Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-312-RC2, 2017

Geoscientific Model Development Discussions

Interactive comment on "STRAPS v1.0: Evaluating a methodology for predicting electron impact ionisation mass spectra for the aerosol mass spectrometer" *by* David O. Topping et al.

Anonymous Referee #2

Received and published: 18 February 2017

General Comments

It is acknowledged that techniques are required to help illuminate the highly complex chemical and physical properties and processes associated with SOA formation, evolution and composition. One such considerably powerful instrumental technique is Aerosol Mass Spectrometry (AMS), now used in many areas of atmospheric, climate and air quality science. As stated by the authors, AMS is able to capture all mass but unfortunately is unable to provide the required high level of speciation; thus generally a cumbersome ensemble of techniques is required to fully test measurements with state-of-the-art models. A method to help improve enhanced chemical speciation of AMS spectral composition, and hence empower model vs. measurement comparisons

to enhance fundamental understanding, would be welcomed.

In this work the authors present a proof of concept study around the use of trained supervisionary regression methods to predict the composition of AMS spectra of organic aerosol components. The authors present a well thought out series of sequential results comparing model outputs using various supervised learning methods with experimental data for single component spectra and more complex chemically driven spectra. The authors provide a direct link to the code developed.

This work is very novel and has importance for model and mechanism development in the laboratory and in ambient air (e.g. model vs. measurement comparison for elucidation of fundamental processes); it also has the potential to add to the widely used Positive Matrix Factorisation (PMF) technique and to be employed by other analytical methods. As such it is recommended that this manuscript be published in Geoscientific Model Development subject to the authors addressing the following Specific and Technical Comments.

Specific Comments

P2, Line 2: Could the authors state the rationale for choosing to assess performance simply with cosine angles?

P2, Line 15: could the authors suggest which other analytical techniques could potentially benefit from the method and include appropriate references?

P4, Section 2: It is unclear from this text whether the fingerprint is in fact a single mass spectrum of m/z vs abundance. Can this be section be rephrased slightly with a more explicit statement, please? Further to this, should each column refer to a given m/z, is this enough information, or are other concomitant spectral features required to validate the presence of a certain functional group? If each single column "key" is able to contain sufficient information, can the authors clarify and appropriately state this in the text (further to references e.g. Ulbrich et al.)?

P6, Section 3.1: Can the authors comment on the sensitivity of the technique to the various functional groups listed on line 30, for example? Are there any inherent instrumental sensitivity issues with certain functional groups that might limit the effectiveness of the technique at a top level?

Owing to composition dependence acknowledged by the authors, it would be nice to see additional data c.f. Figures 5 and 6, for other single precursors. Are these data available?

P7, Lines 30 – 33: Regarding the statement – "This reflects sensitivity to information used in the training process and how similarity between performances should be taken with caution in prescribing which method to take forward", as this represents a limitation, could the authors expand their discussion slightly, i.e. potential magnitude of uncertainty associated with inaccurate method prescription? Further, could the authors clarify the sensitivity of the technique to user required experience and expertise?

P8, Lines 18 – 20: When the authors refer to addition of data from mixed systems, are they referring to an ensemble photo-oxidation study, or simply an inert multicomponent mixture? Did the authors consider a test intermediate in complexity, e.g. the obvious intermediate between a single compound mass spectrum and a chamber photo-oxidation experiment would be an analysis of a mixture of 2-3 compounds, without the complex oxidative chemistry. Was this considered?

Regarding the AMS data employed (e.g. Figure 7): How were these data treated? Were they experiment averaged, summed, normalised? Despite the reference to Alfarra et al., 2013, it may be useful to briefly state this on introduction of the experimental data in order to provide context.

Please check reference formatting throughout, e.g. spaces between text and parentheses and improper use of chronological ordering of multiple citations.

Technical Corrections

P2, Line 21: Please add more indicative primary source references; this paper is rather specific

P2, Line 30: Reference repeated

P3, Line 14: "...air and in THE laboratory..."

P4, Line 26: "now" rather than "new"

P4, Line 28: "than" rather than 'that"?

P5, Line 10: "than" rather than "that"

P5, Line 30: Full-stop missing after "3.2"

P6, Lines 14 – 16: Rewrite to facilitate ease of reading

P8, Lines 9 – 11: Sentence is awkward, I suggest it is rewritten for clarity

P8, Line 21: "Fingerprints"

P9, Line 1: Repeated word - "value values"

P14, Tables 1 and 2 legends: Right [parenthesis missing

P17, Figures 4: axis labels are too small and potentially unreadable in final print, please increase the text size

P17, Figures 4 legends: Right [parenthesis missing

P19, Figure 5: Axis labels missing

P20, Figure 6: Axis labels missing

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-312, 2017.