Geophysical Research Abstracts
Vol. 19, EGU2017-2099, 2017
EGU General Assembly 2017

STRAPS v1.0: Evaluating a methodology for predicting electron impact ionisation mass spectra for the aerosol mass spectrometer

David Topping (1,2), James Allan (1,2), Rami Alfarra (1,2), and Bernard Aumont (3)

(1) University of Manchester, Centre for Atmospheric Science, Manchester, United Kingdom (david.topping@manchester.ac.uk), (2) National Centre for Atmospheric Science, University of Manchester, Manchester, M13 9PL, United Kingdom, (3) LISA, UMR CNRS 7583, Universite Paris Est Creteil et Universite Paris Diderot, Creteil, France

Our ability to model the chemical and thermodynamic processes that lead to secondary organic aerosol (SOA) formation is thought to be hampered by the complexity of the system. While there are fundamental models now available that can simulate the tens of thousands of reactions thought to take place, validation against experiments is highly challenging. Techniques capable of identifying individual molecules such as chromatography are generally only capable of quantifying a subset of the material present, making it unsuitable for a carbon budget analysis. Integrative analytical methods such as the Aerosol Mass Spectrometer (AMS) are capable of quantifying all mass, but because of their inability to isolate individual molecules, comparisons have been limited to simple data products such as total organic mass and O:C ratio. More detailed comparisons could be made if more of the mass spectral information could be used, but because a discrete inversion of AMS data is not possible, this activity requires a system of predicting mass spectra based on molecular composition.

In this proof of concept study, the ability to train supervised methods to predict electron impact ionisation (EI) mass spectra for the AMS is evaluated. Supervised Training Regression for the Arbitrary Prediction of Spectra (STRAPS), is not built from first principles. A methodology is constructed whereby the presence of specific mass-to-charge ratio (m/z) channels are fit as a function of molecular structure before the relative peak height for each channel is similarly fit using a range of regression methods. The widely-used AMS mass spectral database is used as a basis for this, using unit mass resolution spectra of laboratory standards.

Key to the fitting process is choice of structural information, or molecular fingerprint. Initial results suggest the generic public 'MACCS' fingerprints provide the most accurate trained model when combined with both decision trees and random forests with median cosine angles of 0.94-.0.97 between modelled and measured spectra. There is some sensitivity to choice of fingerprint, but most sensitivity is in choice of regression technique. Support Vector Machines perform the worst, with median values of 0.78-0.85 and lower ranges approaching 0.4 depending on the fingerprint used. More detailed analysis of modelled versus mass spectra demonstrates important composition dependent sensitivities on a compound-by-compound basis. This is further demonstrated when we apply the trained methods to a model α-pinene SOA system, using output from the GECKO-A model. This shows that use of a generic fingerprint referred to as 'FP4' and one designed for vapour pressure predictions ('Nanoolal') give plausible mass spectra, whilst the use of the MACCS keys perform poorly in this application, demonstrating the need for evaluating model performance against other SOA systems rather than existing laboratory databases on single compounds.