-comparison of results between different analysis methods. Because modern analytical instrumentation produces large datasets, computer assisted analysis is unavoidable. A major question in the field is if the results are robust, reproducible, etc. That the same major patterns and groupings appear (with minor differences) regardless of which technique is used is an important result that provides legitimacy to the vast amount of work currently being done with PMF and other techniques in atmospheric chemistry. Therefore I recommend this paper for publication.

My major questions include the following.

- 1) The pre-light mixing period is a substantial fraction of the whole experiment- is this interesting? The inclusion of this time period seems to have significant impact on the algorithm results.
- 2) Mostly this paper seems to show that various different algorithms distinguish similarly between primary and secondary VOCs. Is this a useful reduction of chamber data, to group all secondary VOCs into one or two blocs? From Table 1 it seems that the two matrix factorization techniques result in perhaps three oxidized factors, but this is hardly discussed in the text. It is not clear to me if the three oxidized factors are consistent between the two techniques. It would be helpful to have more interpretation of these factors, especially if it were supported by a more detailed connection to the chemistry of the system.
NMF also seems to result in some mixing of primary and secondary emissions (FN4) which is probably unphysical.
- 3) The way the mass spectra (chemical composition of groupings) are presented is very difficult to interpret and compare. From Table 1 the authors have assigned a consistent identity to factors resulting from each technique. Can you show a direct comparison of "Factor 2" for example, perhaps by plotting one mass spectra against the other so that it is easy to see which VOCs are similarly enhanced, and which may be different?

Specific comments:

Page 3, line 3-4: I disagree with this statement, “PMF was originally developed for field measurement data sets where real changes in factors are expected to be much slower than e.g. the noise in the data.” In the ambient environment, VOC composition and concentrations can actually change very quickly (on the order of seconds), especially when plumes of highly-concentrated primary emissions are intercepted. The abrupt changes in conditions during a chamber experiment therefore do not present a special challenge to PMF, compared to ambient measurements.

Page 3, lines 26-29: A few more details are needed here (instead of in the supplement):

What was the typical concentration of total VOC (or of a few key VOC e.g. aromatics)?

What concentration of α -pinene was added?

What was the VOC-NO_x ratio, and how was it adjusted?

Why were these specific concentrations of vehicle exhaust VOC and α -pinene chosen?

Were vehicle exhaust and pinene added just at the beginning of the experiment, or were they continuously injected?

Was the chamber continuously refilled (and with clean air or with fresh emissions?) to replace air taken by the mass spectrometers, or did the volume of the chamber decrease over time?

Section 3.1.2: EFA seems very similar to PMF. Could you please explain the major relevant difference(s) between EFA and PMF, and how they would affect the resulting dimensionality reduction?

Page 7 line 16 (and elsewhere): At multiple points in the manuscript it is mentioned that the rapid changes associated with lights-on cause problems when implementing the various dimensionality reduction techniques. Would it make more sense to exclude data prior to $t=0$? Was there a reason this was not done?

Page 10 lines 28-31: Since Figure 1 relies on a comparison of BIC and SRMR, it would be helpful here to provide more detail on how these two metrics are calculated, what the relevant differences are, and why one may be preferred over the other. What was the purpose of calculating both metrics and why were these particular metrics chosen?

Equations 11 and 12: There are several errors in these equations which are likely a copying error from Brunet et al. 2004. The authors should check that the actual implementation was done correctly. The equations should read:

$$H_{au} \leftarrow H_{au} \frac{\sum_i W_{ia} X_{iu} / (WH)_{iu}}{\sum_k W_{ka}}$$
$$W_{ia} \leftarrow W_{ia} \frac{\sum_u H_{au} X_{iu} / (WH)_{iu}}{\sum_v H_{av}}$$

I also suggest here to use “ k ” as the row index for \mathbf{W} (in the denominator term), to avoid confusion with “ p ” being the factor rank, and for consistency with Lee and Seung, 2001.

Page 11 lines 29-33: Given that the update functions (11) and (12) are derived from the divergence cost function $D(X||WH)$ (Lee and Seung, 2001, Eq. 3), I suggest that this cost function is monitored as a function of p , analogously to $Q/Q_{exp}(p)$ for PMF. The termination condition for NMF wasn’t described in Section 3.1.4, but presumably it is not dependent on p ; if this is the case then the divergence of the end solution can be compared for each value of p .

The residual sum of squares is not an appropriate metric, as this was not the cost function used for the NMF implementation.

Page 12 line 2: What is meant by “not achieved only by change?”

Page 13 line 5: Why not compare the absolute value of the residual?

Pages 14 and 15: For other researchers which would like to use this paper as a guide, it would be helpful to indicate the range of values that are acceptable. For example Page 14 line 4-5, what value of residual would be considered not acceptable? Page 14 line 32, what is the Kaiser limit and what range of values are considered “close”? Page 15 line 11 are these considered large or small residuals?

Page 14 line 1 and page 15 line 2: Can you show please how it is determined that the additional component is not a new component with different properties but rather a mixture of previous components?

Page 17 Lines 16-17, 23-30: The signal following error is essentially introducing a smoothness constraint, which doesn't seem appropriate given that you know there are sharp changes due to experimental conditions. Is it recalculated for each data resampling? Why not resample the error along with the data? Is it possible to split the data into time periods whose start and stop are defined by sample injection, lights on, etc. so that the running standard deviation does not include these sharp changes?

Additionally, lines 26-27: Ambient data often has fast changes that are due to real variability. This is one of the reasons why fast online techniques such as PTR-MS are used for ambient measurements, because they allow the observation of these changes.

Page 22 line 20: This indicates that the error estimation should be revised, or that these compounds should be downweighted. PMF and NMF are extremely similar techniques with the crucial difference being only the inclusion of an error matrix, so it does not seem likely that the difference in performance is due to the size of the dataset. The extremely small values of NMF residuals also seem suspect. The authors should check that residuals for PMF and NMF were calculated in exactly the same way so as to enable the direct quantitative comparison.

Page 24 line 13-14: Is this correct? I read this paper as well. I thought they had PTRMS and AMS. Additionally, PTR-TOF is a subset of TOF-CIMS, or?

Figures 3,4: In the plot caption or legend it would be helpful to have a brief description of the interpretation of each factor, e.g. “pinene”, “car exhaust”, “background”. Additionally the display in plot (b) of factor contribution as a function of m/z doesn't add much to the paper; I wouldn't expect the factor contribution to depend on m/z in any particularly meaningful way. Since (I believe) your PTRMS has multiple peaks resolved at each nominal mass, showing a unit-mass stick spectrum here is also not especially meaningful. If you want to show mass spectra I strongly suggest to break panel b into 4 separate spectra, one for each factor, so that they can be examined separately.

Figure 9: Where does isoprene come from in this experiment? Is this more likely to be a hydrocarbon from vehicle exhaust? Cycloalkanes in fossil fuel are known to create PTR ions at $C_5H_9^+$, see e.g. Yuan et al. Chem Rev. 2017 doi.org/10.1021/acs.chemrev.7b00325.

Figure 13: The factor time series for the most part do not look realistic. What physical process could lead to the non-smooth behavior and multiple maxima?

Minor/Technical corrections:

Page 2 line 22, "alike" -> "like"

Page 3 line 22, "was" -> "were"

Page 13 line 24, "described in section 0" -> "described in section 3"

Page 15 line 19, "gab" -> "gap"

Page 19 line 27, "much" -> "many"