

## ***Interactive comment on “A Novel Approach for Simple Statistical Analysis of High-Resolution Mass Spectra” by Yanjun Zhang et al.***

**Anonymous Referee #2**

Received and published: 17 April 2019

This paper introduces a new method of interpreting ambient mass spectrometer data, binPMF. This is an extension of existing PMF techniques to sub-unity resolution mass spectra without the usual peak fitting stage before this. The advantage of this approach is that it can utilise the extra mass spectrometric information without having to specify the peaks that are expected to be present. I can foresee a number of applications for this technique; it could conceivably be used deliver more accurate factorisations than UMR-PMF, but I strongly suspect that it may prove more useful in identifying peaks to use for HR-PMF. However, time will tell on that.

Presented as a proof-of-concept, this is likely to stimulate activity and further development within the community of users of ARI/Tofwerk instruments, but I would anticipate that this may have applications beyond this. The paper is certainly relevant to AMT and

C1

is very well written and presented. I have a few reservations (see below) but these are minor, so subject to these I recommend publication.

Comments:

Generally, the authors are very bullish about the capabilities of the technique, but I can foresee a number of fundamental limitations. In the interests of properly exploring this method on a conceptual level, I would recommend that the authors discuss these so as to create realistic expectations of what this is capable of. The issues I can think of are as follows, but there may be others:

\* Peak shape - It must be explicitly stated that this technique assumes that the peak shape model is consistent throughout a dataset. While this should remain constant if the instrument is working properly, if it drifts somehow, then this will likely cause unpredictable behaviour here.

\* Aliasing - the act of binning the mass spectra through over- followed by under-sampling may introduce artificial smoothing of the data. This is unlikely to be an issue for peaks at low  $m/z$  ratios where the fundamental resolution of the instrument is high, but where the resolution of the mass spectrometer starts to become comparable to the target bin width, I imagine this could be an issue. This would be unlikely to cause problems if the mass calibration of the spectrometer were constant, but as this is known to subtly drift over time, this means that any aliasing artefacts could (in theory) be variable with time, even if the changes in calibration were properly accounted for, which in turn could create artificial factors in the dataset. This should be discussed, if only conceptually.

\* Complexity - the analysis may not be able to adequately factorise systems where there is a large number of degrees of freedom in the chemistry, e.g. studying SOA formation in a chamber. This is a fundamental limitation of PMF and applies equally to the other established methods as well. However, I suppose this technique could still be of use in peak identification, even if it can't explain all the variance.

C2

\* Covariance - as with all PMF, the ability to separate components is contingent on them showing different trends. If (hypothetically) two adjacent peaks were to covary, then this technique would fail to separate them.

Line 196: If the noise is based on signal-free regions of the mass spectrum, would this not be underestimated because of the thresholding applied by the data acquisition system?

Line 535: I disagree that mass calibration could be accounted for by an error term; according to the PMF data model, the errors are supposed to be random and independent of one another, however a shift in mass calibration would cause deviations that are dependent on adjacent points.

---

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2019-59, 2019.