Responses to Reviewers: Ch3MS-RF

Reviewer 1

This work presents a new method for predicting organic molecule properties (carbon number, mean oxidation state, oxygen-to-carbon ratio, vapor pressure) for compounds measured by gas chromatography and mass spectrometry but not listed in mass spectral databases. The novel idea is to use a statistical model trained on compounds listed in such databases together with parameters extracted from measurement, which is available regardless of the compound being listed in such a database. The authors include the caveat that this approach requires two-dimensional gas chromatography measurements that capture both volatility and polarity and on dimension is insufficient, which does require more complicated instrumentation than commonly deployed. Nonetheless, the general idea can be useful for the atmospheric science community, and is recommended for publication in Atmospheric Measurement Techniques. I note a few comments regarding the generalizability below.

    We thank the reviewer for this helpful summary and these comments. We do wish to clarify that two-dimensional gas chromatography is only required for quantification factor prediction which utilized second dimension retention information, as all chemical properties were well predicted using exclusively first dimension retention index (equivalent to standard GC-MS retention index) and mass spectrum. We clarify this point in lines 636 to 642 of newly added section 5.3, *Considerations for Adaptation Across Instruments and Methods.*

The authors refer to their "extrapolation set" in they are not included in the training set, but in reality it appears that new samples span a subset of the feature domain spanned by the training set. The consideration of whether extrapolation in this sense is happening or likely to happen in new data sets is relevant because random forest is not capable of such extrapolations - and this would limit the model's utility substantially. Can the authors clarify this point?

We thank the reviewer for this very helpful note. As noted, random forest modelling does not effectively extrapolate beyond the feature space of the training data set, and as such our terminology was somewhat unclear. We added the following text to section 3.2 *Training, test, and extrapolation set curation* (lines 376-385) to clarify that the extrapolation set is intended to test whether the training set is sufficiently similar to the sample medium of interest to indicate that the model will be capable of appropriately modelling the target compound properties, as follows:

    "The methodology described in this work cannot effectively extrapolate beyond the feature space of the training data set, and the identifiable organic compounds in the Amazonian aerosol samples are defined as an "extrapolation set" not because they test the abilities of the model to extrapolate beyond the feature space boundaries of the external standard training data, but because they represent the true range of individual isomer-specific identities observed in ambient samples. These compounds test the model's ability to extrapolate property prediction beyond the compound groups included in the external standard and indicate whether the sample is sufficiently similar to the training data to make this approach appropriate for the target sample medium, as extremely high prediction inaccuracies indicate compound classes too dissimilar from the training data to be appropriately modelled using Ch3MS-RF."

Thank you for this helpful comment. The protocol utilized in this methodology normalizes retention times to an alkane internal standard series such that retention indices are indicative of when compounds elute relative to known compounds (normal alkanes). Using a retention index is a very standard technique in chromatography to compare results for samples run on different instruments. In cases where elution times are normalized to known compounds such as normal alkanes, and similar phase columns are utilized, this technique is highly adaptable across instruments and techniques. To more quantitatively address this point, we have added section 5.3 *Considerations for Adaptation Across Instruments and Methods* and figure 8, in which we test how sensitive prediction performances for each property are to drifts in retention indices. The results of this analysis indicate that O:C and average carbon oxidation state predictions are not significantly affected by retention time drifts between the training set and the test set, while carbon number and vapor pressure predictions are more sensitive but still robust within drifts the equivalent of up to 1 carbon number unit. Section 5.3 now addresses this and other concerns, highlighting that spectra and retention times/indices produced by other instruments may be used, so long as retention index drifts can be normalized to less than the equivalent of the elution time between two linear alkanes difference, the column type is standardized, and the retention times and spectra of oxygenated species are consistently either derivatized or underivatized.

5.3 Considerations for Adaptation Across Instruments and Methods

The approach presented in this work prioritizes continuity between training, test, and sample data by exclusively training the model on data produced by a single instrument using a standardized methodology. This approach was selected to ensure that the patterns identified by Ch3MS-RF modelling in the training data were as directly relevant as possible to the unidentifiable sample compounds of interest. However, in some cases, accumulation of a representative external standard spanning the entire feature domain of unidentifiable compounds of interest may not be practical or possible. Electron ionization (70 eV) mass spectrometry is an extremely well characterized and consistent technique, but chromatographic retention times and indices can vary. In order for data produced by multiple instruments and techniques to be integrated within Ch3MS-RF, it is therefore important to establish the tolerance of prediction performance to drifts in retention index.

To test sensitivity to retention index or retention time shifts across instruments and methods, the vapor pressure, carbon number, $\overline{OS_c}$, and O:C of the external standard test set compounds were predicted using retention index inputs that were shifted from their observed retention indices. A broad range of shifts from -200 (indicating the equivalent of a two-carbon number shift, for example if in the test sample heptadecane were to elute at the time that pentadecane eluted in the training standard run) to +200 were tested (including -200, -150, -100, -50, -25, +25, +50, +100, +150, +200). A new mean average error was calculated for each set of predictions based on the shifted retention indices and compared to the unshifted mean average error to calculate the %

increase in mean average error as a function of test set retention index shift. These results are visualized in Figure 8. The two measures of oxidation, $\overline{OS_c}$ and O:C were relatively insensitive to retention index shifts, as their mean average errors increased by less than 10% at a retention index shift of ±200 and by < 5% within retention index shifts of ±100. Carbon number and vapor pressure predictions were more sensitive to retention index shifts, as would be expected given that retention times are more directly physically related to these two properties. At retention index shifts of + 200, mean average error of carbon number prediction increased by 44%, while a shift of -200 produced vapor pressure predictions that increased by 39%, both of which significantly decrease the utility of the produced predictions. However, within retention index shifts of ±100, increases in vapor pressure and carbon number prediction errors are modest, with all calculated MAE % error increases < 10%, with the exception of a 12% increase in error for vapor pressure predictions at a retention index shift of -100. Vapor pressure prediction in fact appears to slightly improve at shifts of +<25-50, but these improvements are extremely modest (<3%), are attributable to the generally higher uncertainties in vapor pressure prediction, and are not significantly different from predictions produced at a retention index shift of 0. Reported n-alkane normalized kovats indices of compounds within standardized column types (semistandard non-polar, standard non-polar, etc.) typically vary by <50, meaning that where methodologies allow test compound kovats or retention indices to be calculated, predictions utilizing training data from instruments and analysis protocols not used on the test samples are likely to be robust, particularly for O:C and $\overline{OS_c}$. For methodologies that do not use internal standards and that cannot otherwise easily yield kovats indices, protocols using similar columns and temperature ramps would likely produce retention times that could be substituted for retention indices in the Ch3MS-RF methodology. This approach would be usable across multiple instrumentations, provided it could be established that the retention times of any given compound produced by the training and test instrument drift by less than 1 carbon number equivalent.

In summary, training and/or test data from multiple instruments and protocols can be combined to meet user needs, provided the following criteria are met: 1) the same ionization energy (typically 70 eV) is used 2) retention index or retention time drifts between instruments or protocols can be normalized to less than the difference of the elution time between two sequential linear alkanes (retention index drift of <100) 3) similar phase columns are used (semistandard nonpolar, standard nonpolar, etc) 4) samples and training data are consistently either derivatized or underivatized, and if derivatized use a consistent derivatization agent. It is also important to keep in mind that the training data must span the anticipated feature space of the use data set, and that in cases of doubt this can be tested by adding extrapolation set compounds identified from the sample medium. For chemical properties modelling, this approach can be adapted from the GCxGC approach presented for any instrument using chromatography- electron ionization-mass spectrometry that has the capacity to yield at least unit resolution mass spectra and for which spectra can be sufficiently deconvoluted to yield clean analyte spectra. The model structure and provided sample code are highly flexible and could be utilized to predict any property of interest that might reasonably be expected to be reflected in the combination of compound mass spectra and chromatographic retention time, although performance evaluation is always important for ensuring that the patterns are sufficiently strong to enable accurate property prediction using Ch3MS-RF.
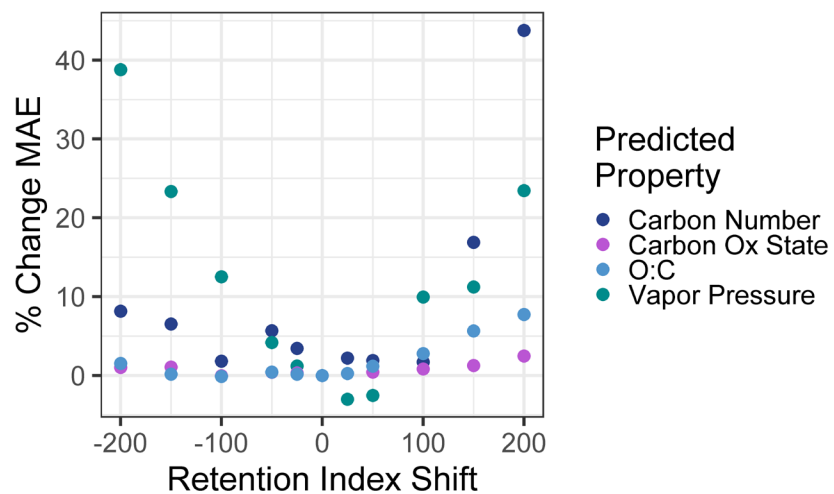
Figure 8: % increases of mean average error in chemical property prediction as a function of shift in test set retention index relative to training set retention index. Retention indices are normalized to a linear alkane series, making an increment of 100 indicate the retention time differences between two linear alkanes separated by 1 carbon number.