Challenges and Opportunities in Advancing Speciated Characterization of Atmospheric Organics

E.B. Franklin[1,2], L. Yee[3], A. Goldstein[3], and D. Farmer[4]

[1]Colorado State University, Fort Collins, CO 80523; 303-551-3812, E-mail: emily.franklin@colostate.edu
[2]University of California at Berkeley, Berkeley, CA 94720
[3]University of California at Berkeley, Department of Environmental Science, Policy, and Management, Berkeley, CA 94720
[4]Colorado State University, Department of Chemistry, Fort Collins, CO 80523

The composition of atmospheric organics ranging from volatiles to aerosol are incredibly complex, spanning an estimated hundreds of thousands of unique chemical products. Relatively few of these compounds have been structurally identified and catalogued in mass spectral databases. While some properties of this complex organic material can be characterized by bulk analysis, speciated analysis, in particular structure-specific speciated analysis, remains critical for mechanistically probing atmospherically relevant aerosol formation and oxidation processes. Analyzing the speciated composition of highly complex organic mixtures, especially when most separated species cannot be identified, poses both instrumental and analytical challenges. This results in underutilization of the full scope of information made available by advances in instrumentation. This underutilization biases our understanding of the complexity of ambient aerosol formation dynamics, skewing analysis away from complex source groups in which mass is distributed among a larger number of individual contributors. Increased use of data science and machine learning-based techniques has the potential to advance speciated characterization of the atmosphere across environments without depending on synthesis of new products for structural confirmation. From remote forested environments to megacities, cataloging unidentifiable organics throughout the globe has the capacity to streamline identification of key process-specific tracers that currently lie beyond our boundaries of knowledge. Newly developed machine learning-based models allow us to place unidentifiable compounds in chemical property spaces commonly used throughout the atmospheric chemistry community, a process which provides critical context for designing targeted chamber experiments and highlights chemically distinct populations of products that lie beyond the current bounds of knowledge.

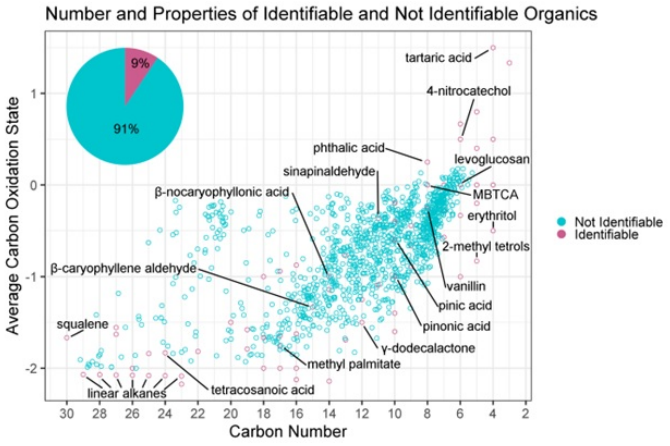

Figure 1. Chemical properties distributions in carbon number- average carbon oxidation state space of identifiable (pink) and not identifiable (teal) organic compounds identified in submicron aerosol collected at the GoAmazon field campaign. Identifiable compounds of interest are labelled in black. The percentage of identifiable and not identifiable compounds compared to the 1325 traced is illustrated by the pie chart in the top right. Carbon numbers of not identifiable compounds are predicted by the Ch3MS-RF model on a continuous scale and predicted carbon numbers are therefore not restricted to integer values.

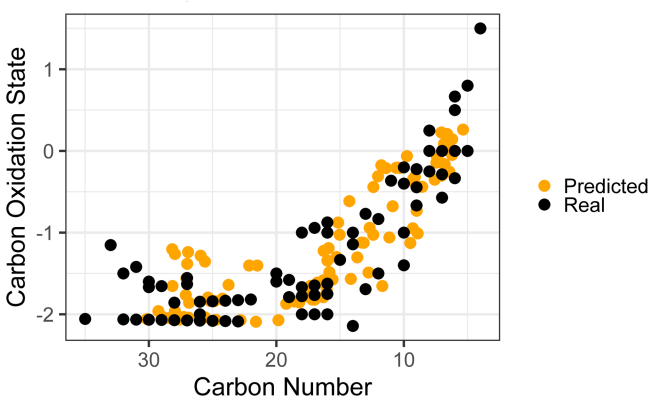

Figure 2. True versus predicted chemical properties distribution of identifiable organic species isolated from aerosol collected at the GoAmazon field campaign within a carbon number - average carbon oxidation state space. True properties are illustrated in black and properties predicted by the Ch3MS-RF model are illustrated in orange.