

Atmos. Meas. Tech. Discuss., referee comment RC2
<https://doi.org/10.5194/amt-2022-99-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on amt-2022-99

Anonymous Referee #2

Referee comment on "Ch3MS-RF: a random forest model for chemical characterization and improved quantification of unidentified atmospheric organics detected by chromatography–mass spectrometry techniques" by Emily B. Franklin et al., Atmos. Meas. Tech. Discuss., <https://doi.org/10.5194/amt-2022-99-RC2>, 2022

This study developed new machine learning techniques to characterize unidentifiable organic compounds using GC-MS and GCxGC-MS techniques. The authors provided a detailed discussion and demonstration of this model and its potential to improve the current understanding of undefined organic species in the atmosphere. This new method is able to improve the quantification accuracy compare with manual proxy modeling, which will lead to a better understanding of atmospheric organic aerosols formation and chemical properties. I'm supportive of this paper and recommend for publication in Atmospheric Measurement Techniques.

Here are a few minor comments:

Authors have mentioned that the vapor pressures were calculated for model training and evaluation, a few external standards test set and extrapolation set species were incompatible with vapor pressure prediction, can the author provide more explain more about how it is incompatible and the evaluation process for vapor pressure? Based on Figure 4, it seems like the predicted vapor pressure has more variability than other parameters, and the more accurate vapor pressure can improve the model accuracy.

Author mentioned that the model underestimated the high carbon oxidation state region and the high carbon number region, but there is no predicted data shown in the plot? it also seems like the model is a little bit overpredicted for the carbon number region between 20 and 30, can author comment on that?

Does this model capable of any GC-MS system or is there any specific requirement for the instrument? Can author add some discussion of the limitation of this model as well?

