Title

Anthropogenic Influences on Coastal and Tropical Biogenic Aerosols: Advancing Data-Science-Driven Chemical Analysis for Climate and Public Health

Author

Franklin, Emily Barnes

Publication Date

2022

Peer reviewed|Thesis/dissertation

Anthropogenic Influences on Coastal and Tropical Biogenic Aerosols: Advancing Data-Science-Driven Chemical Analysis for Climate and Public Health

By

Emily Barnes Franklin

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Civil and Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Allen H. Goldstein, Chair
Professor Thomas W. Kirchstetter
Professor Robert A. Harley
Professor Carl Boettiger

Summer 2022

Abstract

Anthropogenic Influences on Coastal and Tropical Biogenic Aerosols: Advancing Data-Science-Driven Chemical Analysis for Climate and Public Health

by

Emily Barnes Franklin

Doctor of Philosophy in Civil and Environmental Engineering

University of California, Berkeley

Professor Allen H. Goldstein, Chair

Human activity significantly impacts the quantities, properties, and formation mechanisms of aerosols derived from biogenically produced organic chemicals. In this work, new methods are developed to expand speciated analysis of complex mixtures, and these methods are applied to two classes of human-impacted ambient aerosols: coastal marine aerosol, and tropical organic aerosol. The organic composition of both marine and tropical organic aerosol are largely uncharacterized, with over 85% of individual species separated and catalogued in each data set not present in current mass spectral libraries. Previously utilized methods for quantifying and characterizing novel atmospheric organics rely on manual judgements by individual researchers and are therefore highly inefficient and subject to errors that are difficult to quantify but assumed to be significant. To address this challenge, in Chapter 2 this work presents Ch3MS-RF, a machine learning-based model for predicting the chemical characteristics and instrument response factors of novel atmospheric organics based on their mass spectral fragmentation pattern and chromatographic retention. Chemical properties successfully modeled by Ch3MS-RF include carbon number, oxygen/carbon ratio, average carbon oxidation state, and volatility. This model achieves significant improvements in quantification accuracy over previous methods and enables novel atmospheric organics to be visualized in important chemical properties spaces for atmospheric chemistry, including the volatility basis set and Kroll diagram. Chapter 3 investigates the composition of the organic fraction of sea spray aerosol over a mesocosm phytoplankton bloom experiment conducted using coastal sea water. Results indicate that anthropogenic pollutants, including personal care products, oils, and urban compounds, significantly contribute to the organic fraction of sea spray aerosol, and that biological activity can transform this carbon pool by producing new biogenic species and transforming anthropogenic compounds. Chapter 4 focuses on a single class of anthropogenic coastal pollutants from the same experiment, the benzothiazoles. Benzothiazole is found to be emitted from ocean water in both gas and aerosol phases, and gas phase benzothiazole has the capacity to contribute to secondary aerosol formation when oxidized in the atmosphere. In the primary sea spray

aerosol, a diverse suite of benzothiazole-containing species are observed, in concentrations and speciations that are not reflective of those observed in the dissolved organic phase in seawater. Chapter 5 applies similar methods to aerosol samples collected at a semiremote field site in the central Amazon which is impacted by both fires and urban emissions. A high degree of interseasonal uniqueness was observed in secondary products formed in the atmosphere, indicating significant seasonal dependencies of secondary aerosol formation processes. Unique products observed under pristine conditions in the Amazonian wet season and fire impacted conditions in the dry season are not currently included in mass spectral libraries and are not replicated using common laboratory oxidation techniques, highlighting the importance of expanding chamber oxidation studies to simulate a wider range of ambient conditions to elucidate important ambient reaction mechanisms. A chemically speciated view of how human activity alters the properties of terrestrial tropical and marine aerosols will improve our mechanistic understanding of anthropogenic effects on aerosol properties, thereby improving our ability to predict selected aspects of aerosol-climate feedbacks to changing human behavior.

Table of contents

Dedication and Acknowledgments

This work is dedicated to Matthew James Franklin, my best friend, my husband, and my light through the tunnel. Thank you also to my parents for their unwavering support, and to my friends, both the ones I met through and during grad school and the ones who stuck with me along the way. Thank you to Jahmel for being the best brother I never had, and to Mike, Kath and Jon for teaching me to science through chaos with humor and grace. Thank you to Lindsay for unwavering patience, guidance, and support- representation in leadership matters, and I would not have finished without you. Finally, thank you to Allen for opening doors, for creating opportunities, and for always believing in me much more than I ever believed in myself.

1 Introduction

Atmospheric aerosols, suspensions of liquid or solid particulate matter in air, are a critical area of research to improve fundamental understanding of atmospheric chemistry, earth's climate, and environmental factors impacting public health. Uncertainties in aerosol impacts on radiative forcing contribute the largest fraction to uncertainties in total radiative forcing since preindustrial conditions. Knowledge gaps in sources, transformation, and fate of ambient aerosols significantly impede climate models' ability to quantify the current status of ambient radiative forcing and predict future changes.[1] Additionally, ambient aerosols are associated with negative outcomes in public health including premature death and are the leading global environmental health risk as of 2019, surpassing other risks such as unsafe water and unsafe sanitation.[2] Human emissions of pollutants into the environment impact ambient aerosols both directly and indirectly, as the emissions themselves can contribute to aerosol mass and properties, and the presence of pollutants can change the types of products formed compared to natural or biogenic aerosol formation mechanisms. In this work, human influences on two types of ambient aerosol, specifically coastal marine aerosol and tropical organic aerosol, are explored and chemically characterized. Knowledge gaps in each field are assessed, and new methodologies are developed to advance untargeted speciated organic analysis within atmospheric chemistry applications, all with the aim of advancing organic aerosol knowledge for chemistry, climate, and public health.

1.1 Motivation: Aerosol Radiative Forcing

Aerosols influence the energy balance of the earth's atmosphere through both direct and indirect mechanisms. Biogenic and naturally produced aerosols influence climate and earth's radiative balance, and anthropogenic aerosols and anthropogenic perturbations to natural aerosols change how these processes occur, contributing to anthropogenic radiative forcing. These human impacts on earth's energy balance via aerosol perturbations are termed "aerosol radiative forcing."[3] The direct mechanism by which aerosols impact radiative forcing is interactions between the aerosols and light and is highly dependent upon aerosol optical properties (color). Light colored aerosols such as sulfate and most organic aerosols reflect light and increase albedo, leading to a cooling effect, while dark aerosols such as black carbon absorb radiation and lead to warming. The extent to which different classes of anthropogenic aerosol emissions impact direct radiative forcing is accounted for differently across climate models, leading to uncertainties and disagreements. For example, some models attribute a positive radiative forcing effect to biomass burning aerosols, while others attribute a negative radiative forcing effect to these species,[4] and the radiative properties of organic aerosol can change from reflective to absorptive due to the production of "brown" organic aerosol during oxidative ageing.[5]

In addition to directly interacting with solar and terrestrial radiation, aerosols impact radiative forcing through their impacts on clouds, specifically through altering cloud radiative properties and lifetimes.[6,7,5] Secondary marine aerosol, which is discussed

in full in "Marine Aerosols: Production and Anthropogenic Influences" was recently determined to play a dominant role in cloud formation over the ocean,[8] rendering human impacts on secondary marine aerosol formation particularly relevant to radiative forcing through the aerosol indirect effect. The emissions and properties of natural aerosols under pre-industrial conditions, including sea spray aerosols and biogenic organic aerosol, are still incompletely understood and contribute significantly (on the order of 50%) to total uncertainties in aerosol indirect forcing.[9] An improved understanding of aerosol composition and properties in environments at the boundary between human emissions and significant natural sources of aerosols adds to the body of knowledge required to improve projections and reduce uncertainties in how human alterations of natural aerosols impact climate.

1.2 Motivation: Aerosol Impacts on Public Health

Aerosols are also an important area of environmental engineering research because of their impacts on public health. As of 2019, ambient particulate matter is the leading environmental health risk and a top 10 health risk overall, making it a more significant contribution to global health risks than unsafe water and sanitation and similar to metabolic health risks such as high cholesterol.[2] Fine particulate matter, defined as the fraction of ambient aerosol particulates with an aerodynamic diameter less than 2.5 μm ($PM_{2.5}$) or 1 μm (PM_1), depending on context, does not in most surface environments dominate aerosol mass, but it is the most important fraction for human health. Unlike coarse aerosols, which are due to their momentum trapped in the nose and throat, fine particulates are able to penetrate deep into lung tissue; this leads to inflammation of the lungs and oxidative stress, and exposes community members to any carcinogens, toxins, or otherwise hazardous constituents of the particulates themselves.[10,11] These mechanistic insights into how aerosols impact lung tissue and exposure are confirmed by epidemiological exposure-response studies, which find that aerosols elevate mortality and contribute to the millions of annual deaths attributable to air pollution.[11,12]

Both tropical organic aerosol and coastal marine aerosol have been specifically identified as threats to public health, although through different mechanisms. Coastal sea spray aerosol can expose coastal communities to toxins and pollutants in coastal water, such as toxins from harmful algal blooms and carcinogens from wastewater discharge.[13–15] The importance of sea spray aerosol as an exposure route by which coastal communities are impacted by marine pollutants is difficult to establish epidemiologically as coastal communities are also likely to be exposed via consumption of fish, but it has been proposed as an important driver of cancer and poor respiratory outcomes in coastal regions.[16] In tropical regions such as the Amazon rainforest, anthropogenic activity, specifically fires that are used to clear forest for agricultural uses, significantly alters ambient aerosol concentrations, often leading $PM_{2.5}$ levels to exceed WHO guidelines.[17–19] In the Amazon rainforest, fire activity is associated with thousands of premature deaths and is one of the key causes of respiratory hospitalizations for indigenous people.[20,21]

1.3 Challenges and Knowledge Gaps in Speciated Organic Aerosol Characterization

To advance knowledge of how human emissions impact aerosol organic composition, this work employs techniques that separate the components of ambient organic aerosol material so that they can be structurally identified and characterized at an isomer-specific level. Isomer-specific identification is critical to many atmospheric chemistry applications, as different isomers of a given molecule can vary significantly in important properties such as volatility, and structure-specific identification is critical in identifying reaction mechanisms.[22,23] However, this approach is complicated by the extreme chemical complexity and novelty of organic aerosol material. There are an estimated millions of unique gas and aerosol-phase atmospheric organic constituents, and this composition is extremely variable, even at a given fixed site.[24,25] Recent efforts have achieved significant progress in approaching mass closure in tracing reactive carbon over multiple generations of oxidation using an ensemble of bulk and speciated measurements, but there has been far less progress towards isomer-specific mass closure.[26–28] The majority of organic compounds separated and detected by chromatography-mass spectrometry techniques like the ones utilized in this work are not listed in commonly available mass spectral databases and have never been synthesized and made available as authentic standards, making them impossible to definitively identify.[29,30] This remains true, despite the rapidly increasing size of the commonly available NIST/EPA/NIH mass spectral database, with on the order of tens of thousands of new spectra added between releases at intervals of less than 5 years.[31]

A typical ambient organic aerosol sample analyzed using GCxGC-MS (two dimensional gas chromatography coupled with electron ionization mass spectrometry), the primary instrument utilized in this work, contains on the order of high hundreds to low thousands of unique organic compounds above detection limits, the vast majority of which, as discussed in chapters 3 and 5, cannot be definitively identified. A common simplification choice in addressing such complex mixtures is to restrict analysis to the fraction of compounds that can be definitively identified, or a selected subset of identifiable compounds. However, in cases where knowledge biases exist between groups, meaning that a disproportionate share of a given functionality or source group are identifiable, restricting analysis to identifiable constituents can bias speciated analysis of complex organic mixtures. To address this challenge and analyze its potential influence on the study of coastal and tropical organic aerosol, this work presents a model to allow unidentifiable organics to be chemically characterized and explores the composition and sources of unidentifiable atmospheric organics. Chapter 2 presents a new methodology for characterizing unidentifiable organic constituents measured in complex environmental mixtures. Chapters 3 and 5 present analysis of compositional and source-based knowledge biases in the organic composition of marine and tropical organic aerosol materials respectively, highlighting which sources and which types of organic products are observed in the environment but not included in current mechanistic representations of aerosol

chemistry to guide future laboratory experiments, synthesis efforts, and comparisons across field sites.

1.4 Marine Organic Aerosols: Production and Anthropogenic Influences

The ocean contributes to the global aerosol burden through both primary and secondary mechanisms. Sea spray aerosols (SSA) are particulates and droplets directly emitted from the ocean's surface by wind shear and bubble bursting activity, while secondary marine aerosols (SMA) are produced from the oxidation of volatile organic compounds (VOCs) emitted from the ocean.[32,33,34] While marine aerosols are by mass primarily inorganic due to the high salt content, the organic content of marine aerosols and the transfer of organic material from ocean to atmosphere plays an important role in marine atmospheric chemistry. The organic content of primary SSA forms a coating on the exterior of the salt core; as salt and organics have different cloud and ice nucleation properties, the presence and composition of the organic coating is important for cloud formation over the ocean.[35–38] Marine organics are also transferred into the atmosphere in gas-phase through both volatilization and reactions at the ocean's surface that produce volatile products.[33,39,40] When oxidized in the atmosphere, depending on their volatility and functionalization, these gasses can form low volatility products that form secondary aerosol material.[32] When the organics contain sulfur, as is the case for the dominant single contributor to SMA material, dimethyl sulfide (DMS), both organic and sulfate SMA material is produced; while early investigations of SMA focused on DMS to the exclusion of other precursors, recent studies have found that other precursors play an important role in both the sulfate and total SMA emissions budgets.[8,41,40] In Chapter 4, the gas-phase emissions of a common organic pollutant, benzothiazole, are reported, and its secondary marine aerosol formation potential is explored. This observation contributed to the work of Kilgour et al., 2022, which finds that this compound has the capacity to substantially contribute to the marine sulfur budget of polluted coastal systems.[41] As previously noted, SMA plays a dominant role in cloud formation over the ocean, rendering an improved understanding of SMA precursor emissions from both clean and polluted marine environments critical for improving climate models.

The ocean becomes enriched with organic material through both biogenic and anthropogenic mechanisms. Phytoplankton create organic biomass from CO_2 through photosynthesis, and this organic material is transformed by the marine microbial loop, creating a significant and molecularly diverse source of organic material in ocean water.[42,43] Human emissions also contribute to marine organic material, particularly in coastal and/or polluted environments; significant anthropogenic organic inputs include urban runoff and wastewater discharge, both of which are often enriched with personal care products, trash, and shipping pollution.[44–46] These primary anthropogenic pollutants are subject to consumption and/or transformation by marine microbiology, producing biologically transformed organic products from originally anthropogenic carbon sources.[47–49] In addition to the direct impacts of pollution on SSA and SMA precursor emissions, human activity also influences the organic composition of marine aerosols indirectly through impacts on marine biological activity. Anthropogenic discharges of

fertilizer-enriched runoff and climate change-related changes in ocean temperatures and chemistry can induce intense phytoplankton blooms, which add significant quantities of organic material to ocean surface waters.[43,50–54] An improved understanding of how biological activity influences marine aerosol composition and properties is therefore important, not only for improving knowledge of marine aerosol production under pristine or preindustrial conditions but also for characterizing and predicting how anthropogenic perturbations of marine microbiology are likely to influence atmospheric chemistry and public health.

In Chapter 3, anthropogenic, biogenic, and biologically transformed constituents of primary sea spray aerosol are speciated and characterized from samples collected over a biological bloom mesocosm experiment to advance knowledge of how pollution and biology mediate the organic composition of SSA.

1.5 Tropical Organic Aerosol: Production and Anthropogenic Influences

In densely forested environments such as the Amazon rainforest, fine aerosol material is dominated by organics.[18,55,56] Much of this organic material is secondary in nature and originates from biogenic volatile organic compounds (BVOCs), meaning that gas-phase emissions from plant material are oxidized in the atmosphere to form secondary organic aerosol (SOA).[57] While plants emit many VOCs of differing formulae and structures, biogenic SOA formation is dominated by isoprene, monoterpenes (a diverse group of terpenoids sharing a chemical formula of $C_{10}H_{16}$), and sesquiterpenes (also terpenoids, sharing a chemical formula of $C_{15}H_{24}$).[58] Anthropogenic emissions alter the quantities and composition of organic aerosol in impacted regions through multiple mechanisms. Primary organic aerosol emissions, in particular from combustion sources, add primary aerosol material to the atmosphere, and anthropogenic emissions of volatile organic compounds contribute to SOA formation.[59] Anthropogenic emissions also have indirect impacts on organic aerosol quantities and composition by altering, and in many contexts enhancing, the formation of biogenic SOA. The mechanisms by which these enhancements occur are incompletely understood, but are likely attributable to changes in partitioning due to anthropogenic increases in aerosol surface area, as well as changes in oxidation chemistry due to the introduction of anthropogenic oxidants that produce higher quantities of functionalized, low volatility products than would be formed under pristine conditions.[60]

In the Amazon rainforest, two types of anthropogenic influence have been observed and modeled to significantly contribute to organic aerosol; urban emissions, and fires.[61,62] While in many environments fires naturally occur and would not necessarily be defined as anthropogenic, in the Amazon basin, fires are considered to be nearly exclusively human caused. Frequent rain and high humidity conditions render natural fire ignition very unlikely, and fire is used as a deforestation method to clear land for cattle and other agricultural purposes.[19,63,64] Fire activity is highest during the Amazonian dry season, and the high aerosol concentrations observed in the dry season are largely attributed to biomass burning sources.[18] However, fires also emit high concentrations of NO_x,[65] and the extent

to which fires may contribute to alterations in biogenic SOA formation in the Amazonian dry season are incompletely understood. The air quality and atmospheric chemistry of the Amazon basin is also impacted by increasing urbanization. Formation of secondary organic aerosol in regions downwind of the city of Manaus has been extensively studied[17,56,66–69] and significant enhancements have been both modeled and observed,[70] but these enhancements are incompletely mechanistically understood and are not currently well replicated with explicit molecular models.[71] In Chapter 5, the influences of urban emissions and fires on submicron aerosol composition at a semiremote site in the central Amazon are investigated across both the wet and dry seasons to identify seasonally unique chemical signatures of anthropogenic influence and provide speciated compositional insights into how pollution-influences biogenic SOA products differ from their counterparts formed under pristine conditions.

1.6 Structure of the Dissertation

This work characterizes two types of ambient aerosols, both of which are naturally produced but significantly influenced by human emissions: coastal marine aerosols, and secondary organic aerosol from tropical forested environments. To address the challenges posed by the high degree of complexity of ambient aerosol organic material, Chapter 2 presents Ch3MS-RF, a machine learning based model capable of predicting properties of organic aerosol constituents that have been separated and detected but cannot be definitively identified. The chemical composition of coastal sea spray aerosol is investigated in Chapter 3, while the ocean-atmosphere transfer of a single class of marine organic pollutants, the benzothiazoles, is described in Chapter 4. Chapter 5 focuses on the organic aerosols in the Amazon rainforest, exploring how urban pollution plumes and smoke influences aerosol quantities and composition. Finally, Chapters 3 and 5 present quantitative analyses of the state of knowledge of different classes of aerosol constituents in coastal and tropical contexts respectively, highlighting priorities for future synthesis and chamber oxidation experiments. Chapter 6 summarizes the major findings and outlines opportunities and priorities for future research, including additional developments of machine learning based data processing methodologies for atmospheric applications, priorities for marine pollution aerosolization monitoring, and priorities for chamber oxidation experiments that can replicate tropical aerosol formation conditions. Together, this work advances the state of knowledge on how human emissions impact aerosol organic properties in natural systems in order to identify potential threats to public health and advance the mechanistic understanding of anthropogenic impacts on aerosol radiative forcing.

1.7 References

(1) Bender, F. A.-M. Aerosol Forcing: Still Uncertain, Still Relevant. *AGU Adv.* **2020**, *1* (3), e2019AV000128. https://doi.org/10.1029/2019AV000128.

(2) Murray, C. J. L; et al. Global Burden of 87 Risk Factors in 204 Countries and Territories, 1990–2019: A Systematic Analysis for the Global Burden of Disease Study 2019. *Lancet* **2020**, *396* (10258), 1223–1249. https://doi.org/10.1016/S0140-6736(20)30752-2.

(3) Charlson, R. J.; Schwartz, S. E.; Hales, J. M.; Cess, R. D.; Coakley, J. A.; Hansen, J. E.; Hofmann, D. J. Climate Forcing by Anthropogenic Aerosols. *Science (80-.).* **1992**, *255* (5043), 423–430. https://doi.org/10.1126/SCIENCE.255.5043.423.

(4) Myhre, G.; Samset, B. H.; Schulz, M.; Balkanski, Y.; Bauer, S.; Berntsen, T. K.; Bian, H.; Bellouin, N.; Chin, M.; Diehl, T.; Easter, R. C.; Feichter, J.; Ghan, S. J.; Hauglustaine, D.; Iversen, T.; Kinne, S.; Kirkeväg, A.; Lamarque, J. F.; Lin, G.; Liu, X.; Lund, M. T.; Luo, G.; Ma, X.; Van Noije, T.; Penner, J. E.; Rasch, P. J.; Ruiz, A.; Seland; Skeie, R. B.; Stier, P.; Takemura, T.; Tsigaridis, K.; Wang, P.; Wang, Z.; Xu, L.; Yu, H.; Yu, F.; Yoon, J. H.; Zhang, K.; Zhang, H.; Zhou, C. Radiative Forcing of the Direct Aerosol Effect from AeroCom Phase II Simulations. *Atmos. Chem. Phys.* **2013**, *13* (4), 1853–1877. https://doi.org/10.5194/ACP-13-1853-2013.

(5) Tsigaridis, K.; Kanakidou, M. The Present and Future of Secondary Organic Aerosol Direct Forcing on Climate. *Curr. Clim. Chang. Reports 2018 42* **2018**, *4* (2), 84–98. https://doi.org/10.1007/S40641-018-0092-3.

(6) Twomey, S. Pollution and the Planetary Albedo. *Atmos. Environ.* **1974**, *8* (12), 1251–1256. https://doi.org/10.1016/0004-6981(74)90004-3.

(7) Haywood, J.; Boucher, O. Estimates of the Direct and Indirect Radiative Forcing Due to Tropospheric Aerosols: A Review. *Rev. Geophys.* **2000**, *38* (4), 513–543. https://doi.org/10.1029/1999RG000078.

(8) Mayer, K. J.; Wang, X.; Santander, M. V.; Mitts, B. A.; Sauer, J. S.; Sultana, C. M.; Cappa, C. D.; Prather, K. A. Secondary Marine Aerosol Plays a Dominant Role over Primary Sea Spray Aerosol in Cloud Formation. *ACS Cent. Sci.* **2020**. https://doi.org/10.1021/acscentsci.0c00793.

(9) Carslaw, K. S.; Lee, L. A.; Reddington, C. L.; Pringle, K. J.; Rap, A.; Forster, P. M.; Mann, G. W.; Spracklen, D. V.; Woodhouse, M. T.; Regayre, L. A.; Pierce, J. R. Large Contribution of Natural Aerosols to Uncertainty in Indirect Forcing. *Nat. 2013 5037474* **2013**, *503* (7474), 67–71. https://doi.org/10.1038/nature12674.

(10) Pöschl, U. Atmospheric Aerosols: Composition, Transformation, Climate and Health Effects. *Angew. Chemie Int. Ed.* **2005**, *44* (46), 7520–7540. https://doi.org/10.1002/ANIE.200501122.

(11) Shiraiwa, M.; Ueda, K.; Pozzer, A.; Lammel, G.; Kampf, C. J.; Fushimi, A.; Enami, S.; Arangio, A. M.; Fröhlich-Nowoisky, J.; Fujitani, Y.; Furuyama, A.; Lakey, P. S. J.; Lelieveld, J.; Lucas, K.; Morino, Y.; Pöschl, U.; Takahama, S.; Takami, A.; Tong, H.; Weber, B.; Yoshino, A.; Sato, K. Aerosol Health Effects from Molecular to Global Scales. *Environ. Sci. Technol.* **2017**, *51* (23), 13545–13567. https://doi.org/10.1021/ACS.EST.7B04417/ASSET/IMAGES/LARGE/ES-2017-04417P_0004.JPEG.

(12) Dockery, D. W.; Pope, C. A.; Xu, X.; Spengler, J. D.; Ware, J. H.; Fay, M. E.; Ferris, B. G. J.; Speizer, F. E. An Association between Air Pollution and Mortality in Six U.S. Cities. *https://doi.org/10.1056/NEJM199312093292401* **1993**, *329* (24), 1753–1759. https://doi.org/10.1056/NEJM199312093292401.

(13) Kirkpatrick, B.; Fleming, L. E.; Squicciarini, D.; Backer, L. C.; Clark, R.; Abraham, W.; Benson, J.; Cheng, Y. S.; Johnson, D.; Pierce, R.; Zaias, J.; Bossart, G. D.; Baden, D. G. Literature Review of Florida Red Tide: Implications for Human Health Effects. *Harmful*

Algae. Elsevier April 1, 2004, pp 99–115. https://doi.org/10.1016/j.hal.2003.08.005.

(14) Grattan, L. M.; Holobaugh, S.; Morris, J. G. Harmful Algal Blooms and Public Health. *Harmful Algae*. Elsevier B.V. July 1, 2016, pp 2–8. https://doi.org/10.1016/j.hal.2016.05.003.

(15) Cincinelli, A.; Stortini, A. M.; Perugini, M.; Checchini, L.; Lepri, L. Organic Pollutants in Sea-Surface Microlayer and Aerosol in the Coastal Environment of Leghorn—(Tyrrhenian Sea). *Mar. Chem.* **2001**, *76* (1–2), 77–98. https://doi.org/10.1016/S0304-4203(01)00049-4.

(16) Walsh, J. J.; Lenes, J. M.; Weisberg, R. H.; Zheng, L.; Hu, C.; Fanning, K. A.; Snyder, R.; Smith, J. More Surprises in the Global Greenhouse: Human Health Impacts from Recent Toxic Marine Aerosol Formations, Due to Centennial Alterations of World-Wide Coastal Food Webs. *Mar. Pollut. Bull.* **2017**, *116* (1–2), 9–40. https://doi.org/10.1016/J.MARPOLBUL.2016.12.053.

(17) Shilling, J. E.; Pekour, M. S.; Fortner, E. C.; Artaxo, P.; De Sá, S.; Hubbe, J. M.; Longo, K. M.; Machado, L. A. T.; Martin, S. T.; Springston, S. R.; Tomlinson, J.; Wang, J. Aircraft Observations of the Chemical Composition and Aging of Aerosol in the Manaus Urban Plume during GoAmazon 2014/5. *Atmos. Chem. Phys.* **2018**, *18* (14), 10773–10797. https://doi.org/10.5194/ACP-18-10773-2018.

(18) De Sá, S. S.; Rizzo, L. V; Palm, B. B.; Campuzano-Jost, P.; Day, D. A.; Yee, L. D.; Wernis, R.; Isaacman-Vanwertz, G.; Brito, J.; Carbone, S.; Liu, Y. J.; Sedlacek, A.; Springston, S.; Goldstein, A. H.; Barbosa, H. M. J.; Alexander, M. L.; Artaxo, P.; Jimenez, J. L.; Martin, S. T.; Paulson, J. A. Contributions of Biomass-Burning, Urban, and Biogenic Emissions to the Concentrations and Light-Absorbing Properties of Particulate Matter in Central Amazonia during the Dry Season. *Atmos. Chem. Phys* **2019**, *19*, 7973–8001. https://doi.org/10.5194/acp-19-7973-2019.

(19) Cardil, A.; de-Miguel, S.; Silva, C. A.; Reich, P. B.; Calkin, D.; S Brancalion, P. H.; Vibrans, A. C.; P Gamarra, J. G.; Zhou, M.; Pijanowski, B. C.; Hui, C.; Crowther, T. W.; Hérault, B.; Piotto, D.; Salas-Eljatib, C.; North Broadbent, E.; Almeyda Zambrano, A. M.; Picard, N.; O C Aragão, L. E.; Bastin, J.-F.; Routh, D.; van den Hoogen, J.; Peri, P. L.; Liang, J. Recent Deforestation Drove the Spike in Amazonian Fires Environmental Research Letters Recent Deforestation Drove the Spike in Amazonian Fires. *Environ. Res. Lett.* **2020**. https://doi.org/10.1088/1748-9326/abcac7.

(20) Alves, L. Amazon Fires Coincide with Increased Respiratory Illnesses in Indigenous Populations. *Lancet Respir. Med.* **2020**, *8* (11), e84. https://doi.org/10.1016/S2213-2600(20)30421-5.

(21) Nawaz, M. O.; Henze, D. K. Premature Deaths in Brazil Associated With Long-Term Exposure to PM2.5 From Amazon Fires Between 2016 and 2019. *GeoHealth* **2020**, *4* (8), e2020GH000268. https://doi.org/10.1029/2020GH000268.

(22) Nozière, B.; Kalberer, M.; Claeys, M.; Allan, J.; D'Anna, B.; Decesari, S.; Finessi, E.; Glasius, M.; Grgić, I.; Hamilton, J. F.; Hoffmann, T.; Iinuma, Y.; Jaoui, M.; Kahnt, A.; Kampf, C. J.; Kourtchev, I.; Maenhaut, W.; Marsden, N.; Saarikoski, S.; Schnelle-Kreis, J.; Surratt, J. D.; Szidat, S.; Szmigielski, R.; Wisthaler, A. The Molecular Identification of Organic Compounds in the Atmosphere: State of the Art and Challenges. *Chem. Rev.* **2015**, *115* (10), 3919–3983. https://doi.org/10.1021/CR5003485/ASSET/IMAGES/CR5003485.SOCIAL.JPEG_V03.

(23) Isaacman-Vanwertz, G.; Aumont, B. Impact of Organic Molecular Structure on the Estimation of Atmospherically Relevant Physicochemical Parameters. *Atmos. Chem. Phys.* **2021**, *21* (8), 6541–6563. https://doi.org/10.5194/ACP-21-6541-2021.

(24) Goldstein, A. H.; Galbally, I. E. Known and Unexplored Organic Constituents in the Earth's Atmosphere. *Environ. Sci. Technol.* **2007**, *41* (5), 1514–1521. https://doi.org/10.1021/ES072476P.

(25) Ditto, J. C.; Barnes, E. B.; Khare, P.; Takeuchi, M.; Joo, T.; Bui, A. A. T.; Lee-Taylor, J.; Eris, G.; Chen, Y.; Aumont, B.; Jimenez, J. L.; Ng, N. L.; Griffin, R. J.; Gentner, D. R. An Omnipresent Diversity and Variability in the Chemical Composition of Atmospheric Functionalized Organic Aerosol. *Commun. Chem. 2018 11* **2018**, *1* (1), 1–13. https://doi.org/10.1038/s42004-018-0074-3.

(26) Heald, C. L.; Kroll, J. H. The Fuel of Atmospheric Chemistry: Toward a Complete Description of Reactive Organic Carbon. *Sci. Adv.* **2020**, *6* (6). https://doi.org/10.1126/SCIADV.AAY8967/SUPPL_FILE/AAY8967_SM.PDF.

(27) Hunter, J. F.; Day, D. A.; Palm, B. B.; Yatavelli, R. L. N.; Chan, A. W. H.; Kaser, L.; Cappellin, L.; Hayes, P. L.; Cross, E. S.; Carrasquillo, A. J.; Campuzano-Jost, P.; Stark, H.; Zhao, Y.; Hohaus, T.; Smith, J. N.; Hansel, A.; Karl, T.; Goldstein, A. H.; Guenther, A.; Worsnop, D. R.; Thornton, J. A.; Heald, C. L.; Jimenez, J. L.; Kroll, J. H. Comprehensive Characterization of Atmospheric Organic Carbon at a Forested Site. **2017**. https://doi.org/10.1038/NGEO3018.

(28) Isaacman-Vanwertz, G.; Massoli, P.; O'Brien, R.; Lim, C.; Franklin, J. P.; Moss, J. A.; Hunter, J. F.; Nowak, J. B.; Canagaratna, M. R.; Misztal, P. K.; Arata, C.; Roscioli, J. R.; Herndon, S. T.; Onasch, T. B.; Lambe, A. T.; Jayne, J. T.; Su, L.; Knopf, D. A.; Goldstein, A. H.; Worsnop, D. R.; Kroll, J. H. Chemical Evolution of Atmospheric Organic Carbon over Multiple Generations of Oxidation. *Nat. Chem. 2018 104* **2018**, *10* (4), 462–468. https://doi.org/10.1038/s41557-018-0002-2.

(29) Hamilton, J. F.; Webb, P. J.; Lewis, A. C.; Hopkins, J. R.; Smith, S.; Davy, P. Atmospheric Chemistry and Physics Partially Oxidised Organic Components in Urban Aerosol Using GCXGC-TOF/MS. *Atmos. Chem. Phys* **2004**, *4*, 1279–1290.

(30) Worton, D. R.; Decker, M.; Isaacman-VanWertz, G.; Chan, A. W. H.; Wilson, K. R.; Goldstein, A. H. Improved Molecular Level Identification of Organic Compounds Using Comprehensive Two-Dimensional Chromatography, Dual Ionization Energies and High Resolution Mass Spectrometry. *Analyst* **2017**, *142* (13), 2395–2403. https://doi.org/10.1039/c7an00625j.

(31) Vinaixa, M.; Schymanski, E. L.; Neumann, S.; Navarro, M.; Salek, R. M.; Yanes, O. Mass Spectral Databases for LC/MS- and GC/MS-Based Metabolomics: State of the Field and Future Prospects. *TrAC Trends Anal. Chem.* **2016**, *78*, 23–35. https://doi.org/10.1016/J.TRAC.2015.09.005.

(32) Myriokefalitakis, S.; Vignati, E.; Tsigaridis, K.; Papadimas, C.; Sciare, J.; Mihalopoulos, N.; Facchini, M. C.; Rinaldi, M.; Dentener, F. J.; Ceburnis, D.; Hatzianastasiou, N.; O'Dowd, C. D.; van Weele, M.; Kanakidou, M. Global Modeling of the Oceanic Source of Organic Aerosols. *Adv. Meteorol.* **2010**, *2010*, 1–16. https://doi.org/10.1155/2010/939171.

(33) O'Dowd, C. D.; De Leeuw, G. Marine Aerosol Production: A Review of the Current

Knowledge. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. Royal Society July 2007, pp 1753–1774. https://doi.org/10.1098/rsta.2007.2043.

(34) De Leeuw, G.; Neele, F. P.; Hill, M.; Smith, M. H.; Vignati, E. Production of Sea Spray Aerosol in the Surf Zone. *J. Geophys. Res. Atmos.* **2000**, *105* (D24), 29397–29409. https://doi.org/10.1029/2000JD900549.

(35) O'Dowd, C. D.; Facchini, M. C.; Cavalli, F.; Ceburnis, D.; Mircea, M.; Decesari, S.; Fuzzi, S.; Young, J. Y.; Putaud, J. P. Biogenically Driven Organic Contribution to Marine Aerosol. *Nature* **2004**, *431* (7009), 676–680. https://doi.org/10.1038/nature02959.

(36) Lee, H. D.; Morris, H. S.; Laskina, O.; Sultana, C. M.; Lee, C.; Jayarathne, T.; Cox, J. L.; Wang, X.; Hasenecz, E. S.; Demott, P. J.; Bertram, T. H.; Cappa, C. D.; Stone, E. A.; Prather, K. A.; Grassian, V. H.; Tivanski, A. V. Organic Enrichment, Physical Phase State, and Surface Tension Depression of Nascent Core-Shell Sea Spray Aerosols during Two Phytoplankton Blooms. *ACS Earth Sp. Chem.* **2020**, *4* (4), 650–660. https://doi.org/10.1021/acsearthspacechem.0c00032.

(37) Hodas, N.; Zuend, A.; Mui, W.; Flagan, R. C.; Seinfeld, J. H. Influence of Particle-Phase State on the Hygroscopic Behavior of Mixed Organic–Inorganic Aerosols. *Atmos. Chem. Phys.* **2015**, *15* (9), 5027–5045. https://doi.org/10.5194/acp-15-5027-2015.

(38) Fuentes, E.; Coe, H.; Green, D.; McFiggans, G. On the Impacts of Phytoplankton-Derived Organic Matter on the Properties of the Primary Marine Aerosol - Part 2: Composition, Hygroscopicity and Cloud Condensation Activity. *Atmos. Chem. Phys.* **2011**, *11* (6), 2585–2602. https://doi.org/10.5194/acp-11-2585-2011.

(39) Carpenter, L. J.; Archer, S. D.; Beale, R. Ocean-Atmosphere Trace Gas Exchange. *Chem. Soc. Rev.* **2012**, *41* (19), 6473–6506. https://doi.org/10.1039/c2cs35121h.

(40) Trueblood, J. V.; Wang, X.; Or, V. W.; Alves, M. R.; Santander, M. V.; Prather, K. A.; Grassian, V. H. The Old and the New: Aging of Sea Spray Aerosol and Formation of Secondary Marine Aerosol through OH Oxidation Reactions. *ACS Earth Sp. Chem.* **2019**, *3* (10), 2307–2314. https://doi.org/10.1021/acsearthspacechem.9b00087.

(41) Kilgour, D. B.; Novak, G. A.; Sauer, J. S.; Moore, A. N.; Dinasquet, J.; Amiri, S.; Franklin, E. B.; Mayer, K.; Winter, M.; Morris, C. K.; Price, T.; Malfatti, F.; Crocker, D. R.; Lee, C.; Cappa, C. D.; Goldstein, A. H.; Prather, K. A.; Bertram, T. H. Marine Gas-Phase Sulfur Emissions during an Induced Phytoplankton Bloom. *Atmos. Chem. Phys.* **2022**, *22* (2), 1601–1613. https://doi.org/10.5194/ACP-22-1601-2022.

(42) Azam, F.; Fenchel, T.; Field, J. G.; Prog, S.; Gray, J. S.; Meyer-Reil L.A.; Thingstad F. The Ecological Role of Water-Column Microbes in the Sea Placing Marine Mixoplankton in Context View Project DOMAINE View Project MARINE ECOLOGY-PROGRESS SERIES The Ecological Role of Water-Column Microbes in the Sea*. *Mar. Ecol. Prog.* **1983**, *10*, 257–263. https://doi.org/10.3354/meps010257.

(43) Cochran, R. E.; Laskina, O.; Trueblood, J. V.; Estillore, A. D.; Morris, H. S.; Jayarathne, T.; Sultana, C. M.; Lee, C.; Lin, P.; Laskin, J.; Laskin, A.; Dowling, J. A.; Qin, Z.; Cappa, C. D.; Bertram, T. H.; Tivanski, A. V.; Stone, E. A.; Prather, K. A.; Grassian, V. H. Molecular Diversity of Sea Spray Aerosol Particles: Impact of Ocean Biology on Particle Composition and Hygroscopicity. *Chem* **2017**, *2* (5), 655–667. https://doi.org/10.1016/j.chempr.2017.03.007.

(44) Tovar-Sánchez, A.; Sánchez-Quiles, D.; Basterretxea, G.; Benedé, J. L.; Chisvert, A.; Salvador, A.; Moreno-Garrido, I.; Blasco, J. Sunscreen Products as Emerging Pollutants to Coastal Waters. *PLoS One* **2013**, *8* (6), e65451. https://doi.org/10.1371/JOURNAL.PONE.0065451.

(45) Zeng, E. Y.; Vista, C. L. Organic Pollutants in the Coastal Environment off San Diego, California. 1. Source Identification and Assessment by Compositional Indices of Polycyclic Aromatic Hydrocarbons. *Environ. Toxicol. Chem.* **1997**, *16* (2), 179–188. https://doi.org/10.1002/etc.5620160212.

(46) Elliott, J. E.; Elliott, K. H. Tracking Marine Pollution. *Science*. American Association for the Advancement of Science May 3, 2013, pp 556–558. https://doi.org/10.1126/science.1235197.

(47) Harayama, S.; Kishira, H.; Kasai, Y.; Shutsubo, K. Petroleum Biodegradation in Marine Environments. *Pet. Biodegrad. 63 J. Molec. Microbiol. Biotechnol* **1999**, *1* (1), 63–70.

(48) Gao, J.; Chi, J. Biodegradation of Phthalate Acid Esters by Different Marine Microalgal Species. *Mar. Pollut. Bull.* **2015**, *99* (1–2), 70–75. https://doi.org/10.1016/J.MARPOLBUL.2015.07.061.

(49) Baena-Nogueras, R. M.; González-Mazo, E.; Lara-Martín, P. A. Degradation Kinetics of Pharmaceuticals and Personal Care Products in Surface Waters: Photolysis vs Biodegradation. *Sci. Total Environ.* **2017**, *590–591*, 643–654. https://doi.org/10.1016/J.SCITOTENV.2017.03.015.

(50) Glibert, P. M.; Maranger, R.; Sobota, D. J.; Bouwman, L. The Haber Bosch-Harmful Algal Bloom (HB-HAB) Link. *Environ. Res. Lett.* **2014**, *9* (10), 105001. https://doi.org/10.1088/1748-9326/9/10/105001.

(51) Kahru, M.; Mitchell, B. G. Ocean Color Reveals Increased Blooms in Various Parts of the World. *Eos (Washington. DC).* **2008**, *89* (18), 170. https://doi.org/10.1029/2008EO180002.

(52) Dale, B.; Edwards, M.; Reid, P. C. Climate Change and Harmful Algal Blooms. *Ecol. Stud.* **2006**, *189*, 367–378.

(53) Prather, K. A.; Bertram, T. H.; Grassian, V. H.; Deane, G. B.; Stokes, M. D.; DeMott, P. J.; Aluwihare, L. I.; Palenik, B. P.; Azam, F.; Seinfeld, J. H.; Moffet, R. C.; Molina, M. J.; Cappa, C. D.; Geiger, F. M.; Roberts, G. C.; Russell, L. M.; Ault, A. P.; Baltrusaitis, J.; Collins, D. B.; Corrigan, C. E.; Cuadra-Rodriguez, L. A.; Ebben, C. J.; Forestieri, S. D.; Guasco, T. L.; Hersey, S. P.; Kim, M. J.; Lambert, W. F.; Modini, R. L.; Mui, W.; Pedler, B. E.; Ruppel, M. J.; Ryder, O. S.; Schoepp, N. G.; Sullivan, R. C.; Zhao, D. Bringing the Ocean into the Laboratory to Probe the Chemical Complexity of Sea Spray Aerosol. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (19), 7550–7555. https://doi.org/10.1073/pnas.1300262110.

(54) Paerl, H.; Huisman, J. Blooms like It Hot. *Science (80-.).* **2008**. https://doi.org/10.1126/science.1156721.

(55) Zhang, H.; Yee, L. D.; Lee, B. H.; Curtis, M. P.; Worton, D. R.; Isaacman-VanWertz, G.; Offenberg, J. H.; Lewandowski, M.; Kleindienst, T. E.; Beaver, M. R.; Holder, A. L.; Lonneman, W. A.; Docherty, K. S.; Jaoui, M.; Pye, H. O. T.; Hu, W.; Day, D. A.; Campuzano-Jost, P.; Jimenez, J. L.; Guo, H.; Weber, R. J.; De Gouw, J.; Koss, A. R.; Edgerton, E. S.; Brune, W.; Mohr, C.; Lopez-Hilfiker, F. D.; Lutz, A.; Kreisberg, N. M.;

Spielman, S. R.; Hering, S. V.; Wilson, K. R.; Thornton, J. A.; Goldstein, A. H. Monoterpenes Are the Largest Source of Summertime Organic Aerosol in the Southeastern United States. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (9), 2038–2043. https://doi.org/10.1073/PNAS.1717513115/SUPPL_FILE/PNAS.1717513115.SD01.RTF.

(56) De Sá, S. S.; Palm, B. B.; Campuzano-Jost, P.; Day, D. A.; Hu, W.; Isaacman-Vanwertz, G.; Yee, L. D.; Brito, J.; Carbone, S.; Ribeiro, I. O.; Cirino, G. G.; Liu, Y.; Thalman, R.; Sedlacek, A.; Funk, A.; Schumacher, C.; Shilling, J. E.; Schneider, J.; Artaxo, P.; Goldstein, A. H.; Souza, R. A. F.; Wang, J.; Mckinney, K. A.; Barbosa, H.; Alexander, M. L.; Jimenez, J. L.; Martin, S. T. Urban Influence on the Concentration and Composition of Submicron Particulate Matter in Central Amazonia. *Atmos. Chem. Phys* **2018**, *18*, 12185–12206. https://doi.org/10.5194/acp-18-12185-2018.

(57) Griffin, R. J.; Cocker, D. R.; Flagan, R. C.; Seinfeld, J. H. Organic Aerosol Formation from the Oxidation of Biogenic Hydrocarbons. *J. Geophys. Res. Atmos.* **1999**, *104* (D3), 3555–3567. https://doi.org/10.1029/1998JD100049.

(58) Mahilang, M.; Deb, M. K.; Pervez, S. Biogenic Secondary Organic Aerosols: A Review on Formation Mechanism, Analytical Challenges and Environmental Impacts. *Chemosphere* **2021**, *262*, 127771. https://doi.org/10.1016/J.CHEMOSPHERE.2020.127771.

(59) Kleindienst, T. E.; Corse, E. W.; Li, W.; McLver, C. D.; Conver, T. S.; Edney, E. O.; Driscoll, D. J.; Speer, R. E.; Weathers, W. S.; Tejada, S. B. Secondary Organic Aerosol Formation from the Irradiation of Simulated Automobile Exhaust. *http://dx.doi.org/10.1080/10473289.2002.10470782* **2011**, *52* (3), 259–272. https://doi.org/10.1080/10473289.2002.10470782.

(60) Hoyle, C. R.; Boy, M.; Donahue, N. M.; Fry, J. L.; Glasius, M.; Guenther, A.; Hallar, A. G.; Huff Hartz, K.; Petters, M. D.; Petäjä, T.; Rosenoern, T.; Sullivan, A. P. A Review of the Anthropogenic Influence on Biogenic Secondary Organic Aerosol. *Atmos. Chem. Phys.* **2011**, *11* (1), 321–343. https://doi.org/10.5194/ACP-11-321-2011.

(61) Artaxo, P.; Rizzo, L. V.; Brito, J. F.; Barbosa, H. M. J.; Arana, A.; Sena, E. T.; Cirino, G. G.; Bastos, W.; Martin, S. T.; Andreae, M. O. Atmospheric Aerosols in Amazonia and Land Use Change: From Natural Biogenic to Biomass Burning Conditions. *Faraday Discuss.* **2013**, *165* (0), 203–235. https://doi.org/10.1039/C3FD00052D.

(62) Martin, S. T.; Artaxo, P.; Machado, L. A. T.; Manzi, A. O.; Souza, R. A. F.; Schumacher, C.; Wang, J.; Andreae, M. O.; Barbosa, H. M. J.; Fan, J.; Fisch, G.; Goldstein, A. H.; Guenther, A.; Jimenez, J. L.; Pöschl, U.; Dias, M. A. S.; Smith, J. N.; Wendisch, M. Introduction: Observations and Modeling of the Green Ocean Amazon (GoAmazon2014/5). *Atmos. Chem. Phys* **2016**, *16*, 4785–4797. https://doi.org/10.5194/acp-16-4785-2016.

(63) Davidson, E. A.; De Araüjo, A. C.; Artaxo, P.; Balch, J. K.; Brown, I. F.; Mercedes, M. M.; Coe, M. T.; Defries, R. S.; Keller, M.; Longo, M.; Munger, J. W.; Schroeder, W.; Soares-Filho, B. S.; Souza, C. M.; Wofsy, S. C. The Amazon Basin in Transition. *Nat. 2012 4817381* **2012**, *481* (7381), 321–328. https://doi.org/10.1038/nature10717.

(64) Urrutia-Pereira, M.; Rizzo, L. V.; Chong-Neto, H. J.; Solé, D. Impact of Exposure to Smoke from Biomass Burning in the Amazon Rain Forest on Human Health. *J. Bras. Pneumol.* **2021**, *47* (5), 11. https://doi.org/10.36416/1806-3756/E20210219.

(65) Potter, C.; Brooks-Genovese, V.; Klooster, S.; Torregrosa, A. Biomass Burning Emissions of Reactive Gases Estimated from Satellite Data Analysis and Ecosystem Modeling for the Brazilian Amazon Region. *J. Geophys. Res. Atmos.* **2002**, *107* (D20), LBA 23-1. https://doi.org/10.1029/2000JD000250.

(66) Cirino, G.; Brito, J.; Barbosa, H. M. J.; Rizzo, L. V.; Tunved, P.; de Sá, S. S.; Jimenez, J. L.; Palm, B. B.; Carbone, S.; Lavric, J. V.; Souza, R. A. F.; Wolff, S.; Walter, D.; Tota, J.; Oliveira, M. B. L.; Martin, S. T.; Artaxo, P. Observations of Manaus Urban Plume Evolution and Interaction with Biogenic Emissions in GoAmazon 2014/5. *Atmos. Environ.* **2018**, *191*, 513–524. https://doi.org/10.1016/J.ATMOSENV.2018.08.031.

(67) Nascimento, J. P.; Bela, M. M.; Meller, B. B.; Banducci, A. L.; Rizzo, L. V.; Liduvino Vara-Vela, A.; Barbosa, H. M. J.; Gomes, H.; Rafee, S. A. A.; Franco, M. A.; Carbone, S.; Cirino, G. G.; Souza, R. A. F.; Mckeen, S. A.; Artaxo, P. Aerosols from Anthropogenic and Biogenic Sources and Their Interactions-Modeling Aerosol Formation, Optical Properties, and Impacts over the Central Amazon Basin. *Atmos. Chem. Phys.* **2021**, *21* (9), 6755–6779. https://doi.org/10.5194/ACP-21-6755-2021.

(68) Palm, B. B.; De Sá, S. S.; Day, D. A.; Campuzano-Jost, P.; Hu, W.; Seco, R.; Sjostedt, S. J.; Park, J. H.; Guenther, A. B.; Kim, S.; Brito, J.; Wurm, F.; Artaxo, P.; Thalman, R.; Wang, J.; Yee, L. D.; Wernis, R.; Isaacman-VanWertz, G.; Goldstein, A. H.; Liu, Y.; Springston, S. R.; Souza, R.; Newburn, M. K.; Lizabeth Alexander, M.; Martin, S. T.; Jimenez, J. L. Secondary Organic Aerosol Formation from Ambient Air in an Oxidation Flow Reactor in Central Amazonia. *Atmos. Chem. Phys.* **2018**, *18* (1), 467–493. https://doi.org/10.5194/ACP-18-467-2018.

(69) Rafee, S. A. A.; Martins, L. D.; Kawashima, A. B.; Almeida, D. S.; Morais, M. V. B.; Souza, R. V. A.; Oliveira, M. B. L.; Souza, R. A. F.; Medeiros, A. S. S.; Urbina, V.; Freitas, E. D.; Martin, S. T.; Martins, J. A. Contributions of Mobile, Stationary and Biogenic Sources to Air Pollution in the Amazon Rainforest: A Numerical Study with the WRF-Chem Model. *Atmos. Chem. Phys* **2017**, *17*, 7977–7995. https://doi.org/10.5194/acp-17-7977-2017.

(70) Shrivastava, M.; Andreae, M. O.; Artaxo, P.; Barbosa, H. M. J.; Berg, L. K.; Brito, J.; Ching, J.; Easter, R. C.; Fan, J.; Fast, J. D.; Feng, Z.; Fuentes, J. D.; Glasius, M.; Goldstein, A. H.; Gomes Alves, E.; Gomes, H.; Gu, D.; Guenther, A.; Jathar, S. H.; Kim, S.; Liu, Y.; Lou, S.; Martin, S. T.; Mcneill, V. F.; Medeiros, A.; De Sá, S. S.; Shilling, J. E.; Springston, S. R.; Souza, R. A. F.; Thornton, J. A.; Isaacman-Vanwertz, G.; Yee, L. D.; Ynoue, R.; Zaveri, R. A.; Zelenyuk, A.; Zhao, C. Urban Pollution Greatly Enhances Formation of Natural Aerosols over the Amazon Rainforest. *Nat. Commun.* **2019**. https://doi.org/10.1038/s41467-019-08909-4.

(71) Mouchel-Vallon, C.; Lee-Taylor, J.; Hodzic, A.; Artaxo, P.; Aumont, B.; Camredon, M.; Gurarie, D.; Jimenez, J. L.; Lenschow, D. H.; Martin, S. T.; Nascimento, J.; Orlando, J. J.; B. Palm, B.; Shilling, J. E.; Shrivastava, M.; Madronich, S. Exploration of Oxidative Chemistry and Secondary Organic Aerosol Formation in the Amazon during the Wet Season: Explicit Modeling of the Manaus Urban Plume with GECKO-A. *Atmos. Chem. Phys.* **2020**, *20* (10), 5995–6014. https://doi.org/10.5194/ACP-20-5995-2020.

2 Ch3MS-RF: A Random Forest Model for Chemical Characterization and Improved Quantification of Unidentified Atmospheric Organics Detected by Chromatography-Mass Spectrometry Techniques

This work is adapted from:
Emily B. Franklin, Lindsay D. Yee, Bernard Aumont, Robert J. Weber, Paul Grigas, Allen Goldstein, "Ch3MS-RF: A Random Forest Model for Chemical Characterization and Improved Quantification of Unidentified Atmospheric Organics Detected by Chromatography-Mass Spectrometry Techniques," *Atmospheric Measurement Techniques* (**2022**)

2.1 Abstract

The chemical composition of ambient organic aerosols plays a critical role in driving their climate and health relevant properties and holds important clues to the sources and formation mechanisms of secondary aerosol material. In most ambient atmospheric environments, this composition remains incompletely characterized, with the number of identifiable species consistently outnumbered by those that have no mass spectral matches in the literature or NIST/NIH/EPA mass spectral databases, making them nearly impossible to definitively identify. This creates significant challenges in utilizing the full analytical capabilities of techniques which separate and generate spectra for complex environmental samples. In this work, we develop the use of machine learning techniques to quantify and characterize novel, or unidentifiable, organic material. This work introduces Ch3MS-RF (Chemical Characterization by Chromatography-Mass Spec Random Forest Modelling), an open-source R-based software tool for efficient machine-learning enabled characterization of compounds separated in chromatography-mass spec applications but not identifiable by comparison to mass spectral databases. A random forest model is trained and tested on a known 130 component representative external standard to predict the response factors of novel environmental organics based on position in volatility-polarity space and mass spectrum, enabling reproducible, efficient, and optimized quantification of novel environmental species. Quantification accuracy on a reserved 20% test set randomly split from the external standard compound list indicate that random forest modelling significantly outperforms the commonly used methods in both precision and accuracy, with a median response factor % error of -2% for modelled response factors compared to >15% for typically used proxy assignment-based methods. Chemical properties modelling, evaluated on the same reserved 20% test set as well as an extrapolation set of species identified in ambient organic aerosol samples collected in the amazon rainforest, also demonstrates robust performance. Extrapolation set property prediction mean absolute errors for carbon number, oxygen to carbon ratio (O:C), average carbon oxidation state ($\overline{OS_c}$), and vapor pressure are 1.8, 0.15, 0.25, and 1.0 (log(atm)), respectively. Extrapolation set Out-of-Sample R^2 for all properties modelled are above 0.75, with the

exception of vapor pressure. While predictive performance for vapor pressure is less robust compared to the other chemical properties modelled, random forest-based modelling was significantly more accurate than other commonly used methods of vapor pressure prediction, decreasing mean vapor pressure prediction error to 0.24 (log(atm)) from 0.55 (log(atm)) (chromatography-based vapor pressure prediction) and 1.2 (log(atm)) (chemical formula-based vapor pressure prediction). The random forest model significantly advances untargeted analysis of the full scope of chemical speciation yielded by GCxGC-MS techniques and can be applied to GC-MS as well. It enables accurate estimation of key chemical properties commonly utilized in the atmospheric chemistry community, which may be used to more efficiently identify important tracers for further individual analysis and to characterize compound populations uniquely formed under specific ambient conditions.

2.2 Introduction

Organic aerosols play a critical role in global radiative forcing and regional aerosol-attributable public health concerns, making up a significant (20-90%) fraction of fine particulate matter around the globe.[1] This organic material is highly complex in terms of chemical composition and constantly changing; Goldstein and Galbally, 2007 estimates the number of gas and aerosol-phase atmospheric organic constituents to lie in the millions, while Ditto et al., 2018 reports molecular level variability of 60-80% between consecutive samples collected at fixed sites for samples comprised of high thousands of resolvable species. While there has been significant progress towards achieving mass closure of atmospheric reactive carbon using an ensemble of both bulk and speciated measurement techniques over the past two decades, speciated and isomer identified mass closure remains challenging.[4–6] A comprehensive review of the challenges and utility of different levels of molecular identification, Nozière et al., 2015, compares the utility of many types of incomplete identification of atmospheric organic compounds, but defines that "An organic compound is fully identified only if its molecular structure is entirely known, including its isomeric and spatial (stereo) configuration." Important chemical information can be gleaned from formula-based identifications and bulk characterization, but isomer-specific identifications provide critical atmospheric chemistry-relevant insights. As described in Isaacman-Vanwertz and Aumont, 2021, different isomers of the same chemical formula vary over orders of magnitude in volatility and Henry's constant, and by a factor of 2 in reactivity with the hydroxyl radical, all critical properties for characterization of aerosol formation and properties. Isomer specific identifications also play a crucial role in elucidation of important chemical reaction mechanisms.

Gas chromatography coupled with electron ionization mass spectrometry (GC-MS) is a commonly utilized technique for isomer specific speciation of atmospheric constituents. Observed ambient species may be matched to authentic standards or mass spectral database entries by both retention index (chromatographic elution time relative to that of a series of alkanes) and mass spectrum. Two dimensional gas chromatography (GCxGC-MS), a methodologically similar technique which achieves advanced separation by passing compounds through multiple GC columns configured for different chemical properties,

increases the scope of isomer-specific identification by separating species that would coelute in single dimension GC-MS applications.[9–11] However, a significant challenge of fully utilizing the data from these techniques is the novelty and diversity of the atmospheric constituents; most observed organic species have never been synthesized and are not in any mass spectral library and are therefore not directly identifiable from GC-MS or GCxGC-MS techniques. Although the size of mass spectral libraries are rapidly increasing, with ~30,000 new compounds added to the NIST/EPA/NIH mass spectral database between the 2011 and 2014 versions (bringing the number of compounds catalogued in NIST14 EI library to ~250,000), the numbers of identifiable constituents in typical atmospheric samples remain low.[12] As described in Hamilton et al., 2004, in an urban aerosol sample analyzed by GCxGC-MS, of >10,000 unique observed species, fewer than 2% were identifiable from authentic standard or mass spectral matching. Low numbers of matched relative to novel ambient species persist; Worton et al., 2017 finds that fewer than 35% of ~500 compounds isolated from aerosol samples collected at a forested site match mass spectral database entries, while this work (as later described) finds that fewer than 10% of ~1500 aerosol phase organic species can be matched to published spectra. As described in Worton et al., 2017, species that cannot be identified are often not included in GC-MS and GCxGC-MS-based analyses, meaning that the majority of acquired data is not fully utilized. Note that in accordance with the definition of complete molecular identification previously quoted from Nozière et al., 2015, "unidentified" compounds are from here on defined as any species that is not identifiable by comparison (on the basis of retention index and mass spectrum) to either authentic standards or mass spectral database entries of positively identified species. Pairing GC-EI-MS systems with complementary measurements such as chemical ionization (described in Bi et al., 2021) or switching to softer election ionization techniques (specifically through employing 14 eV vacuum ultraviolet rather than traditional 70 eV EI, intended to preserve sufficient precursor ion mass for formula identification, as described in Worton et al., 2017) can enable more separated but unidentified compounds to be characterized by formula identification, even where isomer-specific identification remains elusive. That said, these instrumental configurations are rare, and fragmentation under 14 eV is still sufficiently significant to leave many species' formulae not identifiable and therefore still uncharacterized.[11] Recent efforts to embrace a larger fraction of the full complexity of chemical information yielded by highly speciated organic aerosol measurements (on the scale of low to mid 100's of compounds) have categorized unidentified species by likely source groups and chemical families through time series correlations with known tracer species[15] or by manual group assignments by individual researcher judgements based on mass spectral features.[16] These methods are difficult to standardize and reproduce and become prohibitively inefficient when pushing towards the full chemical complexity of speciated observations produced from typical atmospheric samples, which extend into the low to mid thousands of species.

Quantification of unidentified compounds faces similar challenges. Where possible, compounds in GC and GCxGC-MS are directly quantified by calibration curves of authentic standards, but direct quantifications are limited by standard expense and availability, even for species that can be positively identified. Compounds that cannot be

directly quantified, both in GCxGC-MS and in GC-MS, are most commonly quantified by assigning quantification factors from compounds resolved closely in chromatographic space, compounds that are identified as sharing chemical structures, or some interpolation of multiple nearby proxies.[15–18] The errors associated with these assignments/choices are usually estimated from the range of quantification factors of close or chemically similar species and are assumed to be high (up to a factor of 2 depending on degree of certainty in assigning chemical class as described in Jen et al., 2019 and Liang et al., 2021). To our knowledge, this work presents the first quantitative error analysis of these techniques based on applying proxy quantification techniques to compounds with known quantification factors.

Current manual characterization and quantification proxy-assignments are essentially an exercise of pattern recognition, as researchers use experience in analysing spectra and position in chromatographic space to categorize or otherwise characterize unidentifiable species. Given the scale of the novel compound characterization challenge (on the order of hundreds to thousands of species for a given sampling location using current methods), transitioning to automated characterization methods will be necessary to keep up with data acquisition, and will offer co-benefits in increased reproducibility and reduced susceptibility to researcher biases. Decision tree-based machine learning methods including gradient boosting and random forests have demonstrated robust performance in pattern recognition-based regression applications including nonlinear features across a wide range of fields.[19,20] Random forests, a decision tree-based method which generates predictions based on a combination of diverse trees generated by randomized feature selection and resampling on a training data set,[21] are particularly suited to this application and intended audience. They have demonstrated robust performance across a range of applications, including predictions of chemical properties[22] and do not require extensive hyperparameter tuning to achieve high performance.[20] In this work, we develop machine learning models, specifically based on the random forests methodology, that use chromatographic and mass spectral feature inputs to predict a diverse suite of chemical properties, including quantification factor in a TD-GCxGC-MS system, oxygen to carbon ratio (O:C), carbon number, average carbon oxidation state ($\overline{OS_c}$), and vapor pressure. Coinciding with this manuscript, we have released a repository template including an Rmarkdown notebook (https://github.com/ebarnesey/Ch3MS-RF) that enables users with general atmospheric chemistry background, who do not necessarily have special expertise in machine learning data science applications, to tailor our analysis for their specific use cases. As such, robust performance evaluation and ease of applicability to a range of potential use cases are emphasized over extensive application-specific hyperparameter tuning.

In summary, this work aims to provide the GC-MS and GCxGC-MS atmospheric chemistry community with tools to achieve the following objectives:

1) Enable accurate chemical characterization of organic constituents separated in gas chromatographic space but not necessarily published (in mass spectral databases)

2) Improve the quantification accuracy for species that cannot be directly calibrated using authentic standards

2.3 Instrumentation and Data

2.3.1 Calibration Curves Using an External Standard Mixture of Authentic Standards

A custom calibration standard mixture (referred to hereafter as "external standard") was created containing ~130 unique authentic standards selected for maximal coverage of the compounds and compound classes typically observed in atmospheric regions with significant biogenic emissions, as well as influences from anthropogenic activities and biomass burning. The selection of these standard species was informed by previous work targeting similar sample types using the same instrumentation [9,15,23], and covers species including sugars, PAH's, and both monoterpene and isoprene oxidation products. In addition to commercially available external standards, 6 sesquiterpene oxidation products were custom synthesized by collaborators (as described in Bé et al., 2019) for expanded coverage of potentially important chemical tracers. The full list of standard components can be found in Table 2.A1, and the standard property distribution in volatility-polarity space is illustrated in Figure 2.2. The standard was prepared from pure components immediately prior to sample analysis in 1:1 methanol:chloroform solution, replicating the methodology utilized in Zhang et al., 2018. Standards were introduced to the instrument by injecting onto tissuquartz filter material to maximize consistency between filter samples (organic aerosol also collected on tissuquartz filters) and calibration runs. At 5 points throughout sample analysis, 6-point calibration curves (5 loaded points and a zero point) were performed to determine the "quantification factors" (internal standard normalized signal/ng compound) of each external standard species. The internal standard, described in detail in section 2.3.1, is a solution of ~30 deuterated organics applied identically to all sample and calibration analysis runs to enable correction for instrument condition and matrix effects. For efficiency, outlier calibration points (significantly deviating from the slope of other points in the quantification factor, which are often caused by coelution with a contaminant) were removed. A minimum of 3 calibration points above the zero point was maintained to ensure robust quantification factors.

2.3.2 GoAmazon Field Data

The ambient extrapolation data utilized in this work originates from the Green Ocean Amazon (GoAmazon) field campaign which was conducted in central Amazonia in 2014. This campaign and the collection of ambient filters for offline analysis are described in detail in Martin et al., 2016, 2017 and Yee et al., 2018. Briefly, the campaign was conducted at a semi remote site occasionally downwind of the city of Manaus and periodically impacted by smoke from biomass burning. The campaign spanned two intensive operating periods, one during the Amazonian wet season (February through March) and one during the dry season (August through early October). Submicron aerosol samples were collected on tissuquartz filters (Pallflex), stored in pre-baked foil, double

contained in sealed mylar bags, and frozen prior to analysis. The samples were analysed by TD-GCxGC-EI-ToF-MS, as described below.

2.3.3 Instrumentation: TD-GCxGC-EI-ToF-MS

Both external standard species (during calibration runs) and GoAmazon filter samples were analysed by thermal desorption two-dimensional gas chromatography coupled with electron ionization time-of-flight mass spectrometry (TD-GCxGC-EI-ToF-MS, hereafter abbreviated GCxGC-MS). This instrumentation is described in detail in Worton et al., 2017, and instrument specifics including sub-component models, column materials, and temperature settings are described in Franklin et al., 2021. For ambient filter samples, 0.4 cm^2 aliquots of filter material are directly introduced into the instrument. Standards are stored in solution and introduced by injection onto pre-baked quartz filter material. An internal standard (described in section 2.3.3.1) is applied on top of the sample or external standard filter aliquots immediately prior to analysis. Briefly, the instrument functions as follows: a thermal desorption oven heats filter material, causing analytes and standards to evaporate into a flow of helium. The desorbed components are focused on a cooled inlet system (Gerstel CIS), which at the end of the thermal desorption cycle is rapidly heated to simultaneously release all organic species onto the head of the first column. Compounds are separated by both volatility and polarity by two gas chromatography columns in sequence, with the transition of compounds from the first to the second column modulated by a cryogenic focus and rapid thermal release system. Separated analytes are ionized by 70 eV electron ionization (EI) and detected by HR-ToF-MS (TOFWERK, EI-HTOF), with a resolving power of 4000 acquired at 100 Hz. While the mass spectra produced by this technique are high resolution, these high resolution mass spectra are converted to unit mass resolution spectra to increase the applicability of this technique to unit mass resolution techniques. The vertical (polar) axis of separation is extremely short relative to the horizontal (volatility) axis separation with a vertical stride length of 2.3 seconds compared to a retention time of ~ 1 hr for low volatility organics. As a result, GCxGC-MS deuterated alkane normalized retention indices are directly comparable to retention indices (or, with a linear conversion to non-deuterated retention indices, kovats indices) in single dimension GC-MS applications. This instrument's volatility range spans approximately C_{13}-C_{36} *n*-alkane volatility equivalents, covering the atmospherically important transition regime between IVOC (intermediate volatility organic carbon) and LVOC (low volatility organic carbon) species.

During the thermal desorption process, the carrier flow of helium is enriched with the derivatization agent MSTFA (n-methyl-n-trimethylsilyl-trifluoro-acetamide). This silylating reagent replaces the active hydrogen of polar OH groups with a trimethylsilyl group, $-Si(CH_3)_3$, a process which significantly enhances the recovery of polar organics. This approach is critical to increase the scope and degree of oxygenation of species recovered by thermal desorption-gas chromatography techniques [28]. However, it poses some challenges for data interpretation for diverse, complex, and novel chemical mixtures, because in the case of many polar species, the compound that is separated and detected by the GCxGC-MS instrumentation has been chemically altered from the species that was

collected. This can create challenges in compound identification, as not all species have published derivatized spectra, as well as challenges for mapping chemical properties onto the GCxGC-MS space, as the volatility-polarity distributions of derivatized compounds do not directly reflect their underivatized properties.

2.3.3.1 Internal Standard Normalization

Both filter samples and external standard impregnated filters (for calibration curves) were doped with a custom 23 component deuterated internal standard covering the full range of volatility sensitivity and a broad variety of functional group types immediately prior to analysis. The internal standard enables normalization for matrix effects, configuration of retention indices relative to the elution times of a deuterated alkane series, and normalization for instrument condition drift for improved consistency and quantification accuracy throughout intensive instrument use. In prior methods, the selection of internal standard involved either1) assigning each analyte an internal standard nearest in chromatographic space (by retention times) or 2) manual assignment of analytes to their most chemically similar internal standards regardless of proximity in GCxGC space. Analyte signal would then be normalized (divided) by the signal of the selected internal standard obtained during the same chromatographic run. In a new approach employed in this work, in order to maximize the reliability and consistency of normalization across a large number of samples and complex sample media, internal standard signals were each normalized by their own mean signals (throughout the entire analysis period) to yield an indicator of self-normalized instrument sensitivity. Analyte signal was then normalized by the mean self-normalized responses of the three closest internal standard species. This approach has multiple benefits. First, the responses of sample or external standard compounds are not artificially deflated or inflated due to their proximity to internal standard compounds that have higher or lower sensitivities based on their functional groups and derivatization. Second, this approach enables inclusion and utilization of incomplete data; in previous approaches, if an internal standard cannot be recovered in every sample it cannot be used for normalization, as this would create inconsistencies for the species that are otherwise assigned to that compound. Compounds at the very high and very low ends of the volatility space are chemically important but detectable at baseline low levels that can drop below limits of detection during periods of low sensitivity. Having to discard these species due to a few instances of missing corresponding internal standard data causes losses of valuable information. Finally, this approach decreases analysis sensitivity to any errors and noise in internal standard identification or isolation, as erroneously high or low individual internal standard responses are moderated by averaging with the other nearby internal standard species. Volatility-based sensitivity corrections, which can be achieved by raw internal standard normalization, were achieved in this work through normalization by an external standard-determined response curve, as described in “Featurization and Target Selection for Quantification Modelling.”

2.4 Data Preparation and Featurization

The analytical pipeline for data preparation through performance evaluation of this random forest modelling work is illustrated in Figure 2.1. The processes and decision

making around featurization, feature selection, and target selection for both chemical properties modelling and quantification modelling, as well as the curation of the training, test, and extrapolation data sets, is described below.

2.4.1 Featurization, Feature Selection, and Target Selection

As the aim of this work is to develop methods that can be applied to novel species not included in mass spectral databases, features utilized in this analysis rely solely upon the information readily available for unidentifiable species. Given the size and complexity of the intended use data suites, features must also be automatically generatable from the instrument data output and not rely upon any visual or manual categorization by researchers. In order to make these models more broadly useful to the atmospheric community, less common features produced by the GCxGC-MS instrumental setup (e.g. second dimension retention time and high resolution spectra) are not utilized for chemical properties modelling in order to increase the method's applicability to single dimension GC-MS systems and instruments with lower resolution mass spectra.

2.4.1.1 Mass Spectral Featurization

The only chemical information directly produced by GCxGC-MS for unidentified organic species are their locations in GCxGC volatility-polarity space and mass spectra. These sources of information are therefore exclusively utilized in creating and selecting the features for chemical properties modelling. The retention index of each compound was directly utilized as a feature, but the mass spectra require interpretation in order to be used.

The unit mass resolution spectra utilized in this analysis include each charged fragment represented by its measured mass to charge ratio (*m/z*) and a relative signal score out of 1000 (normalized by the most abundant fragment's peak signal). EI is a high energy or "hard" ionization technique which typically leaves only a small fraction of molecular ions intact and creates positively charged ion fragments that are almost all singly charged, with any multiply charged ions at extremely low abundance. This means that the molecular formulae cannot generally be directly determined from the mass spectrum, even when the spectra are high resolution, and measured ions can be assumed to have a single charge. That said, the m/z of charged fragment ions yield useful information into chemical characteristics and functional groupings that can provide critical chemical information; for example, a peak at *m/z*=73 corresponds to a fragment of $Si(CH_3)_3^+$, a derivatization fragment which indicates that the ionized compound contained an OH group which was derivatized (see section 2.3.3). The mass differences between charged peaks also represent important pieces of information, as they can indicate losses of uncharged molecular fragments that similarly point to the structure and characteristics of the original compound. It is important to note that not all neutral mass differences between charged peaks can be interpreted as direct neutral losses as not all high m/z charged fragments directly fragment onto lower m/z charged fragments in a manner that can be directly interpreted from neutrally charged fragment losses. However, frequently occurring neutral differences may still hold value in reflecting a common coordination of neutral loss processes.

The greatest chemical information lies in features that exist in an intermediate range of occurrence frequency in the data set. A feature which appears in all the training species does not provide any useful information in predicting properties of the test species. Neither does a feature which is totally unique to a single species, as it does not provide any information on patterns which can be used to adjust prediction of properties for other species. This logic can be applied to mass spectral featurization; while it would be possible to convert every m/z to a feature and so input the entire raw mass spectrum of each compound as a series of features for the random forest model, this approach would be inefficient, open to error introduced by noise, and miss the important information provided by neutral mass differences between charged fragments.

Multiple approaches for mass spectral featurization were tested to optimize the number of features and representation of features. Given the final choice in model structure (random forest, as described in section 4), inclusion of covarying features or more features than necessary did not introduce significant sources of error. Target-specific feature restriction based on importance is discussed in section 4. The final mass spectral featurization method selected for this analysis, a simplified adaptation of methodology described in Eghbaldar et al., 1998, was as follows: the top 5 charged fragments (mass spectral peaks) from each training set mass spectrum are selected. The mass differences between these 5 peaks (a maximum of 10 numbers, if all fragments occur at differently spaced *m/z*) were then compiled into a list of "neutral losses". The charged fragment lists and the neutral loss lists of all training set external standard compounds were next combined in a frequency list, with each charged fragment or neutral loss quantified by frequency (how many compounds in the external standard test set exhibited that charged fragment or neutral loss among their top 5 peaks). The top 40 most common charged fragments and top 20 most common neutral losses were converted into features. The identities of these 40 most common fragments and 20 most common neutral mass differences (along with possible identities and notes) can be found in tables 2.A2 and 2.A3, respectively. The mass spectra of all training, test, and extrapolation set compounds were then simplified using the previously described method (top 5 peaks extracted and mass differences between those peaks calculated). Each *m/z* feature was assigned the normalized signal of that peak in the mass spectrum if the feature *m/z* was one of the top 5 peaks; otherwise, it was set to zero. Each neutral loss feature was assigned true or false for each compound depending on whether the neutral loss appeared in the mass differences between the 5 most significant peaks. An example mass spectral featurization for the example compound hexadecane can be found in Table 2.A4, and the mass spectral featurization process is included in the open-source R script accompanying this publication.

2.4.1.2 Target Selection for Chemical Properties Modelling

The goal of chemical properties modelling is to enable inclusion of "unidentified" species in aerosol organic analysis that has previously been restricted to species for which the identity or at least chemical formula is known. One way in which complex organic mixtures are visualized and analyzed is through orientation of observed species in chemical properties spaces that have been developed and broadly utilized in the field of aerosol

science. Two such spaces include the Volatility Basis Set (VBS [30]) and the visualization by average carbon oxidation state and carbon number developed in Kroll et al., 2011, hereafter referred to as $\overline{OS_c}$-n_c space. Compounds can be plotted in VBS space by their O:C or $\overline{OS_c}$ (average carbon oxidation state[31]) against some measure of volatility, either log(Vapor Pressure) or log(C_0), where C_0 is the pure component sub-cooled liquid vapor pressure in atm. In $\overline{OS_c}$-n_c space, compounds are plotted by their average carbon oxidation state ($\overline{OS_c}$) against carbon number. The ability to map novel or unidentifiable compounds in these spaces would provide critical information about the properties of the individual species, enable identification of groups of chemically distinct novel compounds deserving particular consideration, and more completely visualize the distribution of chemical characteristics for complex mixtures and potential routes of chemical transformation (e.g. oligomerization, functionalization, fragmentation) beyond the identifiable components. With these goals in mind, the properties selected to be the targets of these modelling efforts were number of carbons (n_c), O:C, $\overline{OS_c}$, and vapor pressure.

Carbon number, O:C, and $\overline{OS_c}$ (based on the equation in Kroll et al., 2011) can all be directly calculated from chemical formula, which was known for each standard and ambient extrapolation compound (see section 2.4.2). Vapor pressure is not directly calculatable from chemical formula and not all identified compounds in the external standard and extrapolation data sets have reliable experimental vapor pressure measurements available, so structurally-based vapor pressure predictions are utilized instead. Isaacman-Vanwertz and Aumont, 2021 finds that of all structure-based vapor pressure prediction methods available, the average of predictions generated by the EVAPORATION,[32] Nannoolal,[33] and Simpol[34] models yields the most accurate vapor pressure prediction. These methods were therefore utilized to predict the vapor pressures of all standard and extrapolation set compounds, and the average structurally predicted vapor pressures were utilized as the "true" vapor pressures for model training and evaluation. Seven of the external standard test set species and fifteen of the extrapolation set species were incompatible with the prediction capabilities of one or more of the three structural vapor pressure prediction methods (most frequently due to functional group types for which the models are not parameterized) and were therefore not utilized in performance analysis. Two additional potential targets, double bond equivalent and H:C ratio, were tested but failed to produce sufficiently robust property predictions.

The final components of the chemical properties random forest models are as follows:

Targets: Carbon number, $\overline{OS_c}$, O:C, vapor pressure (structurally modelled)

Features: Retention index, 40 feature representation of mass spectral charged fragments, 20 feature representation of neutral mass differences between charged fragments

A table listing the entire set of input features for chemical properties modelling of the example compound Hexadecane can be found in table 2.A4, and the instrument-produced mass spectrum for this species can be found in Figure 2.A1.

2.4.1.3 Featurization and Target Selection for Quantification Modelling

Compound quantification factor is significantly and reliably related to retention index across all compound classes tracked, but this relationship is not linear and changes much more rapidly in some retention index windows than others. This phenomenon, caused by incomplete cold inlet trapping of species in the most volatile sensitivity region and incomplete thermal desorption of species in the least volatile sensitivity region, is illustrated in Figure A2 and is consistent with findings presented in Zhang et al., 2018. A variety of retention index corrections were tested, including the following: a) factorizing the retention indices of each compound (rounded to the nearest 100) and including as a feature in model training and testing, and b) normalizing (dividing) each compound by the raw signal of its nearest deuterated alkane internal standard, the method utilized in Zhang et al., 2018. Both methods however performed poorly in the 1600-1900 RI range, where response increases extremely rapidly with RI (Figure 2.A2). The most reliable normalization method and the method selected for this analysis was normalizing (dividing) all compound quantification factors by the average response curve for alkanes, defined by the combination of 2 best fit exponential curves, which intersect at RI ≈ 1950 as illustrated in Figure A2, and training on/predicting this normalized response factor rather than the raw quantification factor. The r^2 of the exponential fit of individual calibration period quantification factors around the response curve in the volatile region is .77, while the r^2 of the curve describing the less volatile region is .65. Note that these fits take into account each quantification factor of each calibration window, and are therefore influenced by the variations in the measured quantification factors of the same compounds measured at different points throughout analysis. RI-normalized response factors were translated back to predicted quantification factors for performance evaluation, as other methods of quantification do not utilize this normalization method.

Unlike in the case of chemical properties modelling, quantification modelling performance was significantly improved by inclusion of second dimension retention time information, and it was therefore included as a feature in response factor prediction. As a result, this approach in its current form is only usable by GCxGC-MS applications, but could be adapted to single dimension chromatography-mass spec.

In this analysis, continuous measurement periods (consecutively collected samples) were analyzed in sequences bounded by calibration curve runs. To preserve the quantification continuity in these consecutive measurements and avoid step changes in calculated concentration that might occur due to switching between quantification factors, the two quantification factors bookending an analysis period are averaged to assign the quantification factors for samples run in that interval. To replicate this approach, the compound quantification factors were sequentially averaged to yield 5 quantification periods (the final calibration curve experienced an instrument failure, and the last calibration period is therefore based solely on the final curve).

The mass spectral featurization is described in “Mass Spectral Featurization” above.

The final components of the quantification model are as follows:

Target: Normalized response factor (RI curve-normalized, calibration period averaged)

Features: Retention index, second dimension retention time, calibration period, 40 feature representation of mass spectral charged fragments, 20 feature representation of neutral mass differences between charged fragments.

2.4.2 Training, Test, and Extrapolation Set Curation

To generate a training and test set from the external standard data, each external standard was assigned to a chemical group (alkane, sugar, PAH, etc), and the list of external standard compounds was randomly split 80:20 (80% of compounds in the training set, 20% in the test set) maintaining the ratios of different chemical groups. 200 possible splits were generated, and the split which demonstrated the greatest similarity in median retention index and median second dimension retention time between the test and training sets was selected to avoid potential extrapolation problems that might occur with a highly skewed distribution of test and training compounds across the GCxGC space. This process is documented in Supporting Information.

The extrapolation set was curated from the compounds isolated from the GoAmazon samples by comparing the spectra and retention indices of compounds to the external standard and matches in the NIST14 mass spectral database. Of the ~1500 unique compounds identified across 11 template samples, 63 were determined to match external standard compounds and an additional 71 compounds were identifiable from the NIST library due to high (>800[11]) mass spectral match factor and retention index agreement with database entries. Based on number of silicon atoms in the assigned formulae from the NIST identification, each chemical formula was converted to its underivatized form. Only the 71 compounds that were identifiable from the NIST library but not from external standards were included in the extrapolation set to ensure that performance metrics for the extrapolation set would not be skewed by the inclusion of species that may have been in the training data, and to ensure that the test set and extrapolation set performance evaluations would be entirely independent. 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. As illustrated in Figure 2.2, the distribution of training, test, and extrapolation set species utilized in this work effectively span the distribution of unknown compounds in GCxGC volatility-polarity space.

2.5 Model Selection, Training, and Tuning

The number and complexity of input features and lack of clear linear relationships between target properties and input features in this analysis is well suited to a decision tree-based analytical approach.[19,20] Random forest and gradient boosting methods were both preliminarily tested for response factor prediction. Random forests demonstrated slightly better performance and was selected for this and additional methodological reasons, as follows. Random Forests are more robust to overfitting than gradient boosting, which is a particular concern in this case given the small number of training compounds (~100) compared to the large numbers of novel environmental organics that are the intended subjects of unverifiable modelling. Additionally, random forests perform well using the default settings and do not require extensive tuning to achieve optimal performance [20]. As the aim of this work is to produce models that the atmospheric science community, including non-experts in machine learning, can easily implement for novel compound analysis, this robustness and simplicity is a significant advantage.

The training and tuning processes for chemical properties prediction are visualized in Figure 1. For each target property, the model was trained on the external standard training set data, the curation of which is described above. As previously referenced, random forests do not require extensive tuning, and for ease of use reasons most parameters were maintained at their default values. Tuning primarily focused on feature restriction. Feature restriction to enforce tree diversity (mtry) was optimized by 5-fold cross validation, with the mtry value that minimized mean absolute error (MAE) selected. Although random forest modelling is comparatively not influenced by the inclusion of features that do not contribute significant predictive capabilities, the inclusion of unnecessary features can contribute to overfitting of the training data which decreases prediction performance for the test and extrapolation data sets. To address this problem, the feature importance (a measure of increase in node purity when this feature is used in a split) of each input feature was extracted from the original predictive model. The importance metrics were normalized by the total importance of all features to generate a percent importance score for each feature. Importance distributions were highly skewed, with a relatively low number of features contributing the majority of decrease in node purity. Features that contributed less than 1% to the total importance score were removed, and the model was re-trained on only the important features. Extrapolation set performance improvements from removal of low importance features was low, with an improvement in OSR^2 (out of sample R^2, defined in detail in section 2.6.1) on the order of 0-0.03. This indicates that this step is not crucial for chemical properties or quantification factor prediction. The cross validation-optimized mtry number, number of important features, and identity of important features for the chemical properties models (one optimized model per property predicted) are summarised in Table 2.1. For quantification modelling, mtry is optimized at 44 features and 46 features meet an importance criterion of >1 %.

2.6 Model Performance Evaluation

2.6.1 Chemical Properties Modelling Performance

Three performance metrics are utilized to evaluate target predictions for the four chemical properties models. The first, out-of-sample r^2 (OSR^2), provides a measure of how significantly a model improves upon a baseline assumption that all target property values are equal to the mean of those values in the training data. It approaches a maximum of 1 for perfect predictions. The second metric, mean absolute error (MAE) provides the mean absolute prediction residual in the units of the target property. This metric is particularly important, as it provides a benchmark for prediction accuracy which can be translated into visualization and utilized to determine which applications are appropriate given prediction errors. The final performance metric, root mean square error (RMSE), is also a scale dependent error metric and provides the quadratic mean of prediction residuals. The equations for these metrics are provided below:

$$OSR^2 = 1 - \frac{\sum_{i=1}^{n}(T_i - P_i)^2}{\sum_{i=1}^{n}(T_i - \bar{R}_T)^2} \quad (1)$$

$$MAE = \frac{\sum_{i=1}^{n}|P_i - T_i|}{n} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - T_i)^2}{n}} \quad (3)$$

In this notation, for each test or extrapolation set compound i summed across a population of n compounds, T_i indicates the true value of the property being tested, P_i indicates the predicted value of that property, and $\bar{R}_T$ indicates the mean of the selected property in the training data set.

The prediction performance for the tuned and trained chemical properties model are evaluated independently on both the external standard test set (Figure 2.3, Table 2.2) and the ambient sample extrapolation set (Figure 2.4, Table 2.3). Both of these performance evaluations are important for different reasons. The external standard contains many series of highly chemically similar species (for example alkane and carboxylic acid series), meaning that the test set is likely to be more chemically similar to the training set than a real distribution of ambient organic species would be. Performance evaluation on the extrapolation set therefore provides a more realistic assessment of likely prediction accuracies on the large number of novel ambient organic compounds that are the intended focus of this modelling effort. That said, prediction performance on the external standard test set also yields important information. The external standard is designed to cover the entire space of anticipated chemical features for the environmental samples and is therefore more diverse relative to the number of compounds included compared to the extrapolation set (which is primarily CHO-type compounds). Performance evaluation on the external

standard test set therefore yields more information about model performance across a broad suite of compound classes.

2.6.1.1 Test Set Performance Evaluation

By all evaluation metrics applied (summarised in Table 2.2), performance for carbon number, O:C, carbon oxidation state, and log(VP) predictions on the external standard test set are robust. The O:C and carbon number predictions are particularly strong, with OSR^2 of .89 and .88 respectively and mean absolute errors of .072 element ratio units and 1.8 carbon number units. For context, given the range in true values from O:C= 0-1 and carbon number = 4-31, both mean absolute errors are approximately 7% of the range of measured values. For $\overline{OS_c}$ and vapor pressure, the mean absolute errors normalized by the measurement range are both approximately 12%. As illustrated Figure 2.3, this means that the distribution of predicted properties usefully and reliably reflects the distribution of true properties and indicates that the random forest-based model provides useful information that allows a wide range of compound classes to be reliably characterized based on mass spectrum and retention index.

2.6.1.2 Extrapolation Set Performance Evaluation

As discussed above, while the external standard test set provides useful information on model performance across a wide range of compound types, its performance is potentially inflated by high degree of chemical similarity between the training and test set compounds. Performance evaluation on the ambient sample extrapolation set is therefore likely a more accurate indicator of prediction performance on novel or uncatalogued species. Of the four properties modelled, the performances for carbon number prediction and carbon oxidation state remain consistent or slightly improve (carbon number OSR^2 increases to .93), while O:C and log(VP) prediction performances decrease, both in terms of OSR^2 and MAE (Table 2.3).

The weakest extrapolation set performance by far is vapor pressure prediction, which drops to an OSR^2 of .68. The correlation between predicted and true properties is also the weakest (as illustrated in Figure 2.4), with particularly large prediction residuals for the highest volatility species. For example, the extrapolation set compound with the highest vapor pressure prediction error is 1,2-Benzenedicarboxylic acid, which has a retention index of <1400 making it more volatile than the most volatile internal standard compound. While this compound does not lie outside of the volatility and polarity boundaries of the external standards in GCxGC space, is significantly more volatile than any diacid compound in the standard mixture, and the influence of double derivitization on its true ambient volatility relative to the chromatographic elution time of its derivatized form may not have been appropriately captured. Unlike the other properties targeted in this analysis, vapor pressure is not directly calculable based on chemical formula and poses challenges for many techniques; as discussed in Isaacman-Vanwertz and Aumont, 2021, molecular structure plays an important role in volatility, which significantly limits the accuracy with which techniques that identify formula but not structure (typically chemical ionization techniques) can predict the true volatility of their measured components. A more complete

comparison between the random forest model's performance in vapor pressure prediction compared to other techniques used throughout the field is therefore required to provide context for vapor pressure prediction errors in the ambient sample extrapolation set (further discussed below in section 2.6.1.3).

For both O:C and $\overline{OS_c}$ (which are highly related properties), extrapolation set prediction performance suffers at the high end of the oxygenation scale, although the performance reduction is far more pronounced for O:C prediction. This is due to the lack of highly oxygenated species in the external standard; random forest models do not extrapolate beyond the range of properties in the training data and therefore cannot predict O:C ratios of higher than 1.5 when that is above the maximum in the training data. The extraneously highly oxidized species for which O:C and $\overline{OS_c}$ prediction accuracy suffers lie almost exclusively in the most volatile region instrument sensitivity, where vapor pressure prediction inaccuracies have been previously described. As a result, extrapolation set property prediction for O:C, $\overline{OS_c}$, and log(VP) were restricted to compounds above a retention index of 1500. As illustrated in Figure 2.A3 and Figure 2.2, the significant majority of ambient analytes were above the 1500 retention index threshold, justifying the decision to restrict prediction of these properties to the retention index window in which their performance is better optimized. In applying these techniques to the larger suite of novel species, maintaining these retention window restrictions is critical to avoid the introduction of significant sources of error.

Given the strong and consistent performance of carbon number and $\overline{OS_c}$ predictions across the majority of the retention index space and between both test and extrapolation sets, the most robust visualization of chemical properties based on random forest predictions is likely to be in $\overline{OS_c}$-n_c space.[31] Predicting the carbon numbers and $\overline{OS_c}$ of the known ambient compounds and superimposing the true and predicted property distributions in the $\overline{OS_c}$-n_c space highlights the strengths and weaknesses of chemical properties modelling. To better represent the prediction capabilities of the full chemical space and the scope of information that would be provided for properties prediction on a complex sample including hundreds of individual species, all identifiable ambient compounds (including those that overlap with the external standard) were included in property prediction and visualization. As illustrated in Figure 2.5, the real and predicted chemical properties spaces for the ambient data set indicate both strengths and weaknesses for this application of chemically properties modelling. As noted earlier, random forest modelling does not extrapolate and has a tendency to underpredict property extremes. This is apparent in both the high $\overline{OS_c}$ region and the high carbon number regions of the $\overline{OS_c}$-n_c space, where high carbon oxidation states and high carbon numbers were each independently underpredicted. These errors could be moderated by adding more oxygenated species and higher carbon number species to the external standard, which would provide the model with more information to predict properties in these regions. In a context of extended continuity of analysis of similar sample media, this suggests an iterative approach in which the addition of new standards to a calibration mixture can be prioritized through analyzing the chemical features of poorly predicted compounds in the

sample media and adding new standards that replicate those features. Despite the prediction errors visualized in Figure 2.5, the overall shape of the true chemical properties space was extremely well represented by predictions. While conclusions based on the presence or absence of extremes in predicted properties would not be appropriate, analyses based on the relative distributions of populations of interest provides valuable insight comparable to other parameterizations of compound properties from incomplete knowledge.

2.6.1.3 Vapor Pressure Modelling: Comparison to Prior Methods

Chromatography using a non-polar column is intended to separate compounds by volatility and has been used to directly predict novel compound vapor pressures in previous studies.[35] It is therefore important in this context to evaluate both how significantly random forest modelling improves upon simple linear modelling of volatility based on retention index as well as how this method compares to other parameterizations of vapor pressure. As illustrated in Figure 2.6 and Table 2.4, the log(VP) prediction residuals for random forest model predictions indicate that random forest-generated predictions are both more accurate and more precise than predictions by the linearized retention index method or from the Li et al., 2016[36] chemical formula-based parameterization, as they demonstrate a tighter distribution that is more centered around zero. The mean absolute error for random forest vapor pressure prediction is significantly lower than errors from both predictions based on retention index (t-test p value = .01) and predictions based on chemical formula (t-test p value = 3.1×10^{-5}).

2.6.2 Quantification Modelling Performance

The approach for evaluating performance for quantification modeling requires slight alterations compared to property prediction. Although the random forest model predicts the residuals of quantification factors around the retention index response normalization curve (Figure 2.A2) rather than directly, these residuals are converted back to quantification factors for both the true and predicted properties for performance evaluation. This serves two purposes; first, other quantification methods do not use this retention index-based normalization so conversion to absolute prediction errors is necessary to compare methods, and second, a direct quantification error assessment provides more useful and applicable information about how significantly quantification errors could influence conclusions based on model-quantified data.

The test set compounds were quantified using two alternative quantification methods, *Manual* or *Closest proxy quantification* (described in Liang et al., 2021, which utilizes a combination of both), to benchmark random forest model performance. *Manual proxy quantification* entails manually assigning a compound to a chemically similar external standard based on researcher judgement on what chemical class the unidentified compound would likely belong to based on some combination of location in GCxGC space and mass spectrum. This is the current preferred method for quantification of compounds that are not in the external standard and in theory should provide the most reliable results in cases where an extremely chemically similar standard is available, but it is highly inefficient and relies upon researcher judgement calls which are difficult to standardize. *Closest proxy*

quantification assigns each compound to its nearest external standard in GCxGC space, or to an average of the nearest standards within a set radius limitation. In this work, the average of the quantification factors of the 3 nearest standard species was used, as this demonstrated improved performance compared to single closest proxy quantification. This method is efficient, but it introduces potentially significant error by assigning species with different chemical characteristics (and therefore different quantification factors) the same response factor if they are sufficiently close in GCxGC space. Each test set compound was assigned to a proxy quantification factor from the training set based on each of these two methods, and each proxy compound's quantification factor at each time point was substituted as a prediction of the test set compound's quantification factor at that calibration window.

The standard performance metrics for quantification factor prediction using the random forest model, manual proxy quantification and closest proxy quantification, are compared in Table 2.5. The random forest model significantly outperforms both other methods; it has a relatively high OSR^2 of .65 compared to negative OSR^2 values for the two proxy methods (indicating that at least on average, assuming all test compounds have the same quantification factor as the average of all training set compounds would have performed better than proxy quantification). MAE and RMSE also indicate improved performance when using the random forest model over other methods. While these metrics provide useful information on model performance, they do not reveal why the performance (particularly of the proxy methods) is so poor and do not provide useful information to evaluate likely propagation of quantification errors. Unlike for the chemical properties modelling, for quantification modelling % error is a much more important metric than absolute error, because it translates directly to how significant total quantification error across a large suite of compounds is likely to be and provides insights into underlying biases in different methods. Figure 2.7 illustrates the quantification factor % error distributions of the three methods and demonstrates the improved performance of random forest modelled quantification predictions on three criteria. First, as illustrated by panel A, random forest modeling produces far fewer and less extreme outlier prediction errors that are orders of magnitude different from the true values. These result when a compound that the instrument is extremely insensitive to (which would have a true extremely low quantification factor) is assigned a moderate or high quantification factor. In practice the influence of these types of quantification inaccuracies is very limited as few ambient species that the instrument is this significantly insensitive to would occur above detection limits, but they could introduce errors nonetheless. Here it is important to keep in mind that each point represents a single quantification from a single calibration period; some outliers therefore indicate compounds that exhibited extremes in quantification factors during a single calibration period. This was most common among standard compounds at the edges of the instrument's sensitivity window, as these species are more significantly impacted by alterations in instrument performance. Second, as illustrated by Figure 2.7 panel B, the error distribution for the random forest model is significantly more centered around zero compared to either proxy model. Median random forest model quantification error is -2%, compared to 17% for closest proxy quantification and 19% for manual proxy quantification.

In practice, this indicates that over a large number of quantified species, random forest modeling is unlikely to introduce biased quantifications that might skew results, while the two proxy methods would likely inflate the apparent mass of novel compounds. Third, also illustrated by Figure 2.7 panel B (though less directly), random forest modeling produces prediction errors more tightly distributed around the median, meaning that the absolute % error distribution for random forest modelling also outperforms the two proxy methods. Median absolute % error for random forest model predictions is 37%, compared to 57% for the closest proxy method and 41% for the manually assigned proxy method. The average % error improvements from random forest modeling compared to both proxy methods are statistically significant (t-test p values both < .0004), but the median absolute % error distributions of the random forest and manually assigned proxy quantifications are not significantly different based on a Mood's median test. The random forest and closest proxy method absolute % error distribution differences are statistically significant, with a Mood's test p value of .001. While critical for contextualizing the potential impact of quantification errors on mass attribution of complex mixtures, a % error-based analysis of prediction accuracy is necessarily asymmetrical, as a predicted quantification factor can produce a minimum of -100 % error (the case if the predicted value were to be zero) but far more than +100% error if the quantification factor is significantly overpredicted. A symmetrical error analysis of log(predicted quantification factor/true quantification factor), illustrated in Figure 2.A4, is required to probe the frequency and dynamics of underprediction in greater depth. Figure 2.A4 demonstrates that the random forest model is more prone to underprediction outliers, but continues to outperform the other methods in achieving a narrow error distribution centered at zero.

A final benefit of random forest modeling-based quantification not captured in the performance metrics is the ability to utilize incomplete data. With proxy quantification, any standard compound that cannot be calibrated for at any point over the course of an analysis cannot be used, as the species that are calibrated by that compound would not be quantifiable during the window with missing calibration data. The random forest-based quantification method relies upon the entire external standard suite to inform corrections for instrument performance over time and can therefore produce robust quantification factor predictions even when individual standard calibration curves are missing for a particular calibration window. This allows for significantly greater flexibility in analysis, as compounds can be added to the external standard if they are observed in initial samples and still be usable to inform quantifications for periods before they were present.

In summary, random forest quantification factor modelling significantly outperforms both closest proxy and manual proxy quantification methods. It is significantly more efficient than manual proxy modeling, exhibits fewer outliers of multi order of magnitude overestimations, produces an error distribution that is more centered around zero (preventing significant biases in total mass over large numbers of quantified and summed species), and exhibits improvements in absolute percent error of predictions.

2.6.3 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 absolute error was calculated for each set of predictions based on the shifted retention indices and compared to the unshifted mean absolute error to calculate the % increase in mean absolute error as a function of test set retention index shift. These results are visualized in Figure 2.8. The two measures of oxidation, $\overline{OS_c}$ and O:C were relatively insensitive to retention index shifts, as their mean absolute 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 absolute 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.

2.7 Conclusions

This work presents a new machine-learning based method for quantifying and predicting chemical properties of novel organic compounds observed in the atmosphere. Based on a relatively small combined training and test set of ~130 known compounds, we are able to predict the carbon numbers, vapor pressures, carbon oxidation states, and O to C ratios of ambient organic compounds with sufficient accuracy to usefully represent compound distributions in chemical property spaces that are important in atmospheric science. That these predictions are generated solely from retention indices and unit mass resolution mass spectra marks a significant step forward in ability to characterize the novel organic components of earth's atmosphere based on measurements generated from a wide range of commonly available atmospheric instrumentation. In GCxGC-MS applications, these methods contribute significant improvements in both accuracy and analytical efficiency for novel compound quantification that enable users to perform untargeted analysis of the rich complexity of data generated by advances in instrumentation. While the untargeted analysis data science techniques described in this work have been developed for and tested on atmospheric applications, they are not structurally limited in scope and could be applied to a wide range of chromatographic-mass spectral data sets to enable

characterization of complex organic mixtures. The open-source R script published in supplement to this work is intended to provide a framework for groups throughout the atmospheric chemistry community to efficiently apply and adapt these methods to broadly enhance our ability to take advantage of the increasingly complex information provided by ever accelerating advances in environmental chemistry instrumentation.

2.8 Acknowledgements

We gratefully acknowledge the support of the National Science Foundation Graduate Research Fellowship Program (DGE- 1752814), as well as the Department of Energy Atmospheric System Research program (DE-SC0020051). The described work was conducted under DOE-ASR support and benefits from samples collected at the GoAmazon field campaign, a DOE-ASR supported observational campaign. Paul Grigas's research is supported in part by NSF AI Institute for Advances in Optimization Award 211253 and NSF Award CMMI-1762744

2.9 References

(1) Jimenez, J. L.; Canagaratna, M. R.; Donahue, N. M.; Prevot, A. S. H.; Zhang, Q.; Kroll, J. H.; DeCarlo, P. F.; Allan, J. D.; Coe, H.; Ng, N. L.; Aiken, A. C.; Docherty, K. S.; Ulbrich, I. M.; Grieshop, A. P.; Robinson, A. L.; Duplissy, J.; Smith, J. D.; Wilson, K. R.; Lanz, V. A.; Hueglin, C.; Sun, Y. L.; Tian, J.; Laaksonen, A.; Raatikainen, T.; Rautiainen, J.; Vaattovaara, P.; Ehn, M.; Kulmala, M.; Tomlinson, J. M.; Collins, D. R.; Cubison, M. J.; E.; Dunlea, J.; Huffman, J. A.; Onasch, T. B.; Alfarra, M. R.; Williams, P. I.; Bower, K.; Kondo, Y.; Schneider, J.; Drewnick, F.; Borrmann, S.; Weimer, S.; Demerjian, K.; Salcedo, D.; Cottrell, L.; Griffin, R.; Takami, A.; Miyoshi, T.; Hatakeyama, S.; Shimono, A.; Sun, J. Y.; Zhang, Y. M.; Dzepina, K.; Kimmel, J. R.; Sueper, D.; Jayne, J. T.; Herndon, S. C.; Trimborn, A. M.; Williams, L. R.; Wood, E. C.; Middlebrook, A. M.; Kolb, C. E.; Baltensperger, U.; Worsnop, D. R. Evolution of Organic Aerosols in the Atmosphere. *Science (80-.).* **2009**, *326* (5959), 1525–1529. https://doi.org/10.1126/SCIENCE.1180353.

(2) Goldstein, A. H.; Galbally, I. E. Known and Unexplored Organic Constituents in the Earth's Atmosphere. *Environ. Sci. Technol.* **2007**, *41* (5), 1514–1521. https://doi.org/10.1021/ES072476P.

(3) Ditto, J. C.; Barnes, E. B.; Khare, P.; Takeuchi, M.; Joo, T.; Bui, A. A. T.; Lee-Taylor, J.; Eris, G.; Chen, Y.; Aumont, B.; Jimenez, J. L.; Ng, N. L.; Griffin, R. J.; Gentner, D. R. An Omnipresent Diversity and Variability in the Chemical Composition of Atmospheric Functionalized Organic Aerosol. *Commun. Chem.* **2018**. https://doi.org/10.1038/s42004-018-0074-3.

(4) Heald, C. L.; Kroll, J. H.; Jimenez, J. L.; Docherty, K. S.; DeCarlo, P. F.; Aiken, A. C.; Chen, Q.; Martin, S. T.; Farmer, D. K.; Artaxo, P. A Simplified Description of

the Evolution of Organic Aerosol Composition in the Atmosphere. *Geophys. Res. Lett.* **2010**, *37* (8). https://doi.org/10.1029/2010GL042737.

(5) Hunter, J. F.; Day, D. A.; Palm, B. B.; Yatavelli, R. L. N.; Chan, A. W. H.; Kaser, L.; Cappellin, L.; Hayes, P. L.; Cross, E. S.; Carrasquillo, A. J.; Campuzano-Jost, P.; Stark, H.; Zhao, Y.; Hohaus, T.; Smith, J. N.; Hansel, A.; Karl, T.; Goldstein, A. H.; Guenther, A.; Worsnop, D. R.; Thornton, J. A.; Heald, C. L.; Jimenez, J. L.; Kroll, J. H. Comprehensive Characterization of Atmospheric Organic Carbon at a Forested Site. **2017**. https://doi.org/10.1038/NGEO3018.

(6) Isaacman-Vanwertz, G.; Massoli, P.; O'Brien, R.; Lim, C.; Franklin, J. P.; Moss, J. A.; Hunter, J. F.; Nowak, J. B.; Canagaratna, M. R.; Misztal, P. K.; Arata, C.; Roscioli, J. R.; Herndon, S. T.; Onasch, T. B.; Lambe, A. T.; Jayne, J. T.; Su, L.; Knopf, D. A.; Goldstein, A. H.; Worsnop, D. R.; Kroll, J. H. Chemical Evolution of Atmospheric Organic Carbon over Multiple Generations of Oxidation. *Nat. Chem. 2018 104* **2018**, *10* (4), 462–468. https://doi.org/10.1038/s41557-018-0002-2.

(7) Nozière, B.; Kalberer, M.; Claeys, M.; Allan, J.; D'Anna, B.; Decesari, S.; Finessi, E.; Glasius, M.; Grgić, I.; Hamilton, J. F.; Hoffmann, T.; Iinuma, Y.; Jaoui, M.; Kahnt, A.; Kampf, C. J.; Kourtchev, I.; Maenhaut, W.; Marsden, N.; Saarikoski, S.; Schnelle-Kreis, J.; Surratt, J. D.; Szidat, S.; Szmigielski, R.; Wisthaler, A. The Molecular Identification of Organic Compounds in the Atmosphere: State of the Art and Challenges. *Chem. Rev.* **2015**, *115* (10), 3919–3983. https://doi.org/10.1021/CR5003485.

(8) Isaacman-Vanwertz, G.; Aumont, B. Impact of Organic Molecular Structure on the Estimation of Atmospherically Relevant Physicochemical Parameters. *Atmos. Chem. Phys.* **2021**, *21* (8), 6541–6563. https://doi.org/10.5194/ACP-21-6541-2021.

(9) Worton, D. R.; Kreisberg, N. M.; Isaacman, G.; Teng, A. P.; McNeish, C.; Górecki, T.; Hering, S. V.; Goldstein, A. H. Thermal Desorption Comprehensive Two-Dimensional Gas Chromatography: An Improved Instrument for In-Situ Speciated Measurements of Organic Aerosols. *http://dx.doi.org/10.1080/02786826.2011.634452* **2011**, *46* (4), 380–393. https://doi.org/10.1080/02786826.2011.634452.

(10) Goldstein, A. H.; Worton, D. R.; Williams, B. J.; Hering, S. V; Kreisberg, N. M.; Pani´cpani´c, O.; Górecki, T. Thermal Desorption Comprehensive Two-Dimensional Gas Chromatography for in-Situ Measurements of Organic Aerosols. *J. Chromatogr. A* **2008**, *1186*, 340–347. https://doi.org/10.1016/j.chroma.2007.09.094.

(11) Worton, D. R.; Decker, M.; Isaacman-VanWertz, G.; Chan, A. W. H.; Wilson, K. R.; Goldstein, A. H. Improved Molecular Level Identification of Organic Compounds Using Comprehensive Two-Dimensional Chromatography, Dual Ionization Energies and High Resolution Mass Spectrometry. *Analyst* **2017**, *142* (13), 2395–2403. https://doi.org/10.1039/C7AN00625J.

(12) Vinaixa, M.; Schymanski, E. L.; Neumann, S.; Navarro, M.; Salek, R. M.; Yanes, O. Mass Spectral Databases for LC/MS- and GC/MS-Based Metabolomics: State of the Field and Future Prospects. *TrAC Trends Anal. Chem.* **2016**, *78*, 23–35. https://doi.org/10.1016/J.TRAC.2015.09.005.

(13) Hamilton, J. F.; Webb, P. J.; Lewis, A. C.; Hopkins, J. R.; Smith, S.; Davy, P. Atmospheric Chemistry and Physics Partially Oxidised Organic Components in Urban Aerosol Using GCXGC-TOF/MS. *Atmos. Chem. Phys* **2004**, *4*, 1279–1290.

(14) Bi, C.; Krechmer, J. E.; Frazier, G. O.; Xu, W.; Lambe, A. T.; Claflin, M. S.; Lerner, B. M.; Jayne, J. T.; Worsnop, D. R.; Canagaratna, M. R.; Isaacman-Vanwertz, G. Coupling a Gas Chromatograph Simultaneously to a Flame Ionization Detector and Chemical Ionization Mass Spectrometer for Isomer-Resolved Measurements of Particle-Phase Organic Compounds. *Atmos. Meas. Tech.* **2021**, *14* (5), 3895–3907. https://doi.org/10.5194/AMT-14-3895-2021.

(15) Zhang, H.; Yee, L. D.; Lee, B. H.; Curtis, M. P.; Worton, D. R.; Isaacman-VanWertz, G.; Offenberg, J. H.; Lewandowski, M.; Kleindienst, T. E.; Beaver, M. R.; Holder, A. L.; Lonneman, W. A.; Docherty, K. S.; Jaoui, M.; Pye, H. O. T.; Hu, W.; Day, D. A.; Campuzano-Jost, P.; Jimenez, J. L.; Guo, H.; Weber, R. J.; Gouw, J. de; Koss, A. R.; Edgerton, E. S.; Brune, W.; Mohr, C.; Lopez-Hilfiker, F. D.; Lutz, A.; Kreisberg, N. M.; Spielman, S. R.; Hering, S. V.; Wilson, K. R.; Thornton, J. A.; Goldstein, A. H. Monoterpenes Are the Largest Source of Summertime Organic Aerosol in the Southeastern United States. *Proc. Natl. Acad. Sci.* **2018**, *115* (9), 2038–2043. https://doi.org/10.1073/PNAS.1717513115.

(16) Liang, Y.; Jen, C. N.; Weber, R. J.; Misztal, P. K.; Goldstein, A. H. Chemical Composition of PM2.5 in October 2017 Northern California Wildfire Plumes. *Atmos. Chem. Phys.* **2021**, *21* (7), 5719–5737. https://doi.org/10.5194/ACP-21-5719-2021.

(17) Hatch, L. E.; Luo, W.; Pankow, J. F.; Yokelson, R. J.; Stockwell, C. E.; Barsanti, K. C. Identification and Quantification of Gaseous Organic Compounds Emitted from Biomass Burning Using Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry. *Atmos. Chem. Phys* **2015**, *15*, 1865–1899. https://doi.org/10.5194/acp-15-1865-2015.

(18) Jen, C. N.; Hatch, L. E.; Selimovic, V.; Yokelson, R. J.; Weber, R.; Fernandez, A. E.; Kreisberg, N. M.; Barsanti, K. C.; Goldstein, A. H. Speciated and Total Emission Factors of Particulate Organics from Burning Western US Wildland Fuels and Their Dependence on Combustion Efficiency. *Atmos. Chem. Phys.* **2019**, *19* (2), 1013–1026. https://doi.org/10.5194/ACP-19-1013-2019.

(19) Rokach, L. Decision Forest: Twenty Years of Research. *Inf. Fusion* **2016**, *27*, 111–125. https://doi.org/10.1016/J.INFFUS.2015.06.005.

(20) Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A Comparative Analysis of Gradient Boosting Algorithms. *Artif. Intell. Rev.* **2021**, *54* (3), 1937–1967. https://doi.org/10.1007/S10462-020-09896-5/TABLES/12.

(21) Breiman, L. Random Forests. *Mach. Learn. 2001 451* **2001**, *45* (1), 5–32. https://doi.org/10.1023/A:1010933404324.

(22) Whitmore, L. S.; Davis, R. W.; McCormick, R. L.; Gladden, J. M.; Simmons, B. A.; George, A.; Hudson, C. M. BioCompoundML: A General Biofuel Property Screening Tool for Biological Molecules Using Random Forest Classifiers. *Energy and Fuels* **2016**, *30* (10), 8410–8418. https://doi.org/10.1021/ACS.ENERGYFUELS.6B01952/SUPPL_FILE/EF6B01952_SI_001.XLSX.

(23) Yee, L. D.; Isaacman-VanWertz, G.; Wernis, R. A.; Meng, M.; Rivera, V.; Kreisberg, N. M.; Hering, S. V.; Bering, M. S.; Glasius, M.; Upshur, M. A.; Gray Bé, A.; Thomson, R. J.; Geiger, F. M.; Offenberg, J. H.; Lewandowski, M.; Kourtchev, I.; Kalberer, M.; De Sá, S.; Martin, S. T.; Alexander, M. L.; Palm, B. B.; Hu, W.; Campuzano-Jost, P.; Day, D. A.; Jimenez, J. L.; Liu, Y.; McKinney, K. A.; Artaxo, P.; Viegas, J.; Manzi, A.; Oliveira, M. B.; De Souza, R.; Machado, L. A. T.; Longo, K.; Goldstein, A. H. Observations of Sesquiterpenes and Their Oxidation Products in Central Amazonia during the Wet and Dry Seasons. *Atmos. Chem. Phys.* **2018**, *18* (14), 10433–10457. https://doi.org/10.5194/ACP-18-10433-2018.

(24) Bé, A. G.; Chase, H. M.; Liu, Y.; Upshur, M. A.; Zhang, Y.; Tuladhar, A.; Chase, Z. A.; Bellcross, A. D.; Wang, H. F.; Wang, Z.; Batista, V. S.; Martin, S. T.; Thomson, R. J.; Geiger, F. M. Atmospheric β-Caryophyllene-Derived Ozonolysis Products at Interfaces. *ACS Earth Sp. Chem.* **2019**, *3* (2), 158–169. https://doi.org/10.1021/ACSEARTHSPACECHEM.8B00156/SUPPL_FILE/SP8B00156_SI_001.PDF.

(25) Martin, S. T.; Artaxo, P.; MacHado, L. A. T.; Manzi, A. O.; Souza, R. A. F.;

Schumacher, C.; Wang, J.; Andreae, M. O.; Barbosa, H. M. J.; Fan, J.; Fisch, G.; Goldstein, A. H.; Guenther, A.; Jimenez, J. L.; Pöschl, U.; Silva Dias, M. A.; Smith, J. N.; Wendisch, M. Introduction: Observations and Modeling of the Green Ocean Amazon (GoAmazon2014/5). *Atmos. Chem. Phys.* **2016**, *16* (8), 4785–4797. https://doi.org/10.5194/ACP-16-4785-2016.

(26) Martin, S. T.; Artaxo, P.; Machado, L.; Manzi, A. O.; Souza, R. A. F.; Schumacher, C.; Wang, J.; Biscaro, T.; Brito, J.; Calheiros, A.; Jardine, K.; Medeiros, A.; Portela, B.; De Sá, S. S.; Adachi, K.; Aiken, A. C.; Alblbrecht, R.; Alexander, L.; Andreae, M. O.; Barbosa, H. M. J.; Buseck, P.; Chand, D.; Comstmstmstock, J. M.; Day, D. A.; Dubey, M.; Fan, J.; Faststt, J.; Fisch, G.; Fortner, E.; Giangrande, S.; Gilllles, M.; Goldststein, A. H.; Guenther, A.; Hubbbbe, J.; Jensen, M.; Jimenez, J. L.; Keutstsch, F. N.; Kim, S.; Kuang, C.; Laskskin, A.; McKinney, K.; Mei, F.; Millller, M.; Nascimento, R.; Pauliquevis, T.; Pekour, M.; Peres, J.; Petäjä, T.; Pöhlklker, C.; Pöschl, U.; Rizzo, L.; Schmid, B.; Shilllling, J. E.; Silva Dias, M. A.; Smith, J. N.; Tomlmlinson, J. M.; Tóta, J.; Wendisch, M. The Green Ocean Amazon Experiment (GoAmazon2014/5) Observes Pollution Affecting Gases, Aerosols, Clouds, and Rainfall over the Rain Forest. *Bull. Am. Meteorol. Soc.* **2017**, *98* (5), 981–997. https://doi.org/10.1175/BAMS-D-15-00221.1.

(27) Franklin, E. B.; Alves, M. R.; Moore, A. N.; Kilgour, D. B.; Novak, G. A.; Mayer, K.; Sauer, J. S.; Weber, R. J.; Dang, D.; Winter, M.; Lee, C.; Cappa, C. D.; Bertram, T. H.; Prather, K. A.; Grassian, V. H.; Goldstein, A. H. Atmospheric Benzothiazoles in a Coastal Marine Environment. *Environ. Sci. Technol.* **2021**, acs.est.1c04422. https://doi.org/10.1021/ACS.EST.1C04422.

(28) Isaacman, G.; Kreisberg, N. M.; Yee, L. D.; Worton, D. R.; Chan, A. W. H.; Moss, J. A.; Hering, S. V; Goldstein, A. H. Online Derivatization for Hourly Measurements of Gas-and Particle-Phase Semi-Volatile Oxygenated Organic Compounds by Thermal Desorption Aerosol Gas Chromatography (SV-TAG). *Atmos. Meas. Tech* **2014**, *7*, 4417–4429. https://doi.org/10.5194/amt-7-4417-2014.

(29) Eghbaldar, A.; Forrest, T. P.; Cabrol-Bass, D. Development of Neural Networks for Identification of Structural Features from Mass Spectral Data. *Anal. Chim. Acta* **1998**, *359* (3), 283–301. https://doi.org/10.1016/S0003-2670(97)00663-6.

(30) Donahue, N. M.; Robinson, A.; Stanier, C. O.; Pandis, S. N. Coupled Partitioning, Dilution, and Chemical Aging of Semivolatile Organics. *Environ. Sci. Technol.* **2006**, *40* (8), 2635–2643. https://doi.org/10.1021/ES052297C.

(31) Kroll, J. H.; Donahue, N. M.; Jimenez, J. L.; Kessler, S. H.; Canagaratna, M. R.; Wilson, K. R.; Altieri, K. E.; Mazzoleni, L. R.; Wozniak, A. S.; Bluhm, H.; Mysak,

E. R.; Smith, J. D.; Kolb, C. E.; Worsnop, D. R. Carbon Oxidation State as a Metric for Describing the Chemistry of Atmospheric Organic Aerosol. *Nat. Chem. 2010 32* **2011**, *3* (2), 133–139. https://doi.org/10.1038/nchem.948.

(32) Compernolle, S.; Ceulemans, K.; Müller, J. F. Evaporation: A New Vapour Pressure Estimation Methodfor Organic Molecules Including Non-Additivity and Intramolecular Interactions. *Atmos. Chem. Phys.* **2011**, *11* (18), 9431–9450. https://doi.org/10.5194/ACP-11-9431-2011.

(33) Nannoolal, Y.; Rarey, J.; Ramjugernath, D. Estimation of Pure Component Properties: Part 3. Estimation of the Vapor Pressure of Non-Electrolyte Organic Compounds via Group Contributions and Group Interactions. *Fluid Phase Equilib.* **2008**, *269* (1–2), 117–133. https://doi.org/10.1016/J.FLUID.2008.04.020.

(34) Pankow, J. F.; Asher, W. E. SIMPOL.1: A Simple Group Contribution Method for Predicting Vapor Pressures and Enthalpies of Vaporization of Multifunctional Organic Compounds. *Atmos. Chem. Phys.* **2008**, *8* (10), 2773–2796. https://doi.org/10.5194/ACP-8-2773-2008.

(35) Isaacman-VanWertz, G.; Yee, L. D.; Kreisberg, N. M.; Wernis, R.; Moss, J. A.; Hering, S. V.; De Sá, S. S.; Martin, S. T.; Alexander, M. L.; Palm, B. B.; Hu, W.; Campuzano-Jost, P.; Day, D. A.; Jimenez, J. L.; Riva, M.; Surratt, J. D.; Viegas, J.; Manzi, A.; Edgerton, E.; Baumann, K.; Souza, R.; Artaxo, P.; Goldstein, A. H. Ambient Gas-Particle Partitioning of Tracers for Biogenic Oxidation. *Environ. Sci. Technol.* **2016**, *50* (18), 9952–9962. https://doi.org/10.1021/ACS.EST.6B01674/SUPPL_FILE/ES6B01674_SI_001.PDF.

(36) Li, Y.; Pöschl, U.; Shiraiwa, M. Molecular Corridors and Parameterizations of Volatility in the Chemical Evolution of Organic Aerosols. *Atmos. Chem. Phys* **2016**, *16*, 3327–3344. https://doi.org/10.5194/acp-16-3327-2016.

2.10 Figures and Tables

2.10.1 Figures

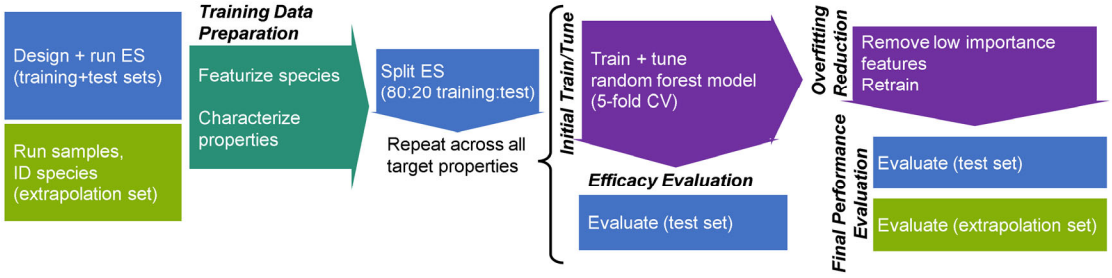


Figure 2.1: Analytical pipeline for chemical properties modelling using a random forest model. ES indicates external standard; CV indicates cross validation

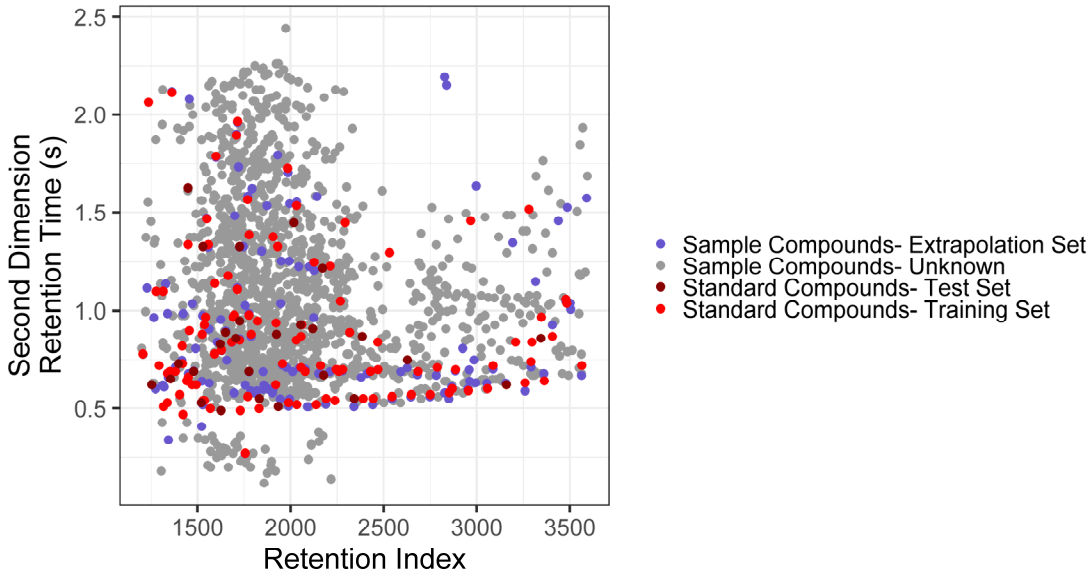


Figure 2.2: Distribution of training, test, extrapolation, and unidentified sample compounds in two-dimensional chromatographic chemical properties space

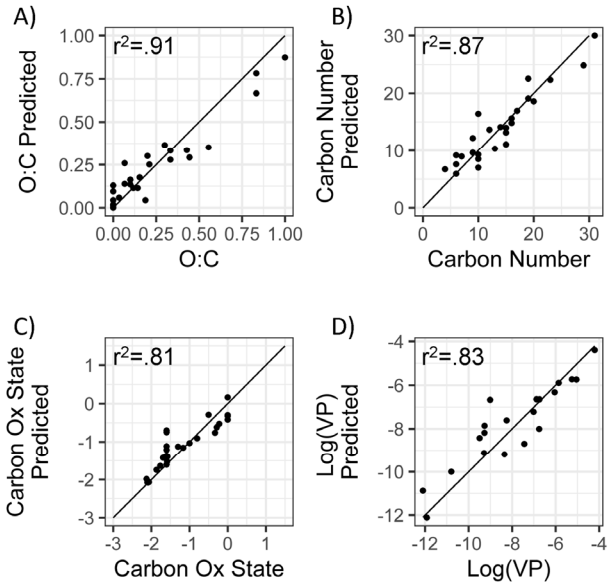


Figure 2.3: External standard test set true and predicted chemical properties from random forest modelling

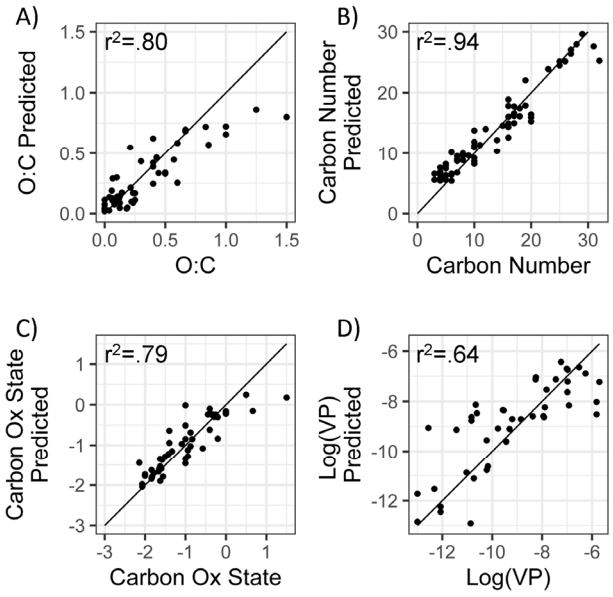


Figure 2.4: Ambient extrapolation set true and predicted chemical properties from random forest modelling

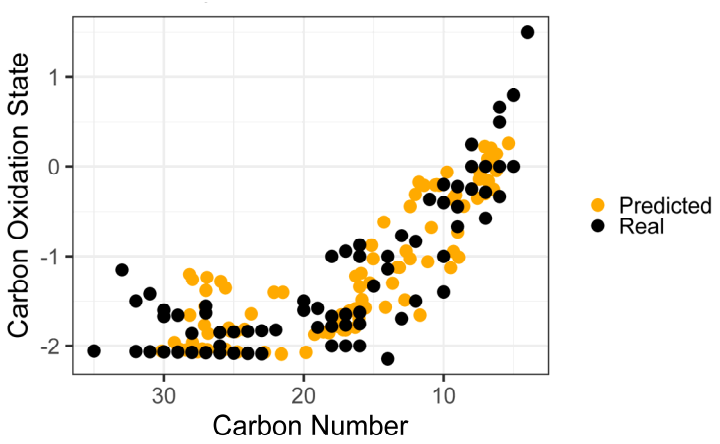


Figure 2.5: True versus predicted chemical properties distribution of ambient sample organic species within a Carbon Number v. Carbon Oxidation State space

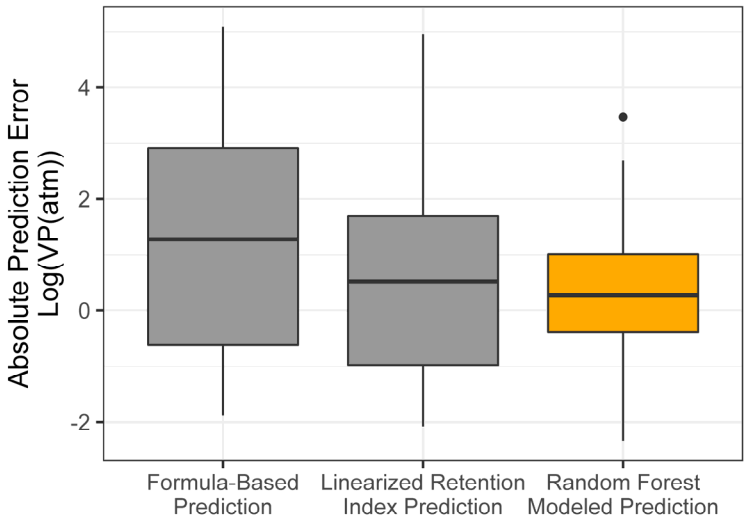


Figure 2.6: Vapor pressure prediction residuals (Log(VP), VP in atm) for vapor pressure predictions of the ambient extrapolation set based on formula-based parameterization (Li et al., 2016), linearized retention index-based modelling, and random forest modelling.

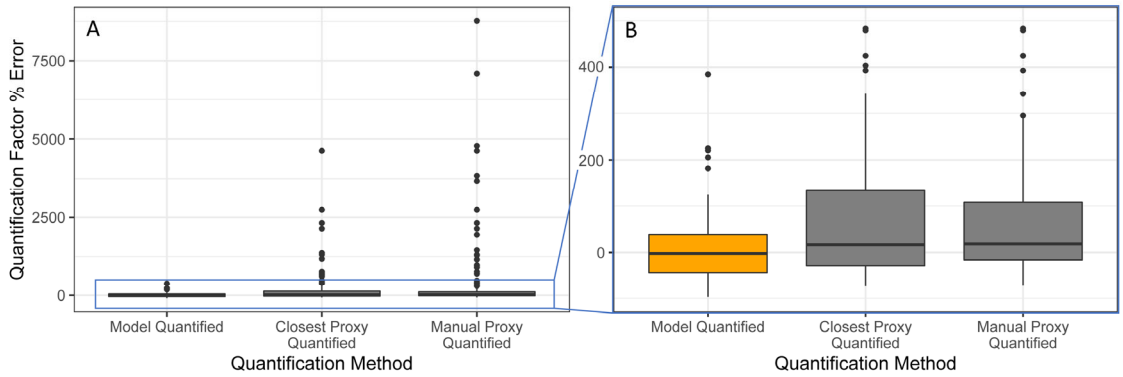


Figure 2.7: Quantification performance comparison between random forest model (orange) and two previously utilized quantification methods, specifically closest proxy quantification and manually assigned proxy quantification. Midline of boxes indicates sample median, while top and bottom indicate 25th and 75th percentiles. Linear "whiskers" extend to the least extreme values within 1.5 × the inner quartile range of the sample. Disconnected dots indicate sample outliers that fall beyond the whisker parameters.

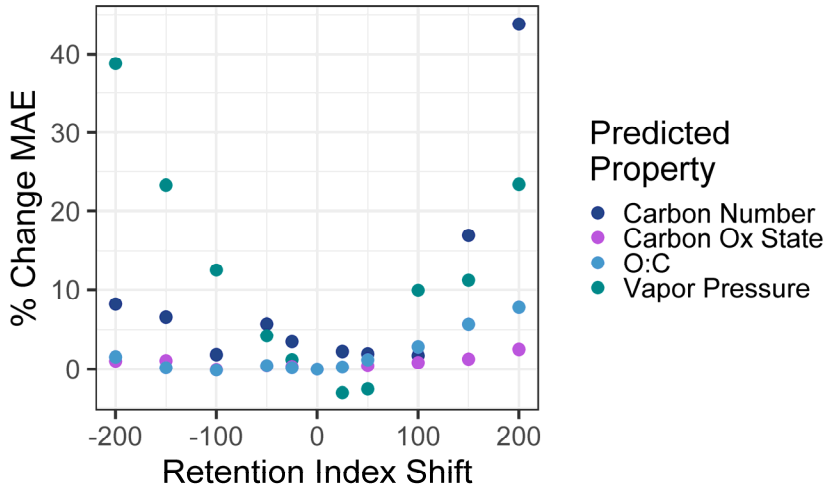


Figure 2.8: % increases of mean absolute error in chemical property prediction as a function of shift in test set retention index relative to training set retention index. Retention indices are normalized to a linear alkane series, making an increment of 100 indicate the retention time differences between two linear alkanes separated by 1 carbon number.

2.10.2 Tables

Table 2.1: Tuning parameters and important features for chemical properties prediction models. m/z indicates charged fragment features and n indicates neutral mass difference features.

Property Model	Optimized mtry	Number of Important Features	Important Features
O:C	4	19	Retention index, m/z 41, m/z 43, m/z 45, m/z 57, m/z 69, m/z 73, m/z 74, m/z 75, m/z 103, m/z 113, m/z 147, m/z 189, m/z 204, m/z 217, n 2, n 15, n 28, n 30
Carbon Number	6	9	Retention index, m/z 41, m/z 45, m/z 55, m/z 57, m/z 73, m/z 99, n 14
Average Carbon Oxidation State	4	17	Retention index, m/z 41, m/z 43, m/z 45, m/z 55, m/z 57, m/z 69, m/z 73, m/z 75, m/z 91, m/z 93, m/z 117, m/z 119, m/z 147, n 1, n 2, n 30
Log(Vapor Pressure)	9	9	Retention index, m/z 55, m/z 73, m/z 75, m/z 129, m/z 145, m/z 147, n 3, n 30

Table 2.2: Performance metrics for random forest-based modelling of chemical properties of the external standard test set. "Range of true properties" units in units of property: O:C in unitless atom#/atom#, Carbon Number in atom#, average carbon oxidation state in mean charge, and Log(Vapor Pressure) in Log(atm).

Property	Out of Sample R^2	Mean Absolute Error	Root Mean Square Error	Range of True Properties
O:C	.89	.072	.094	0-1
Carbon Number	.88	1.8	2.4	4-31
Average Carbon Oxidation State	.79	.24	.33	(-2.1)- 0
Log(Vapor Pressure)	.82	.72	.93	(-12)-(-4.2)

Table 2.3: Performance metrics for random forest-based modelling of chemical properties of the ambient aerosol sample extrapolation set. "Range of true properties" units in units of property: O:C in unitless atom#/atom#, Carbon Number in atom#, average carbon oxidation state in mean charge, and Log(Vapor Pressure) in Log(atm).

Property	Out of Sample R^2	Mean Absolute Error	Root Mean Square Error	Range of True Properties
O:C*	.78	.11	.17	0-1.5
Carbon Number	.93	1.8	2.2	3-32
Average Carbon Oxidation State*	.80	.25	.37	(-2.1)- (1.5)
Log(Vapor Pressure)*	.68	1.1	1.4	(-13)- (-5.7)

***Restricted to retention index > 1500**

Table 2.4. Error distribution metrics random forest model, retention index linear model, and formula-based predictions of vapor pressure. All reported errors in units of log(VP(atm)).

Vapor Pressure Prediction Method	Mean Error	Median Error	Mean Absolute Error	Median Absolute Error
Random Forest Model	.24	.21	1.1	.76
Retention Index Linear Model	.55	.52	1.5	1.1
Formula-Based Parameterization	1.2	1.3	2.0	1.3

Table 2.5. Performance metrics for quantification factor prediction for three methods of unidentified compound quantification: random forest modelling, manually assigned proxy quantified, and closest proxy quantified.

Quantification Method	Out of Sample R^2	Mean Absolute Error	Root Mean Square Error
Random Forest Model	.65	.00085	.0021
Manually Assigned Proxy Quantified	-4.1	.0036	.0080
Closest Proxy Quantified	-1.8	.0026	.0059

2.11 Supporting Information

2.11.1 Appendix A: Supplementary Tables and Figures

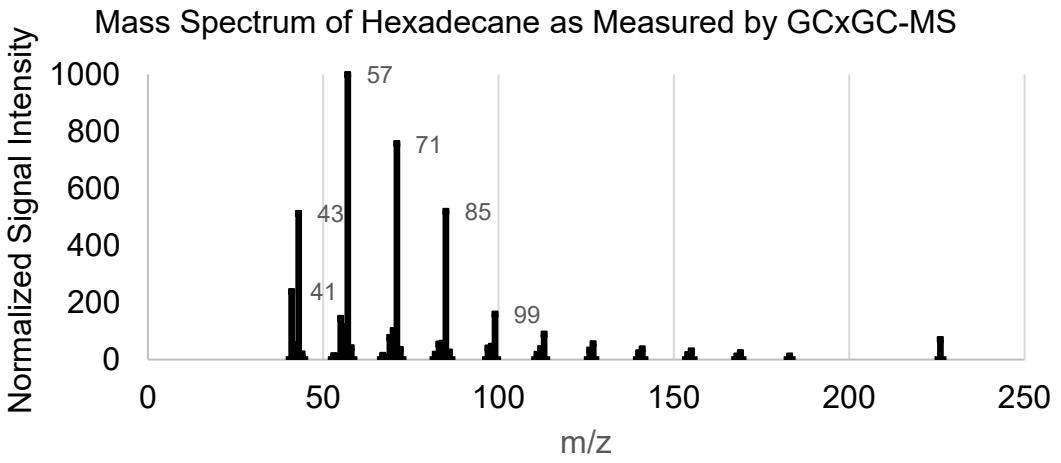


Figure 2.A1: Mass Spectrum of Hexadecane as measured by GCxGC-MS and featurized in table A2.

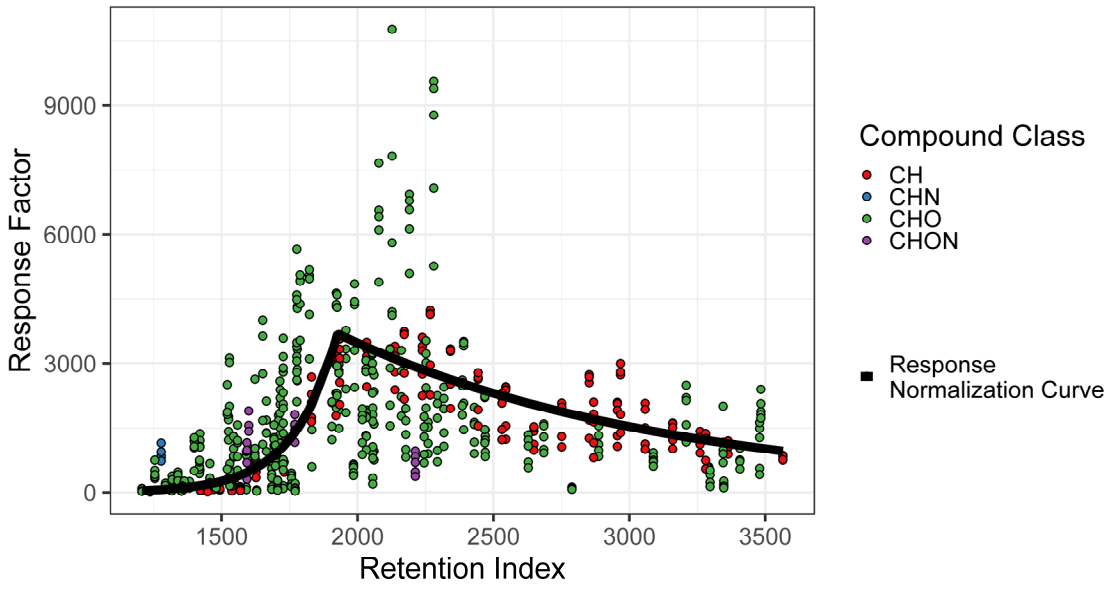


Figure 2.A2: Quantification factor normalization curve based on average response factors of alkanes

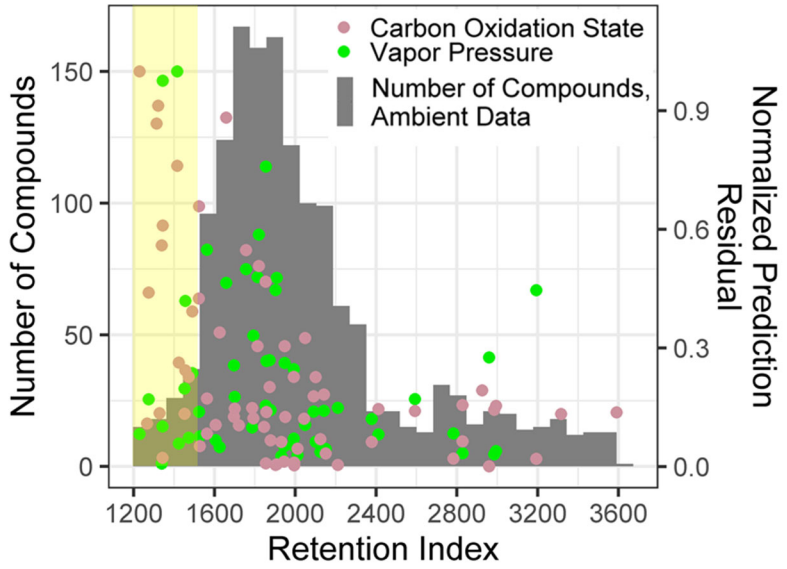


Figure 2.A3: Normalized prediction residuals of carbon oxidation state and vapor pressure v. retention index for ambient data compound property predictions set, overlaid with compound number distribution over the retention index for ambient data set. The yellow highlighted region indicates compounds below a retention index of 1500.

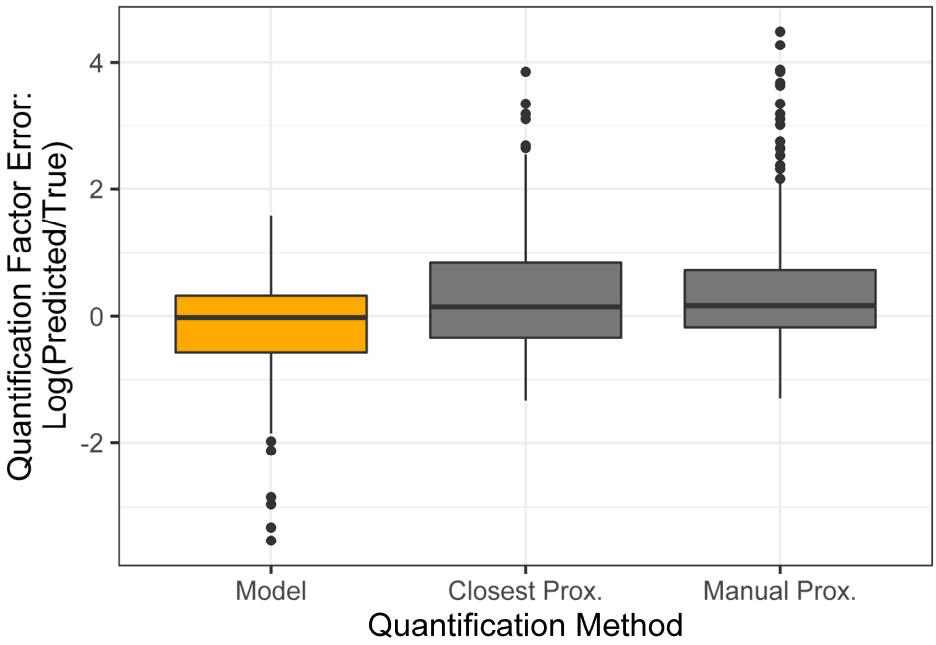


Figure 2.A4: Quantification factor prediction errors expressed in Log(predicted quantification factor/true quantification factor) for test set quantification factors predicted by random forest model (orange), closest proxy, and manual proxy methods.

Table 2.A1: External standard names, formulae (underivatized), retention indexes, split (training set versus test set), and manually assigned quantification proxies.

Name	Chemical Formula	Retention Index*	Split	Manual Proxy
12-OH C18 acid	C18H36O3	2470	Train	
16-OH C16 acid	C16H32O3	2429	Train	
2-ketoglutaric acid	C5H6O5	1629	Train	
3-5-dimethoxyphenol	C8H10O3	1525	Train	
4, 4 dimethoxy-benzophenone	C15H14O3	2293	Train	
4-hydroxybenzoic acid	C7H6O3	1651	Test	2-ketoglutaric acid
4-nitrocatechol	C6H5NO4	1769	Train	
4-terpineol	C10H18O	1206	Train	
9H-florenone	C13H8O	1778	Train	
a-amyrin	C30H50O	3479	Train	
abietic acid	C20H30O2	2468	Train	
anthraquinone	C14H8O2	2017	Test	xanthone
benzophenone	C13H10O	1664	Train	
beta-caryophyllene aldehyde	C15H24O2	1715	Train	
beta-caryophyllinic acid	C14H22O4	2060	Train	
beta-caryophyllonic acid	C15H24O3	1931	Train	
beta-nocaryophyllinic acid	C13H20O5	2127	Train	

beta-nocaryophyllone aldehyde	C14H22O3	1757	Train	
beta-nocaryophyllonic acid	C14H22O4	1985	Train	
beta-sitosterol	C29H50O	3406	Train	
bisabolol	C15H26O	1770	Train	
borneol	C10H18O	1254	Test	nonanol
C10 carboxylic acid	C10H20O2	1479	Test	dimethyl glutaric acid
C10 diacid (sebacic acid)	C10H18O4	1922	Train	
C12 diacid	C12H22O4	2120	Test	beta-caryophyllinic acid
C13 acid	C13H26O2	1776	Test	vanillic acid
C14 alkane	C14H30	1422	Train	
C14 diacid	C14H26O4	2317	Train	
C16 alkane	C16H34	1626	Train	
C16 acid	C16H32O2	2078	Train	
C17 alkane	C17H36	1730	Test	C17 alkane
C17 acid	C17H34O2	2177	Test	linoleic acid
C18 alkane	C18H38	1830	Train	
C18 acid	C18H36O2	2280	Train	
C19 alkane	C19H40	1934	Test	C20 alkane
C20 alkane	C20H42	2034	Train	
C21 alkane	C21H44	2137	Train	
C22 alkane	C22H46	2238	Train	
C22 acid	C22H44O2	2684	Train	
C23 alkane	C23H48	2341	Test	C24 alkane
C24 alkane	C24H50	2443	Train	
C24 acid	C24H48O2	2886	Train	
C25 alkane	C25H52	2545	Train	
C26 alkane	C26H54	2649	Train	
C26 acid	C26H52O2	3088	Train	
C27 alkane	C27H56	2750	Train	
C28 alkane	C28H58	2852	Train	
C28 acid	C28H56O2	3291	Train	
C29 alkane	C29H60	2955	Train	
C30 alkane	C30H62	3058	Train	
C31 alkane	C31H64	3159	Test	C30 alkane
C32 alkane	C32H66	3259	Train	
C33 alkane	C33H68	3363	Train	
C35 alkane	C35H72	3564	Train	
C7 acid	C7H14O2	< 1400	Train	
C8 acid	C8H16O2	1293	Train	
C9 acid	C9H18O2	1381	Train	
C9 diacid (azelaic acid)	C9H16O4	1822	Train	

cholesterol	C27H46O	3209	Train	
chrysene	C18H12	2531	Train	
cis-vaccenic acid	C18H34O2	2259	Train	
citronellol	C10H20O	1338	Train	
cycloisolongifolene	C15H24	1355	Test	pyrocatechol
DEET	C12H17NO	1600	Train	
deoxycholic acid	C24H40O4	3347	Train	
dibenz(ah)anthracene	C22H14	3280	Train	
dimethyl glutaric acid	C7H12O4	1456	Train	
dodecyl benzene	C18H30	1920	Train	
eicosanol	C20H42O	2390	Train	
ergosterol	C28H44O	3296	Train	
erythreitol	C4H10O4	1528	Train	
FAME16 (methyl palmitate)	C17H34O2	1957	Train	
FAME18 (methyl stearate)	C19H38O2	2161	Train	
farnesol	C15H26O	1832	Test	bisabolol
galactosan	C6H10O5	1684	Train	
gamma dodecalactone	C12H22O2	1709	Train	
glyceric acid	C3H6O4	1352	Train	
hexadecanamide	C16H33NO	2212	Train	
hexadecanol	C16H34O	1989	Train	
homosalate	C16H22O3	2054	Test	beta-caryophyllinic acid
hydroquinone	C6H6O2	1420	Train	
ionone	C13H20O	1449	Train	
isoeugenol	C10H12O2	1591	Train	
isopimaric acid	C20H30O2	2385	Test	C14 Diacid
ketopinic acid	C10H14O3	1530	Test	pinonic acid
levoglucosan	C6H10O5	1726	Train	
linoleic acid	C18H32O2	2245	Train	
lupeol	C30H50O	3483	Train	
maltol	C6H6O3	1316	Train	
mannosan	C6H10O5	1706	Test	galactosan
MBTCA	C8H12O6	1776	Train	
Me-OH-glutatric acid	C6H10O5	1623	Test	2-ketoglutaric acid
monopalmitin	C19H38O4	2628	Test	monostearin
monostearin	C21H42O4	2788	Train	
nonanol	C9H20O	1318	Train	
octadecanal	C18H36O	2056	Train	
octadecanol	C18H38O	2191	Train	
octadecanone	C18H36O	2031	Train	
oleic acid	C18H34O2	2251	Train	

palmitoleic acid	C16H30O2	2056	Train	
p-anisic acid (4-methoxybenzoic acid)	C8H8O3	1544	Train	
pentadecanone	C15H30O	1726	Test	pinic acid, isomer 1
perylene	C20H12	2967	Train	
phthalic acid	C8H6O4	1714	Train	
phthalimide	C8H5NO2	1593	Train	
pinic acid, isomer 1	C9H14O4	1692	Train	
pinic acid, isomer 2	C9H14O4	1697	Train	
pinonic acid	C10H16O3	1550	Test	hexadecanamide
pyrene	C10H16	2171	Train	
pyrocatechol	C6H6O2	1339	Train	
quinoline	C9H7N	1278	Train	
resorcinol	C6H6O2	1399	Test	hydroquinone
retene	C18H18	2267	Train	
Sesquiterpene 1†	C15H24	1404	Train	
Sesquiterpene 2†	C15H24	1442	Test	Sesquiterpene 3
Sesquiterpene 3†	C15H24	1449	Train	
Sesquiterpene 4†	C15H24	1451	Train	
Sesquiterpene 5†	C15H24	1471	Train	
Sesquiterpene 6†	C15H24	1493	Train	
Sesquiterpene 7†	C15H24	1537	Train	
Sesquiterpene 8†	C15H24	1569	Train	
Sesquiterpene 9†	C15H24	1610	Train	
sinapinaldehyde	C11H12O4	2032	Train	
squalene	C30H50	2868	Train	
stigmasterol	C29H48O	3344	Test	ergosterol
syringaldehyde	C9H10O4	1726	Test	9H-florenone
syringic acid	C9H10O5	1924	Test	C10 diacid (sebacic acid)
syringol	C8H10O3	1418	Train	
threitol	C4H10O4	1521	Test	erythreitol
triacetin	C9H14O6	1362	Train	
tridecanal	C13H26O	1537	Train	
vanillic acid	C8H8O4	1789	Train	
vanillin	C8H8O3	1558	Train	
verbenone (-)	C10H14O	1237	Train	
xanthone	C13H8O2	1906	Train	

*Normalized by deuterated alkane standard series

†Isomer identity undetermined, only quantification factor and properties related to chemical formula included in modelling

Table 2.A2: 40 most common charged fragments featurized for mass spectral featurization, with possible formulae and implications of published peaks.

Fragment m/z	Possible Formulae	Notes
41	C3H5+	
43	C3H7+, C2H3O+	Propyl group, ketone indicator
45	CHO2	Carboxyl indicator, underivitized
55		
56		
57	C_4H_9+, C3H5O+	Signature alkane fragment, ketone/ester
67		
69		
71	C4H7O+	Ketone/ester
73	$Si(CH_3)_3$+	Indicates derivatization and therefore presence of OH group
74		
75		
77	C6H5+	phenyl
79		
81		
83		
85		
91		
92		
93	C6H5O+	Oxygenated aromatics
95		
99		
103		
105		
107		
109		
111		
113		
117		
119		
121		
129		
131		
132		
135		
145		
147		
189		
204	Si2C8H20O2+	Indicative of sugars
217	Si2C9H21O2+	Indicative of sugars

Table 2.A3: 20 most common neutral mass differences between charged peaks, selected for mass spectral featurization, with possible formulae and implications of commonly reported neutral losses.

Neutral Loss/Mass Difference (amu)	Probable Formulae/ Interpretation	Notes
1	Loss of H	
2		
3		
4		

6		
8		
10		
11		
12		
13		
14		
15	CH3	Methyl
16	O	Alcohol- derivatization agent loss
18		
20		
26		
27		
28	CO	Carbonyl
30		
42		

Table 2.A4. Full chemical properties modelling features for Hexadecane

Feature	Feature Class	Feature Input
Retention Index (d-alkane normalized)	Chromatography	1627
m/z 41	Mass Spectrum Common Fragment	238
m/z 43	Mass Spectrum Common Fragment	512
m/z 45	Mass Spectrum Common Fragment	0
m/z 55	Mass Spectrum Common Fragment	144
m/z 56	Mass Spectrum Common Fragment	116
m/z 57	Mass Spectrum Common Fragment	999
m/z 67	Mass Spectrum Common Fragment	0
m/z 69	Mass Spectrum Common Fragment	0
m/z 71	Mass Spectrum Common Fragment	757
m/z 73	Mass Spectrum Common Fragment	0
m/z 74	Mass Spectrum Common Fragment	0
m/z 75	Mass Spectrum Common Fragment	0
m/z 77	Mass Spectrum Common Fragment	0
m/z 79	Mass Spectrum Common Fragment	0
m/z 81	Mass Spectrum Common Fragment	0
m/z 83	Mass Spectrum Common Fragment	0
m/z 85	Mass Spectrum Common Fragment	519
m/z 91	Mass Spectrum Common Fragment	0
m/z 92	Mass Spectrum Common Fragment	0
m/z 93	Mass Spectrum Common Fragment	0
m/z 95	Mass Spectrum Common Fragment	0
m/z 99	Mass Spectrum Common Fragment	0
m/z 103	Mass Spectrum Common Fragment	0
m/z 105	Mass Spectrum Common Fragment	0
m/z 107	Mass Spectrum Common Fragment	0
m/z 109	Mass Spectrum Common Fragment	0
m/z 111	Mass Spectrum Common Fragment	0
m/z 113	Mass Spectrum Common Fragment	89
m/z 117	Mass Spectrum Common Fragment	0
m/z 119	Mass Spectrum Common Fragment	0

m/z 121	Mass Spectrum Common Fragment	0
m/z 129	Mass Spectrum Common Fragment	0
m/z 131	Mass Spectrum Common Fragment	0
m/z 132	Mass Spectrum Common Fragment	0
m/z 135	Mass Spectrum Common Fragment	0
m/z 145	Mass Spectrum Common Fragment	0
m/z 189	Mass Spectrum Common Fragment	0
m/z 204	Mass Spectrum Common Fragment	0
m/z 217	Mass Spectrum Common Fragment	0
loss of 1	Mass Spectrum Neutral Loss/Mass Diff.	FALSE
loss of 2	Mass Spectrum Neutral Loss/Mass Diff.	TRUE
loss of 3	Mass Spectrum Neutral Loss/Mass Diff.	FALSE
loss of 4	Mass Spectrum Neutral Loss/Mass Diff.	FALSE
loss of 6	Mass Spectrum Neutral Loss/Mass Diff.	FALSE
loss of 8	Mass Spectrum Neutral Loss/Mass Diff.	FALSE
loss of 10	Mass Spectrum Neutral Loss/Mass Diff.	FALSE
loss of 11	Mass Spectrum Neutral Loss/Mass Diff.	FALSE
loss of 12	Mass Spectrum Neutral Loss/Mass Diff.	FALSE
loss of 13	Mass Spectrum Neutral Loss/Mass Diff.	FALSE
loss of 14	Mass Spectrum Neutral Loss/Mass Diff.	TRUE
loss of 15	Mass Spectrum Neutral Loss/Mass Diff.	FALSE
loss of 16	Mass Spectrum Neutral Loss/Mass Diff.	TRUE
loss of 18	Mass Spectrum Neutral Loss/Mass Diff.	FALSE
loss of 20	Mass Spectrum Neutral Loss/Mass Diff.	FALSE
loss of 26	Mass Spectrum Neutral Loss/Mass Diff.	FALSE
loss of 27	Mass Spectrum Neutral Loss/Mass Diff.	FALSE
loss of 28	Mass Spectrum Neutral Loss/Mass Diff.	TRUE
loss of 30	Mass Spectrum Neutral Loss/Mass Diff.	TRUE
loss of 42	Mass Spectrum Neutral Loss/Mass Diff.	FALSE

3 Anthropogenic and Biological Influences on the Organic Composition of Coastal Submicron Sea Spray Aerosol

This work is adapted from:
Emily B. Franklin, Sarah Amiri, Daniel Crocker, Clare Morris, Kathryn Mayer, Jonathan Sauer, Christopher Lee, Francesca Malfatti, Christopher D. Cappa, Timothy H. Bertram, Kimberly A. Prather, Allen H. Goldstein, "Anthropogenic and Biological Influences on the Organic Composition of Coastal Submicron Sea Spray Aerosol," which is currently undergoing its second round of reviews from co-authors and will be submitted to Environmental Science and Technology in July 2022.

3.1 Abstract

The organic composition of coastal sea spray aerosol is important for both atmospheric chemistry and public health but remains poorly characterized. Coastal waters become enriched with organic material through both anthropogenic processes such as wastewater discharge and biological activity. Here we probe the chemical composition of the organic fraction of sea spray aerosol over the course of the 2019 SeaSCAPE mesocosm experiment, in which a phytoplankton bloom was facilitated in natural coastal water from San Diego, California. We apply untargeted two-dimensional gas chromatography to analysis of submicron nascent sea spray aerosol samples, reporting 754 unique organic species traced over a 19-day phytoplankton bloom experiment. Categorization and quantitative compositional analysis reveal three major findings; first, that anthropogenic species made up 30% of total submicron nascent sea spray aerosol organic mass under pre-bloom condition. Second, biological activity drove large changes within the carbon pool, transforming primary anthropogenic chemicals and creating novel biogenic chemicals. Third, biogenic marine organics are underrepresented in mass spectral databases in comparison to marine organic pollutants.

3.2 Introduction

Sea spray aerosols (SSA), the salty particles and droplets emitted from waves and bubble bursting at the ocean's surface, play an important role in atmospheric chemistry and climate over the ocean and in coastal areas. Sea spray aerosol organic content increases with decreasing aerosol size below an aerodynamic radius of 2.5 μm. Observations from field studies report values of the total submicron organic mass fraction ranging from 6%[1] to 23%[2] of submicron aerosol mass. Sea spray aerosol organic content plays a critical role in marine atmospheric chemistry, as the organic content forms a film on the exterior of salt crystal cores in Salt-Organic Carbon type aerosols, which dominate submicron sea spray aerosol.[3,4,5] Organic material has hygroscopicity, solubility, and ice nucleation-relevant properties that significantly differ from those of NaCl crystals, thus the organic content of

sea spray aerosols influences their atmospheric chemistry and climate impacts over the ocean.[4,5] In addition to altering chemical properties, the organic content of SSA is important in coastal areas, as coastal communities can be exposed to toxins such as those produced by harmful algal blooms[6,7] and hazardous marine pollutants.[8]

Coastal ocean water becomes enriched with organic material through both natural biological processes and anthropogenic pollution. Biogenic sources of organic material include ocean micro algae, which convert CO_2 to biomass that is either grazed by larger organisms or dies and is degraded by heterotrophic bacteria.[9] Anthropogenic sources of organics include urban runoff, personal care products including suncreens, wastewater discharge, trash, and shipping pollution.[10,11,12,13] In addition to directly producing new organic material through photosynthesis, marine microbes also produce biologically transformed products from biodegradation of terrestrial and anthropogenic organic precursors.[14,15,16] Organic material is not evenly distributed throughout the ocean water column; surfactants, hydrophobic organics, and other organic constituents collect at the ocean's surface to form a thin layer known as the sea surface microlayer (SSML).[17–19] Both biogenic and anthropogenic organics can be significantly enriched in the SSML, and enrichment factors of common anthropogenic pollutants including phthalates, polycyclic aromatic hydrocarbons, and heavy metals can be as high as multiple orders of magnitude.[19,20,21,22] The organic composition of sea spray aerosols, and in particular submicron film drops, is significantly influenced by the composition of the SSML, making the enrichment of some chemical classes in the SSML important in determining the organic composition of sea spray aerosol and the transfer of organic pollutants from ocean to atmosphere.[23,22,9]

There has been significant recent progress in characterizing the organic content of sea spray aerosol, including multiple methods of fractional functional group analysis.[9,24,25,26] However, a large fraction remains uncharacterized, particularly at the compound-specific level. This work leverages a mesocosm experiment in which real coastal water was used to facilitate a bloom of naturally occurring phytoplankton and bacteria species to investigate how anthropogenic pollution and marine microbes influence the organic composition of submicron sea spray aerosol in a coastal context.

3.3 Materials and Methods

3.3.1 Experimental Campaign and Sample Collection:

Samples analyzed in this work originate from the 2019 SeaSCAPE (Sea Spray Chemistry And Particle Evolution) experimental campaign organized by CAICE (the Center for Aerosol Impacts on Chemistry of the Environment) based at UCSD. This campaign is described in detail in Sauer et al., 2022.[27] Briefly, water collected at Ellen Browning Scripps Memorial Pier (hereafter Scripps Pier) in La Jolla, California was transported into the Scripps Institute of Oceanography Hydraulics Laboratory wave channel facility (hereafter called the wave channel) which uses mechanically generated breaking waves to

generate realistic sea spray aerosol. Scripps Pier is located close to multiple potential sources of coastal pollution including a high use beach, multiple stormwater discharge points, and municipal wastewater treatment outflow.[28] Nutrients were added to the natural costal water to induce a micro algae bloom and subsequent bacterial and viral blooms of the species naturally present within the coastal water. The chemical and biological properties of the water, including chlorophyll-a concentrations and abundance of phytoplankton, bacteria, and viruses, were extensively monitored both before and throughout the bloom, as documented in Sauer et al. 2022.[27] These indicators are visualized in Figure 3.S6. Phytoplankton enumeration methods are summarized in 3.SI.5.

Collection of submicron aerosol samples for speciated organic analysis at this campaign has been previously described in Franklin et al., 2021.[28] Briefly, submicron aerosols were collected on 47 mm quartz fiber filters (Pallflex Tissuequartz) using a custom designed humidity-controlled sequential sampler at a frequency of two samples per day (light synchronized, one "day" and one "night" sample per day). Samples were frozen (-18ºC) for offline compositional analysis, as described below.

3.3.2 Offline Sample Analysis

Submicron aerosol samples collected at SeaSCAPE were analyzed for speciated organic composition by thermal desorption two-dimensional gas chromatography coupled with electron ionization high resolution time of flight mass spectrometry (TD-GCxGC-EI-HR-MS, hereafter abbreviated to GCxGC-MS). This instrument separates and detects organic constituents collected on filter material as follows: organic constituents are thermally desorbed from filter material, separated by both volatility and polarity by two gas chromatography columns in sequence, and detected by an electron ionization (70 eV) high resolution time of flight mass spectrometer. This instrument is described in detail in Worton et al., 2017.[29]

3.3.3 Data Analysis

From the 38 samples collected, 754 unique organic compounds were compiled into a custom library of mass spectra and retention indices (position relative to the deuterated alkane series in the internal standard). The majority of these organics were identified in either the first sample or the sample corresponding with peak chlorophyll-a concentrations in the bulk water (see Figure 3.1 for chlorophyll-a trace). Using the custom mass spectral library, all 754 catalogued species were traced over the entire bloom. Each observation of each compound was normalized by the mean of the three nearest internal standard components to correct for drifts in instrument sensitivity over the analysis period and matrix effects, using methodology described in detail in Franklin et al., 2022.[30] Finally, the mass spectrum of each compound was searched against the NIST-14 mass spectral database. Of the 754 compounds catalogued, 14% were identifiable by database match. Compounds were defined as "identifiable" if they produced a match factor >750 with an entry in the NIST14 mass spectral database and an approximate match with a retention index for that compound published in the database or the literature.

Quantification of detected and separated organics is described in detail in Supporting Information section 1 (3.SI.1). A representative external standard containing ~130 known components was analyzed in 6-point calibration curves immediately preceding and following sample analysis. Compounds present in both the sample and standard or with clear chemical proxies in the standard were directly quantified by the calibration curves of the external standard, while the quantification factors of all other species were predicted using Ch3MS-RF.[30] Model performance and uncertainties of both modeled and directly calculated quantification factors are described and illustrated in 3.SI.1.

Segments of the filter samples utilized for speciated organic analysis were also analyzed for total organic carbon to determine the fraction of collected organic material to which GCxGC-MS analysis is sensitive to. Results indicate that 40±15% of collected submicron organic mass was recovered and quantified by GCxGC-MS analysis, consistent with prior applications of this methodology to organic aerosol collected in other contexts.[31] Examples of previously characterized contributors to submicron sea spray aerosol organic composition to which this technique is not sensitive include polysaccharides and proteinaceous material.[32,33] Methodological details and validation by complementary measurements are described in 3.SI.2.

Compounds were grouped into clusters of similar temporal variability by dynamic time warping hierarchical clustering,[34] as described in 3.SI.3. The 100 most abundant organics were grouped by cluster analysis, with results optimized by a 7-group solution as illustrated in Figure 3.S4 and Figure 3.S5. All other compounds detected with sufficient frequency for time series construction were assigned to the cluster with which they exhibited the highest Pearson correlation. Clusters were then categorized as anthropogenic (resulting from human emissions and not enhanced by biological activity), biogenic (enhanced by biological activity, either through direct production or biotransformation of previously existing species), or mixed (some combination of the two, or undetermined) influence through two methods: first, comparison of their mean cluster temporal profiles to indicators of biological activity, including chlorophyll-a concentrations and the abundances (cell/L) of microbes including phytoplankton phenotypes, heterotrophic bacteria, and virales; second, through a literature review of characteristic identifiable constituents within each cluster (Table 3.1). Justifications for these assignments are addressed in "Results and Discussion" below. The average temporal variability of each cluster relative to chlorophyll-a and heterotrophic bacteria concentrations is illustrated in Figure 3.1, while Figure 3.2 illustrates the mean Pearson correlation coefficient between each individual aerosol-phase organic and each biological activity or biomass indicator grouped by cluster. Finally, chemical properties of unidentifiable compounds, specifically average carbon oxidation state ($\overline{OS_c}$) and carbon number (n_c) were predicted using Ch3MS-RF,[30] as described in detail in 3.SI.6.

3.4 Results and Discussion

3.4.1 Pre-Bloom Contributions of Anthropogenic Compounds to Submicron Aerosol Mass

Over the first full day of analysis, 73% of the recovered organic carbon pool was attributed to compounds from anthropogenic clusters (Figure 3.3). Given the approximate organic material recovery rate of 40% (discussed in 3.SI.2), this implies that at least 30% of submicron organic material collected during this period is attributable to anthropogenic source material. Importantly, another 15% of the recovered carbon pool is attributed to mixed cluster categories that include anthropogenic compounds, as summarized in Table 3.1, meaning that the total recovered mass fraction attributable to anthropogenic organics is greater than 73%. This anthropogenic material was highly diverse, consisting of over 400 individual organic species, which is than half of total organic compounds catalogued and traced over the experimental bloom. Identities and dynamics of specific groups of interest are described below.

One compound class within the anthropogenic fraction that substantially contributed to total recovered mass was a complex mixture of aliphatic material. This population can be seen in Figure 3.5 panel A as the tightly grouped distribution of low carbon oxidation state species in the n_c range between 18 and 22. While aliphatic material in marine environments can have both biogenic and anthropogenic sources, the aliphatic signature identified in the sea spray aerosol samples was identified as petrochemical rather than biogenic for three reasons. First, the oil signature was identified as petrochemical by comparison to ambient marine oil identified in analysis of the Deep Water Horizon oil spill, which was previously analyzed on the instrument utilized in this work and reported in Drozd et al., 2015.[35] As described in the methodology of Tran et al., 1997, biogenic aliphatic material typically presents as distinct products while oils from fossilized sources presents as an unresolved complex mixture, as was identified in this case. It should be noted that Tran et al, 1997 identified aliphatic material in effluent into the ocean off the coast of San Diego as primarily terrestrial biogenic in origin, but this was due to a distinct product signature not observed in the sea spray aerosol samples collected in this study. Finally, as described in Crocker et al., 2022, the isotopic values of the submicron SSA samples collected at SeaSCAPE were significantly negative, indicating a high degree of anthropogenic influence. While these negative isotope values cannot be attributed to any single compound or group of compounds, they lend credence to an anthropogenic/petrochemical source attribution for this product class. Potential sources of petrochemical aliphatic organics at the sampling location include wastewater discharge, urban runoff, and shipping, including from small research vessels launched from Scripps Pier.[12,13] The potential of an influence from the natural petrochemical seeps at Coal Oil Point and Redondo Beach cannot be entirely ruled out, but they are unlikely to have contributed to the observed signal due to the location of the seeps relative to the sampling location (both seeps North of Scripps Pier, while the Southern California Eddy runs predominantly South to North along

the shoreline) and the distance of approximately 120 miles between Scripps Pier and the nearest seep.[13] The petrochemical signal contributed approximately 20% of the total recovered anthropogenic signal and was almost entirely attributed to cluster Anthro1. Following the behavior of this cluster, the petrochemical aliphatic signal decreased rapidly in the transition period lagging peak chlorophyll, likely indicating that these species were consumed and/or transformed by one or more of the bacterial species which increased in concentration over the bloom.[36]

Other identified and quantified organics which contributed to the anthropogenic mass fraction included compounds attributable to personal care products, as well as phthalates and PAH's, both of which have been previously reported in sea spray aerosol originating from polluted ocean water.[22] Sunscreen in particular stood out as an important potential source; 2-Ethylhexyl salicylate and a mixture of Homosalate isomers, both compounds that have been attributed to sunscreen pollution when identified in coastal waters,[10,37,38] accounted for >80% of the Antro2 cluster mass during the first day of analysis, which equated to 26% of the total anthropogenic organic mass and 19% of total recovered organic mass over the first day. While these compounds disappeared rapidly (Figures 3.1 and 3.3), which may have been attributable to volatilization or loss to surface films, their abundance at the beginning of the bloom is notable. While the presence of sunscreen products in water collected adjacent to a popular surf beach is not surprising, to our knowledge their aerosolized emissions from the coastal ocean have not been previously reported, particularly in such significant quantities.

Contributions of PAH's to pollution in coastal regions, and in particular to SSML organic composition, have been described in a range of environments including Italy, American Samoa, Southern California, and even Antarctica,[22,39,40,41] and enrichment of PAH's in sea spray aerosol compared to both dissolved and SSML phases has also been previously described.[22] In this work, 8 PAH's were isolated and identified in nascent sea spray aerosol, all of which were assigned to group Anthro1, meaning they were observed in relatively high concentrations at the beginning of the experiment before rapidly declining during the transition period between peak chlorophyll-a and peak heterotrophic bacteria in the wave channel water. A full list of the PAH compounds observed in sea spray aerosol at SeaSCAPE is provided in supporting information Table 3.S4. These compounds contributed relatively little to the total anthropogenic carbon pool, making up 1.5% of the anthropogenic attributed mass collected over the first full day of analysis.

In addition to those included in the previously described clustering and quantitative analysis, the following important anthropogenic compounds were observed. Benzophenone, another commonly reported sunscreen pollutant, was observed in the aerosol phase samples, but due to coelution with a contaminant was not able to be accurately traced and quantified. Five siloxanes were observed with temporal variability that would have placed them in either anthropogenic or mixed clusters, but they were not

included in the standard and were too chemically distinct for their quantification factors to be predicted by Ch3MS-RF and were therefore excluded from quantitative analysis.

3.4.2 Evidence for Biological Transformation of Submicron Organic Carbon Pool

Over the course of the SeaSCAPE experimental bloom, the submicron organic composition undergoes transformation from mostly anthropogenic to mostly biogenic. As described above, prior to the onset of the biological blooms, the contributions of anthropogenic organics to total submicron aerosol mass are up 73% of recovered organic mass compared to 15% mixed influence and 13% biogenic averaged over the first full day of sample collection. By the final day of the experiment however, the recovered carbon pool compositional breakdown has completely changed, and is comprised of 73% biogenic material, 19% mixed influence material, and only 8 % primary anthropogenic material. This transformation is illustrated in Figure 3.3. While the rapid loss of compounds in the Anthro2 cluster cannot be directly attributed to biological activity and contributes to the significance of differences between compositional indicators at the beginning and end of the bloom, the cumulative growth of the biogenic/biologically transformed compounds, in both absolute (Figure 3.S7) and relative (Figure 3.3) terms, demonstrates strong temporal connections to the progression of the biological bloom.

The period of most rapid turnover corresponds to the window between August 3rd and August 5th, directly between the peaks in chlorophyll-a and heterotrophic bacteria concentrations and a predatory dinoflagellate grazing event (*P. bipes*) on diatoms. Although the correlations between the bulk water chlorophyll-a concentration and increases in both biogenic organic clusters are weak, as illustrated in Figure 3.2 there are stronger associations between the two clusters and the microbial abundances. Cluster Biogenic1 contains constituents that are positively correlated with concentrations of viruses, heterotrophic bacteria, microzooplankton, and diatom dominated aggregates, while cluster Biogenic2 contains constituents that are positively correlated with diatoms and diatom-dominated aggregates. Although negative correlations between bulk water cell concentrations and biogenic organic constituents in sea spray aerosols could reasonably occur for biogenic species released from cellular matrices, which would necessarily rise only as cells were destroyed, both assigned anthropogenic clusters and the primary mixed cluster, Mixed1, peak before all major microbiological subspecies and are therefore not experiencing this phenomenon. The timings and averaging periods of the biological cell/L and sea spray aerosol samples are also relevant. Water was collected for biological activity analysis in the mornings at approximately 8:00, while the aerosol "day" samples that were compared to the water-side measurements were collected continuously from 7:00-21:00. As a result, any lagging within a 14-hour time scale would be averaged out and present as a positive unlagged Pearson's correlation. Given the complexities in the evolution of the phytoplankton, bacterial, and viral communities, the covariance between the temporal variability of different species, and the sensitivities of the aerosol-based measurements to

variables including pressure and temperature, positive correlations cannot be used to definitively propose causal relationships between individual microbial species and individual compounds. That said, the presence of relatively strong average correlation coefficients between compounds assigned to the biogenic source groups and negative or very weak correlations between the other source groups and species-specific biomass concentrations supports the assignments of the two biogenic clusters.

There is additional molecular-level evidence for biological transformation of the submicron carbon pool. One identifiable constituent in cluster Biogenic1, phthalic anhydride, has been previously identified as a product of bacterial degradation of phthalates, which were observed in the anthropogenic clusters (Table 3.1). Additional identity-based analysis of the biogenic fraction is made difficult by the high proportion of biogenic compounds that are not identifiable by database match, as discussed below. Importantly, more commonly reported biogenic products including fatty acids were observed, but as summarized in Table 3.1 they followed temporal variability that more closely matched that of compounds with anthropogenic sources and were therefore assigned to a mixed influence cluster. The effects of the biological turnover on the distribution of mass in chemical properties space is visualized in Figure 3.4. Between the beginning of the bloom and the primary transition period between peak chlorophyll and peak heterotrophic bacteria (panels A and B), changes in property distribution can be observed, as anthropogenic aliphatic compounds with high n_c and low $\overline{OS_c}$ are lost while lower carbon number and more oxidized mixed influence and biogenic species appear or increase in importance. There is some evidence of the production of high carbon number slightly more oxidized products in the high carbon number biogenic compounds visible in panel B, some of which could be indicative of biotransformation products of the aliphatic signature, but it is important to note that previous studies have identified that GCxGC analysis is less sensitive to oxidized oil biodegradation products than to unoxidized precursors.[42] Finally, on the last day of analysis (depicted in panel C), both anthropogenic and mixed influence compounds are diminished in comparison to two unidentifiable compounds in the cluster Biogenic1, discussed in greater detail below.

3.4.3 Knowledge Bias Against Biologically Transformed and Produced Organics

As previously noted, analysis of the biogenic influence clusters is made more difficult by the fact that the biogenic cluster organic mass pool is significantly under characterized compared to the anthropogenic and mixed fractions. As illustrated in Figure 3.5, over 50% of organic mass attributed to the anthropogenic and mixed influence clusters was attributable to compounds that could be identified by database match compared to less than 15% of the summed biogenic clusters. Given the transition from anthropogenic-dominated to biogenic-dominated organic carbon over the course of the bloom, this led to a consistent decrease in the fraction of organic material collected in each sample that could be identified, ranging from approximately 60% of the mass in the first sample to 25% in the

last (Figure 3.S8). This phenomenon is in part attributable to the presence and abundance of two compounds in cluster Biogenic 1 which dominated recovered mass at the end of the bloom, neither of which could be identified. The mass spectra of these two compounds, along with their n-alkane equivalent Kovats indices in a semistandard nonpolar column and predicted n_c and $\overline{OS_c}$ are discussed in 3.SI.8. The spectra of these compounds will be made available through UCB-GLOBES (University of California, Berkeley Goldstein Library of Biogenic and Environmental Spectra) along with all other compounds traced over the SeaSCAPE bloom for use by the community in comparing field samples and identifying environmentally important species.

3.5 Acknowledgements

We thank the NSF Center for Aerosol Impacts on Chemistry of the Environment (an NSF Chemical Innovation Center, (CHE-1801971)) and the entire experimental campaign team which participate d in the SeaSCAPE for their support of this work. We gratefully acknowledge support from the National Science Foundation Graduate Research Fellowship Program (DGE- 1752814).

3.6 References

(1) Quinn, Patricia; Bates, Timothy; Schulz, Kristen; Coffman, D.; Frossard, A.; Russell, L.; Keene, W.; Kieber, D. Contribution of Sea Surface Carbon Pool to Organic Matter Enriuchment in Sea Spray Aerosol. *Nat. Geosci.* **2014**.

(2) Cravigan, L. T.; Mallet, M. D.; Vaattovaara, P.; Harvey, M. J.; Law, C. S.; Modini, R. L.; Russell, L. M.; Stelcer, E.; Cohen, D. D.; Olsen, G.; Safi, K.; Burrell, T. J.; Ristovski, Z. Sea Spray Aerosol Organic Enrichment, Water Uptake and Surface Tension Effects. *Atmos. Chem. Phys.* **2020**, *20* (13), 7955–7977. https://doi.org/10.5194/acp-20-7955-2020.

(3) Tervahattu, H.; Juhanoja, J.; Kupiainen, K. Identification of an Organic Coating on Marine Aerosol Particles by TOF-SIMS. *J. Geophys. Res. Atmos.* **2002**, *107* (D16), ACH 18-1. https://doi.org/10.1029/2001JD001403.

(4) Prather, K. A.; Bertram, T. H.; Grassian, V. H.; Deane, G. B.; Stokes, M. D.; DeMott, P. J.; Aluwihare, L. I.; Palenik, B. P.; Azam, F.; Seinfeld, J. H.; Moffet, R. C.; Molina, M. J.; Cappa, C. D.; Geiger, F. M.; Roberts, G. C.; Russell, L. M.; Ault, A. P.; Baltrusaitis, J.; Collins, D. B.; Corrigan, C. E.; Cuadra-Rodriguez, L. A.; Ebben, C. J.; Forestieri, S. D.; Guasco, T. L.; Hersey, S. P.; Kim, M. J.; Lambert, W. F.; Modini, R. L.; Mui, W.; Pedler, B. E.; Ruppel, M. J.; Ryder, O. S.; Schoepp, N. G.; Sullivan, R. C.; Zhao, D. Bringing the Ocean into the Laboratory to Probe the Chemical Complexity of Sea Spray Aerosol. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (19), 7550–7555. https://doi.org/10.1073/pnas.1300262110.

(5) Lee, H. D.; Morris, H. S.; Laskina, O.; Sultana, C. M.; Lee, C.; Jayarathne, T.; Cox,

J. L.; Wang, X.; Hasenecz, E. S.; Demott, P. J.; Bertram, T. H.; Cappa, C. D.; Stone, E. A.; Prather, K. A.; Grassian, V. H.; Tivanski, A. V. Organic Enrichment, Physical Phase State, and Surface Tension Depression of Nascent Core-Shell Sea Spray Aerosols during Two Phytoplankton Blooms. *ACS Earth Sp. Chem.* **2020**, *4* (4), 650–660. https://doi.org/10.1021/acsearthspacechem.0c00032.

(6) Kirkpatrick, B.; Fleming, L. E.; Squicciarini, D.; Backer, L. C.; Clark, R.; Abraham, W.; Benson, J.; Cheng, Y. S.; Johnson, D.; Pierce, R.; Zaias, J.; Bossart, G. D.; Baden, D. G. Literature Review of Florida Red Tide: Implications for Human Health Effects. *Harmful Algae*. Elsevier April 1, 2004, pp 99–115. https://doi.org/10.1016/j.hal.2003.08.005.

(7) Grattan, L. M.; Holobaugh, S.; Morris, J. G. Harmful Algal Blooms and Public Health. *Harmful Algae*. Elsevier B.V. July 1, 2016, pp 2–8. https://doi.org/10.1016/j.hal.2016.05.003.

(8) Walsh, J. J.; Lenes, J. M.; Weisberg, R. H.; Zheng, L.; Hu, C.; Fanning, K. A.; Snyder, R.; Smith, J. More Surprises in the Global Greenhouse: Human Health Impacts from Recent Toxic Marine Aerosol Formations, Due to Centennial Alterations of World-Wide Coastal Food Webs. *Marine Pollution Bulletin*. Elsevier Ltd March 15, 2017, pp 9–40. https://doi.org/10.1016/j.marpolbul.2016.12.053.

(9) Cochran, R. E.; Laskina, O.; Trueblood, J. V.; Estillore, A. D.; Morris, H. S.; Jayarathne, T.; Sultana, C. M.; Lee, C.; Lin, P.; Laskin, J.; Laskin, A.; Dowling, J. A.; Qin, Z.; Cappa, C. D.; Bertram, T. H.; Tivanski, A. V.; Stone, E. A.; Prather, K. A.; Grassian, V. H. Molecular Diversity of Sea Spray Aerosol Particles: Impact of Ocean Biology on Particle Composition and Hygroscopicity. *Chem* **2017**, *2* (5), 655–667. https://doi.org/10.1016/j.chempr.2017.03.007.

(10) Tovar-Sánchez, A.; Sánchez-Quiles, D.; Basterretxea, G.; Benedé, J. L.; Chisvert, A.; Salvador, A.; Moreno-Garrido, I.; Blasco, J. Sunscreen Products as Emerging Pollutants to Coastal Waters. *PLoS One* **2013**, *8* (6), e65451. https://doi.org/10.1371/JOURNAL.PONE.0065451.

(11) Vila-Costa, M.; Cerro-Gálvez, E.; Martínez-Varela, A.; Casas, G.; Dachs, J. Anthropogenic Dissolved Organic Carbon and Marine Microbiomes. *ISME Journal*. Springer Nature October 1, 2020, pp 2646–2648. https://doi.org/10.1038/s41396-020-0712-5.

(12) Tran, K.; Yu, C. C.; Zeng, E. Y. Organic Pollutants in the Coastal Environment off San Diego, California. 2. Petrogenic and Biogenic Sources of Aliphatic Hydrocarbons. *Environ. Toxicol. Chem.* **1997**, *16* (2), 189–195. https://doi.org/10.1002/ETC.5620160213.

(13) DiGiacomo, P. M.; Washburn, L.; Holt, B.; Jones, B. H. Coastal Pollution Hazards in Southern California Observed by SAR Imagery: Stormwater Plumes, Wastewater Plumes, and Natural Hydrocarbon Seeps. *Mar. Pollut. Bull.* **2004**, *49* (11–12), 1013–1024. https://doi.org/10.1016/J.MARPOLBUL.2004.07.016.

(14) Harayama, S.; Kishira, H.; Kasai, Y.; Shutsubo, K. Petroleum Biodegradation in Marine Environments. *Pet. Biodegrad. 63 J. Molec. Microbiol. Biotechnol* **1999**, *1* (1), 63–70.

(15) Gao, J.; Chi, J. Biodegradation of Phthalate Acid Esters by Different Marine Microalgal Species. *Mar. Pollut. Bull.* **2015**, *99* (1–2), 70–75. https://doi.org/10.1016/J.MARPOLBUL.2015.07.061.

(16) Baena-Nogueras, R. M.; González-Mazo, E.; Lara-Martín, P. A. Degradation Kinetics of Pharmaceuticals and Personal Care Products in Surface Waters: Photolysis vs Biodegradation. *Sci. Total Environ.* **2017**, *590–591*, 643–654. https://doi.org/10.1016/J.SCITOTENV.2017.03.015.

(17) Cunliffe, M.; Engel, A.; Frka, S.; Gašparović, B. Ž.; Guitart, C.; Murrell, J. C.; Salter, M.; Stolle, C.; Upstill-Goddard, R.; Wurl, O. Sea Surface Microlayers: A Unified Physicochemical and Biological Perspective of the Air–Ocean Interface. *Prog. Oceanogr.* **2013**, *109*, 104–116. https://doi.org/10.1016/J.POCEAN.2012.08.004.

(18) Engel, A.; Bange, H. W.; Cunliffe, M.; Burrows, S. M.; Friedrichs, G.; Galgani, L.; Herrmann, H.; Hertkorn, N.; Johnson, M.; Liss, P. S.; Quinn, P. K.; Schartau, M.; Soloviev, A.; Stolle, C.; Upstill-Goddard, R. C.; van Pinxteren, M.; Zäncker, B. The Ocean's Vital Skin: Toward an Integrated Understanding of the Sea Surface Microlayer. *Front. Mar. Sci.* **2017**, *4* (MAY), 165. https://doi.org/10.3389/FMARS.2017.00165/BIBTEX.

(19) Hardy, J. T. The Sea Surface Microlayer: Biology, Chemistry and Anthropogenic Enrichment. *Progress in Oceanography*. Pergamon January 1, 1982, pp 307–328. https://doi.org/10.1016/0079-6611(82)90001-5.

(20) Marty, J. C.; Saliot, A.; Buat-Ménard, P.; Chesselet, R.; Hunter, K. A. Relationship between the Lipid Compositions of Marine Aerosols, the Sea Surface Microlayer, and Subsurface Water. *J. Geophys. Res. Ocean.* **1979**, *84* (C9), 5707–5716. https://doi.org/10.1029/JC084IC09P05707.

(21) Wurl, O.; Obbard, J. P. A Review of Pollutants in the Sea-Surface Microlayer (SML): A Unique Habitat for Marine Organisms. *Mar. Pollut. Bull.* **2004**, *48* (11–12), 1016–1030. https://doi.org/10.1016/J.MARPOLBUL.2004.03.016.

(22) Cincinelli, A.; Stortini, A. M.; Perugini, M.; Checchini, L.; Lepri, L. Organic Pollutants in Sea-Surface Microlayer and Aerosol in the Coastal Environment of Leghorn—(Tyrrhenian Sea). *Mar. Chem.* **2001**, *76* (1–2), 77–98. https://doi.org/10.1016/S0304-4203(01)00049-4.

(23) Wang, X.; Deane, G. B.; Moore, K. A.; Ryder, O. S.; Stokes, M. D.; Beall, C. M.; Collins, D. B.; Santander, M. V.; Burrows, S. M.; Sultana, C. M.; Prather, K. A. The Role of Jet and Film Drops in Controlling the Mixing State of Submicron Sea Spray Aerosol Particles. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (27), 6978–6983. https://doi.org/10.1073/pnas.1702420114.

(24) Cravigan, L. T.; Mallet, M. D.; Vaattovaara, P.; Harvey, M. J.; Law, C. S.; Modini, R. L.; Russell, L. M.; Stelcer, E.; Cohen, D. D.; Olsen, G.; Safi, K.; Burrell, T. J.; Ristovski, Z. Sea Spray Aerosol Organic Enrichment, Water Uptake and Surface Tension Effects. *Atmos. Chem. Phys. Discuss.* **2019**, 1–35. https://doi.org/10.5194/acp-2019-797.

(25) Bertram, T. H.; Cochran, R. E.; Grassian, V. H.; Stone, E. A. Sea Spray Aerosol Chemical Composition: Elemental and Molecular Mimics for Laboratory Studies of Heterogeneous and Multiphase Reactions. *Chem. Soc. Rev.* **2018**, *47* (7), 2374–2400. https://doi.org/10.1039/C7CS00008A.

(26) Cochran, R. E.; Laskina, O.; Jayarathne, T.; Laskin, A.; Laskin, J.; Lin, P.; Sultana, C.; Lee, C.; Moore, K. A.; Cappa, C. D.; Bertram, T. H.; Prather, K. A.; Grassian, V. H.; Stone, E. A. Analysis of Organic Anionic Surfactants in Fine and Coarse Fractions of Freshly Emitted Sea Spray Aerosol. *Environ. Sci. Technol.* **2016**, *50* (5), 2477–2486. https://doi.org/10.1021/ACS.EST.5B04053/SUPPL_FILE/ES5B04053_SI_001.PDF.

(27) Sauer, J. S.; Mayer, K. J.; Lee, C.; Alves, M. R.; Amiri, S.; Bahaveolos, C. J.; Franklin, E. B.; Crocker, D. R.; Dang, D.; Dinasquet, J.; Garofalo, L. A.; Kaluarachchi, C. P.; Kilgour, D. B.; Mael, L. E.; Mitts, B. A.; Moon, D. R.; Moore, A. N.; Morris, C. K.; Mullenmeister, C. A.; Ni, C. M.; Pendergraft, M. A.; Petras, D.; Simpson, R. M. C.; Smith, S.; Tumminello, P. R.; Walker, J. L.; Demott, P. J.; Farmer, D. K.; Goldstein, A. H.; Grassian, V. H.; Jaffe, J. S.; Malfatti, F.; Martz, T. R.; Slade, J. H.; Tivanski, A. V.; Bertram, T. H.; Cappa, C. D.; Prather, K. A. The Sea Spray Chemistry and Particle Evolution Study (SeaSCAPE): Overview and Experimental Methods. *Environ. Sci. Process. Impacts* **2022**, *24* (2), 290–315. https://doi.org/10.1039/D1EM00260K.

(28) Franklin, E. B.; Alves, M. R.; Moore, A. N.; Kilgour, D. B.; Novak, G. A.; Mayer, K.; Sauer, J. S.; Weber, R. J.; Dang, D.; Winter, M.; Lee, C.; Cappa, C. D.;

Bertram, T. H.; Prather, K. A.; Grassian, V. H.; Goldstein, A. H. Atmospheric Benzothiazoles in a Coastal Marine Environment. *Environ. Sci. Technol.* **2021**, *55* (23), 15705–15714. https://doi.org/10.1021/ACS.EST.1C04422/SUPPL_FILE/ES1C04422_SI_001.PDF.

(29) Worton, D. R.; Decker, M.; Isaacman-VanWertz, G.; Chan, A. W. H.; Wilson, K. R.; Goldstein, A. H. Improved Molecular Level Identification of Organic Compounds Using Comprehensive Two-Dimensional Chromatography, Dual Ionization Energies and High Resolution Mass Spectrometry. *Analyst* **2017**, *142* (13), 2395–2403. https://doi.org/10.1039/c7an00625j.

(30) Franklin, E. B.; Yee, L. D.; Aumont, B.; Weber, R. J.; Grigas, P.; Goldstein, A. Ch3MS-RF: A Random Forest Model for Chemical Characterization and Improved Quantification of Unidentified Atmospheric Organics Detected by Chromatography-Mass Spectrometry Techniques. *Atmos. Meas. Tech- Discuss.* **2022**. https://doi.org/10.5194/amt-2022-99.

(31) Zhang, H.; Yee, L. D.; Lee, B. H.; Curtis, M. P.; Worton, D. R.; Isaacman-VanWertz, G.; Offenberg, J. H.; Lewandowski, M.; Kleindienst, T. E.; Beaver, M. R.; Holder, A. L.; Lonneman, W. A.; Docherty, K. S.; Jaoui, M.; Pye, H. O. T.; Hu, W.; Day, D. A.; Campuzano-Jost, P.; Jimenez, J. L.; Guo, H.; Weber, R. J.; De Gouw, J.; Koss, A. R.; Edgerton, E. S.; Brune, W.; Mohr, C.; Lopez-Hilfiker, F. D.; Lutz, A.; Kreisberg, N. M.; Spielman, S. R.; Hering, S. V.; Wilson, K. R.; Thornton, J. A.; Goldstein, A. H. Monoterpenes Are the Largest Source of Summertime Organic Aerosol in the Southeastern United States. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (9), 2038–2043. https://doi.org/10.1073/PNAS.1717513115/SUPPL_FILE/PNAS.1717513115.SD01.RTF.

(32) Aller, J. Y.; Radway, J. A. C.; Kilthau, W. P.; Bothe, D. W.; Wilson, T. W.; Vaillancourt, R. D.; Quinn, P. K.; Coffman, D. J.; Murray, B. J.; Knopf, D. A. Size-Resolved Characterization of the Polysaccharidic and Proteinaceous Components of Sea Spray Aerosol. *Atmos. Environ.* **2017**, *154*, 331–347. https://doi.org/10.1016/J.ATMOSENV.2017.01.053.

(33) Hasenecz, E. S.; Jayarathne, T.; Pendergraft, M. A.; Santander, M. V.; Mayer, K. J.; Sauer, J.; Lee, C.; Gibson, W. S.; Kruse, S. M.; Malfatti, F.; Prather, K. A.; Stone, E. A. Marine Bacteria Affect Saccharide Enrichment in Sea Spray Aerosol during a Phytoplankton Bloom. *ACS Earth Sp. Chem.* **2020**, *4* (9), 1638–1649. https://doi.org/10.1021/ACSEARTHSPACECHEM.0C00167/ASSET/IMAGES/LARGE/SP0C00167_0005.JPEG.

(34) Łuczak, M. Hierarchical Clustering of Time Series Data with Parametric Derivative Dynamic Time Warping. *Expert Syst. Appl.* **2016**, *62*, 116–130. https://doi.org/10.1016/J.ESWA.2016.06.012.

(35) Drozd, G. T.; Worton, D. R.; Aeppli, C.; Reddy, C. M.; Zhang, H.; Variano, E.; Goldstein, A. H. Modeling Comprehensive Chemical Composition of Weathered Oil Following a Marine Spill to Predict Ozone and Potential Secondary Aerosol Formation and Constrain Transport Pathways. *J. Geophys. Res. Ocean.* **2015**. https://doi.org/10.1002/2015JC011093.

(36) Yakimov, M. M.; Timmis, K. N.; Golyshin, P. N. Obligate Oil-Degrading Marine Bacteria. *Curr. Opin. Biotechnol.* **2007**, *18* (3), 257–266. https://doi.org/10.1016/J.COPBIO.2007.04.006.

(37) Mitchelmore, C. L.; He, K.; Gonsior, M.; Hain, E.; Heyes, A.; Clark, C.; Younger, R.; Schmitt-Kopplin, P.; Feerick, A.; Conway, A.; Blaney, L. Occurrence and Distribution of UV-Filters and Other Anthropogenic Contaminants in Coastal Surface Water, Sediment, and Coral Tissue from Hawaii. *Sci. Total Environ.* **2019**, *670*, 398–410. https://doi.org/10.1016/J.SCITOTENV.2019.03.034.

(38) Bargar, T. A.; Alvarez, D. A.; Garrison, V. H. Synthetic Ultraviolet Light Filtering Chemical Contamination of Coastal Waters of Virgin Islands National Park, St. John, U.S. Virgin Islands. *Mar. Pollut. Bull.* **2015**, *101* (1), 193–199. https://doi.org/10.1016/J.MARPOLBUL.2015.10.077.

(39) Polidoro, B. A.; Comeros-Raynal, M. T.; Cahill, T.; Clement, C. Land-Based Sources of Marine Pollution: Pesticides, PAHs and Phthalates in Coastal Stream Water, and Heavy Metals in Coastal Stream Sediments in American Samoa. *Mar. Pollut. Bull.* **2017**, *116* (1–2), 501–507. https://doi.org/10.1016/J.MARPOLBUL.2016.12.058.

(40) Zeng, E. Y.; Vista, C. L. Organic Pollutants in the Coastal Environment off San Diego, California. 1. Source Identification and Assessment by Compositional Indices of Polycyclic Aromatic Hydrocarbons. *Environ. Toxicol. Chem.* **1997**, *16* (2), 179–188. https://doi.org/10.1002/etc.5620160212.

(41) Cincinelli, A.; Stortini, A. M.; Checchini, L.; Martellini, T.; Bubba, M. Del; Lepri, L. Enrichment of Organic Pollutants in the Sea Surface Microlayer (SML) at Terra Nova Bay, Antarctica: Influence of SML on Superficial Snow Composition. *J. Environ. Monit.* **2005**. https://doi.org/10.1039/b507321a.

(42) Nowak, J. A.; Shrestha, P. M.; Weber, R. J.; McKenna, A. M.; Chen, H.; Coates, J. D.; Goldstein, A. H. Comprehensive Analysis of Changes in Crude Oil Chemical

Composition during Biosouring and Treatments. *Environ. Sci. Technol.* **2018**, *52* (3), 1290–1300. https://doi.org/10.1021/ACS.EST.7B05346/SUPPL_FILE/ES7B05346_SI_001.PDF.

(43) Manodori, L.; Gambaro, A.; Piazza, R.; Ferrari, S.; Stortini, A. M.; Moret, I.; Capodaglio, G. PCBs and PAHs in Sea-Surface Microlayer and Sub-Surface Water Samples of the Venice Lagoon (Italy). *Mar. Pollut. Bull.* **2006**, *52* (2), 184–192. https://doi.org/10.1016/J.MARPOLBUL.2005.08.017.

(44) Wu, Y. L.; Wang, X. H.; Li, Y. Y.; Hong, H. S. Occurrence of Polycyclic Aromatic Hydrocarbons (PAHs) in Seawater from the Western Taiwan Strait, China. *Mar. Pollut. Bull.* **2011**, *63* (5–12), 459–463. https://doi.org/10.1016/J.MARPOLBUL.2011.03.008.

(45) Cross, J. N.; Hardy, J. T.; Hose, J. E.; Hershelman, G. P.; Antrim, L. D.; Gossett, R. W.; Crecelius, E. A. Contaminant Concentrations and Toxicity of Sea-Surface Microlayer near Los Angeles, California. *Mar. Environ. Res.* **1987**, *23* (4), 307–323. https://doi.org/10.1016/0141-1136(87)90024-9.

(46) Tovar-Sánchez, A.; Sánchez-Quiles, D.; Basterretxea, G.; Benedé, J. L.; Chisvert, A.; Salvador, A.; Moreno-Garrido, I.; Blasco, J. Sunscreen Products as Emerging Pollutants to Coastal Waters. *PLoS One* **2013**, *8* (6), e65451. https://doi.org/10.1371/journal.pone.0065451.

(47) Giam, C. S.; Chan, H. S.; Neff, G. S.; Atlas, E. L. Phthalate Ester Plasticizers: A New Class of Marine Pollutant. *Science (80-.).* **1978**, *199* (4327), 419–421. https://doi.org/10.1126/SCIENCE.413194.

(48) Dsikowitzky, L.; Schwarzbauer, J.; Kronimus, A.; Littke, R. The Anthropogenic Contribution to the Organic Load of the Lippe River (Germany). Part I: Qualitative Characterisation of Low-Molecular Weight Organic Compounds. *Chemosphere* **2004**, *57* (10), 1275–1288. https://doi.org/10.1016/J.CHEMOSPHERE.2004.08.052.

(49) Madsen, R. B.; Zhang, H.; Biller, P.; Goldstein, A. H.; Glasius, M. Characterizing Semivolatile Organic Compounds of Biocrude from Hydrothermal Liquefaction of Biomass. *Energy and Fuels* **2017**, *31* (4), 4122–4134. https://doi.org/10.1021/ACS.ENERGYFUELS.7B00160/SUPPL_FILE/EF7B00160_SI_001.PDF.

(50) Wright, R. J.; Bosch, R.; Gibson, M. I.; Christie-Oleza, J. A. Plasticizer Degradation by Marine Bacterial Isolates: A Proteogenomic and Metabolomic

Characterization. *Environ. Sci. Technol.* **2020**, *54* (4), 2244–2256. https://doi.org/10.1021/ACS.EST.9B05228/SUPPL_FILE/ES9B05228_SI_008.XLSX.

(51) Moore, G. E.; Thomas, R. S.; Monkman, J. L. The Routine Determination of Polycyclic Hydrocarbons in Airborne Pollutants. *J. Chromatogr. A* **1967**, *26* (C), 456–464. https://doi.org/10.1016/S0021-9673(01)98903-4.

3.7 Tables and Figures

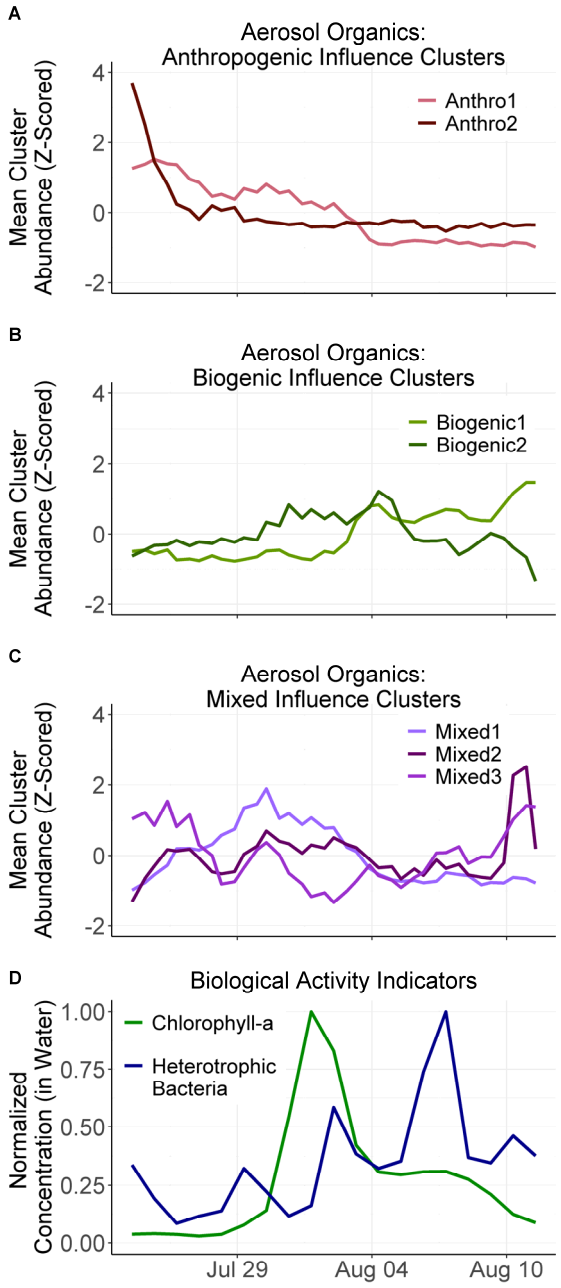


Figure 3.1. Mean temporal profiles of nascent sea spray aerosol organic material clusters over the course of the 2019 SeaSCAPE mesocosm experiment. Panel A illustrates the clusters attributed to anthropogenic influences, panel B illustrates clusters attributed to biogenically influenced production processes, panel C illustrates clusters attributed to mixed influences, and panel D illustrates contextual biological activity markers measured in the wave channel water, specifically chlorophyll-a concentrations and heterotrophic bacteria cell concentrations. The aerosol organic concentration time series of each compound is smoothed by 3-point moving average and z-scored, and all constituents assigned to an influence group are averaged to yield mean factor profiles.

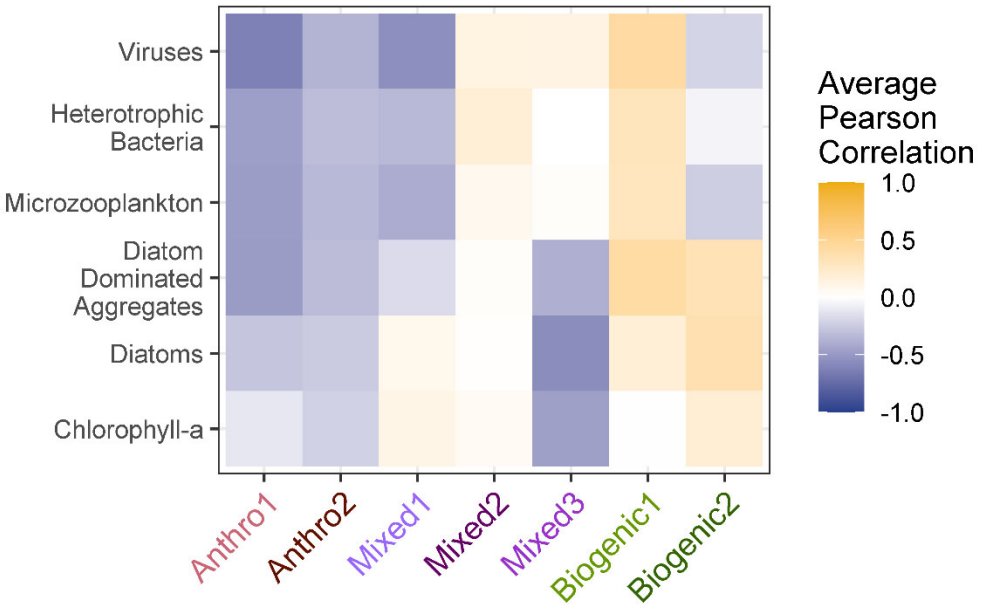


Figure 3.2. Average Pearson correlation coefficients between constituents in aerosol-phase clusters and water-side biological activity indicators during the SeaSCAPe mesocosm experiment.

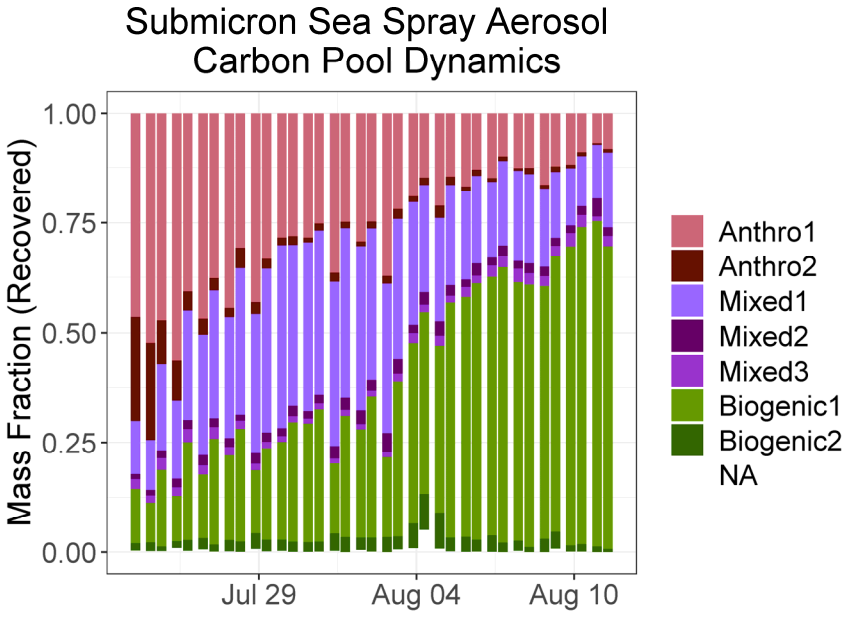


Figure 3.3. Submicron sea spray aerosol carbon pool mass attribution to anthropogenic, mixed, and biogenic influence groupings from sea spray aerosol samples collected over the course of the 2019 SeaSCAPE mesocosm experiment.NA indicates mass attributable to compounds that were too infrequently observed for assignment to time series grouping.

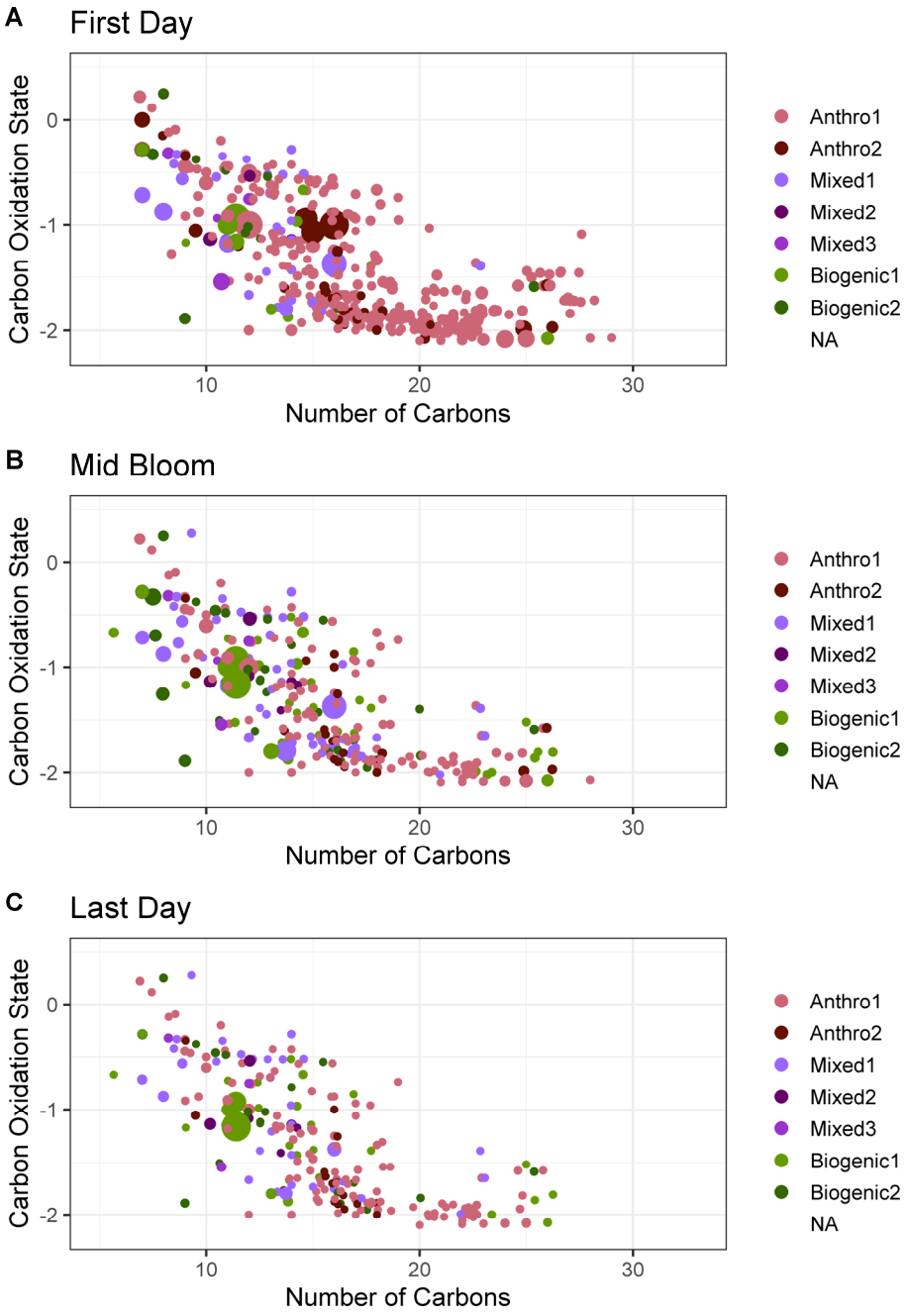


Figure 3.4. Chemical properties distributions in OSc-n_c space and cluster assignments of speciated aerosol-phase organics detected on the first day (panel A), the middle of the bloom (panel B, daytime on August 4th) and the last day of sample collection (panel C). Marker size indicates mass concentration in ng/m^3 scaled to the most abundant compound detected in the sample.

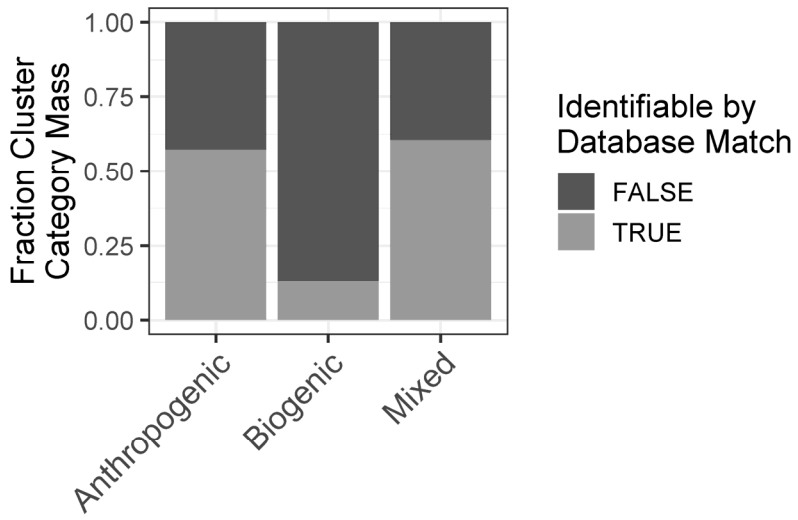


Figure 3.5. Fraction of nascent SSA organic mass (recovered by GCxGC) attributable to identifiable vs unidentifiable organic compounds averaged over the SeaSCAPE mesocosm experiment.

Table 3.1. Identities, potential sources, and references describing representative identifiable constituents of temporal variability clusters used to group submicron nascent sea spray aerosol organic compounds catalogued and traced by GCxGC from samples collected at the SeaSCAPE mesocosm experiment.

Cluster	Representative Compound	Potential Source(s)	References
Anthropogenic 1	7-Methyl-benzanthracene	PAH, incomplete combustion	Manodori, 2006[43] Wu, 2011[44] Cross, 1987[45]
	4-Methyl-pyrene*	PAH, incomplete combustion	
Anthropogenic 2	Homosalate	Sunscreen	Tovar-Sanchez, 2013[46]; Bargar, 2015[38]
	2-Ethylhexyl salicylate	Sunscreen	Mitchelmore, 2019[37]
	Dibutyl phthalate	Plasticizer	Giam, 1978[47]; Cincinelli, 2001[22]
Anthropogenic 3	Diethyl phthalate	Plasticizer	Polidoro, 2017[39]
	2-Butenedioic acid, dibutyl ester	Previously reported in polluted waterways, though source undetermined	Dsikowitzky, 2004[48]
Biogenic 1	Dimethyl quinoline	Marine biooil component*	Madsen, 2017[49]
Biogenic 2	Phthalic anhydride	Bacterial degradation of phthalates	Wright, 2020[50]
Mixed 1	Benzothiazole	Urban effluents	Franklin, 2021[28]

	Palmitic Acid	Marine microbiology (including diatoms)	Tervahattu 2002, Kates 1966
	2,2,4-Trimethyl-1,3-pentanediol diisobutyrate	Plasticizer, commonly used in food packaging	Kempf, 2009
Mixed 2	Trimethyl quinoline	Marine biooil component*	Madsen, 2017
Mixed 3	butylphthalimide	Plasticizer, personal care product component	Dionisio, 2018

* Note that reference refers to products observed in biocrudes rather than direct products, and as such observed species could potentially be instrument thermal decomposition products of chemically distinct original bioproducts.

*Note that references identify the prevalence and sources of the compound class PAH to which this compound belongs rather than identifying this species specifically. This compound has been reported in urban air pollution[51] but has not been explicitly referenced in marine pollution literature to our knowledge.

3.8 Supporting Information

3.SI.1: Quantification: Methods, Performance, and Uncertainties

Compounds directly quantified by authentic external standards using the GCxGC methodology described in this work have a quantification uncertainty of approximately 10%.[1–3] Quantification uncertainty for compounds quantified by Ch3MS-RF is summarized below.

Performance Metric	All Test Set Compounds	Test Set Compounds, > RI 1500*
OSR^2	.17	.99
Mean Average Error	.013	.00069
Root Mean Square Error	.043	.00091
Mean % error	12 %	8.8 %
Median % error	-18 %	-18 %
Absolute % error geometric mean	32 %	28 %

* Restriction of test set from all compounds to those with retention indices above 1500 reduces the number of compounds utilized for performance evaluation from 23 to 19

Table 3.S1. Performance metrics for random forest modelling of quantification factors through methodology described in Franklin, 2022[4]

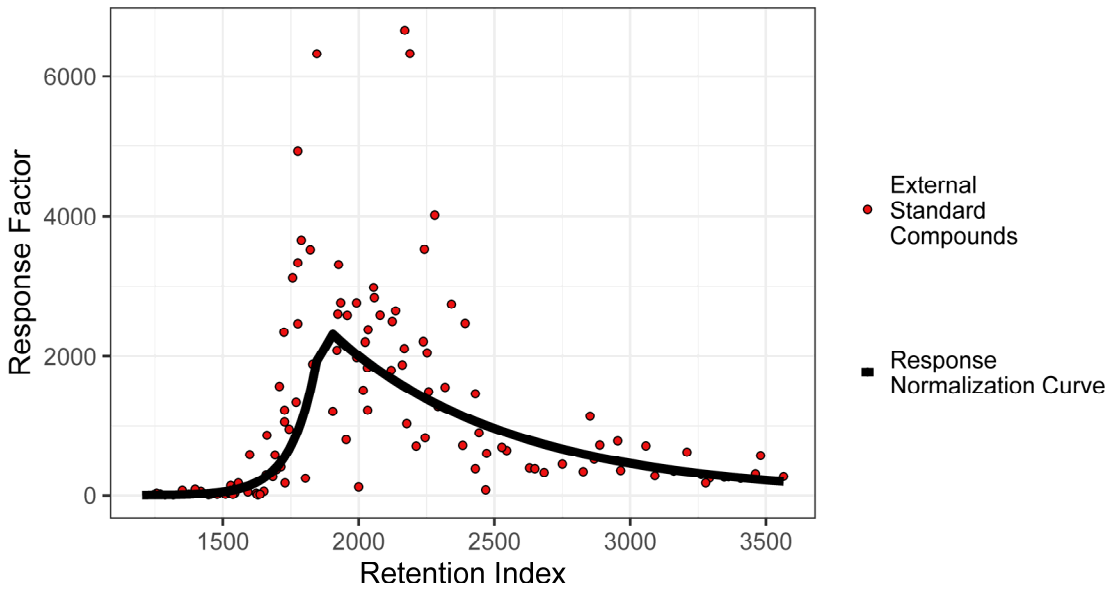


Figure 3.S1. Response normalization curve utilized for quantification factor modelling of unidentifiable compounds measured over the course of the 2019 SeaSCAPE mesocosm experiment.

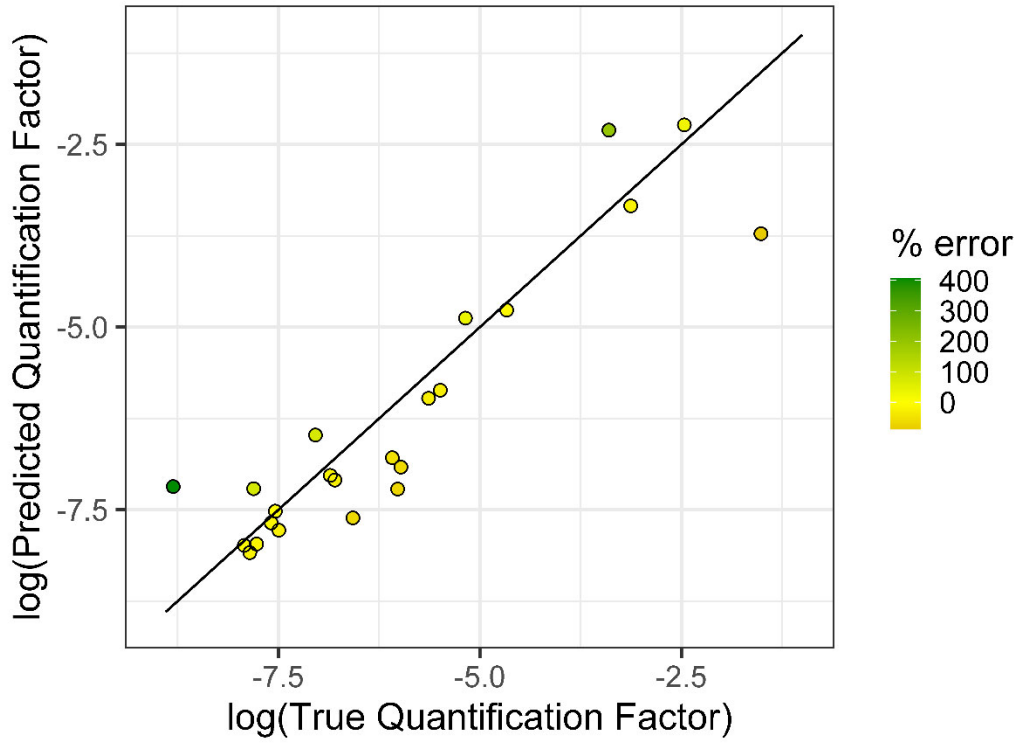


Figure 3.S2. Comparison of true and predicted quantification factors for external standard compounds withheld from quantification model training and utilized for model validation. Coloration indicates predicted quantification factor % error, which directly propagates to potential quantification errors.

3.SI.2: Determination of Organic Carbon Recovery

Segments of each filter sample and 4 field blanks collected during the SeaSCAPE mesocosm experiment and utilized for GCxGC analysis were analyzed for the mass concentrations of organic carbon and elemental carbon per unit filter area by the NIOSH87 protocol (described in Wu, 2016[5]). As anticipated given the source, elemental carbon concentrations were below detection limits in all cases. Organic carbon in the samples was determined by subtracting the geometric mean of organic carbon quantified from the field blanks. As quantification of individual species by GCxGC-MS modeling produces units of total organic mass rather than organic carbon, a correction from organic carbon to organic mass is required to determine what fraction of total collected organic carbon was observable by GCxGC. The organic mass to organic carbon ratio was approximated by calculating the average organic mass to organic carbon ratios of all identifiable constituents within the SeaSCAPE sample, a set of 105 compounds. The average organic mass to organic carbon ratio identified was 1.3, which is consistent with ranges reported in Russell et al., 2003.[6] While there is some variability in the implied mass coverage, which may in part be due to changes in the organic mass to organic carbon ratio over the bloom, average mass coverage was identified to be 40 ± 15%, with 40% indicating the average and 15% indicating the standard deviation of mass coverage percentages calculated for each filter sample. This is consistent with findings reported in Zhang, 2018, which concluded that 39% of total submicron organic material collected in a terrestrial environment was recovered and quantified by GCxGC methodology highly similar to that described in this work.[3]

Organic carbon quantification from analysis of the GCxGC was validated through comparison to organic carbon quantification from filters collected for carbon isotope analysis. These samples were analyzed through combustion and CO_2 production analysis, as described in Crocker et al., 2022. Organic carbon quantification from the filter samples collected for GCxGC analysis are relatively consistent with organic carbon quantification from the carbon isotope filters, although they are on average approximately 35% higher, as illustrated in Figure S3. There are two potential explanations; first, the samples were collected from sampling ports at two different locations. The GCxGC filters were collected closer to the breaking wave, which may have led to a higher concentration of submicron sea spray aerosol in this region. Second, the carbon isotope filters were pumped down to vacuum prior to analysis, which may have led to loss of semivolatile organics.

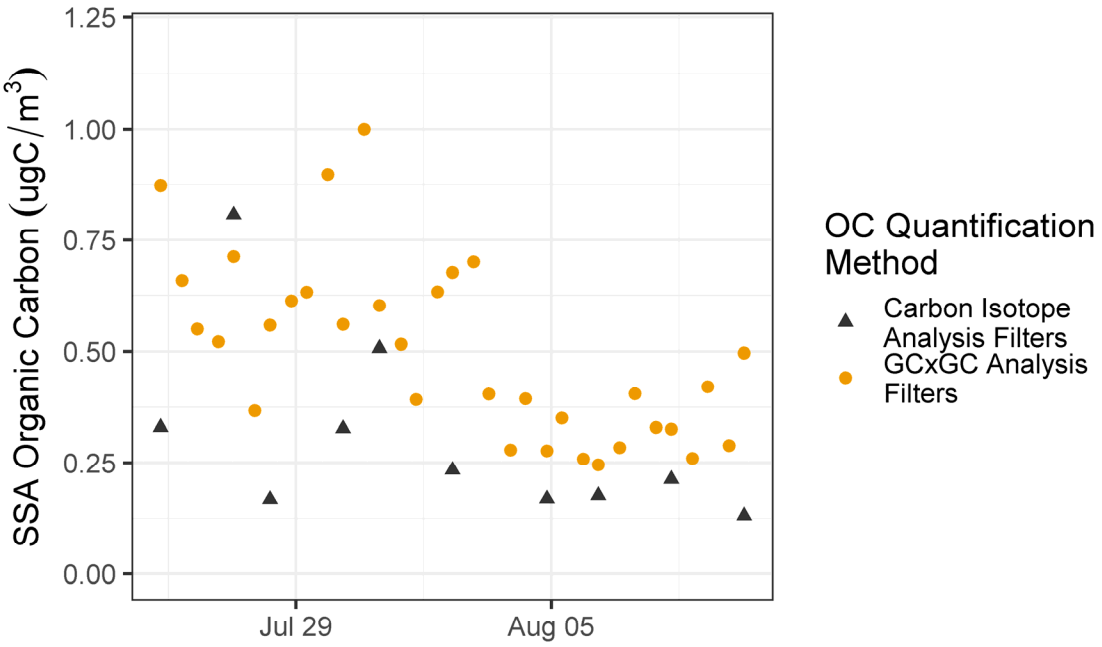


Figure 3.S3. Organic carbon concentration of sea spray aerosol collected during the SeaSCAPE mesocosm experiment bloom 3 as quantified by NIOSH-870 analysis of filters utilized for GCxGC analysis and combustion analysis of a separate filter sample set collected for analysis of carbon isotopes (as described in Crocker et al., 2022).

3.SI.3 Classification of Organic Compounds by Dynamic Time Warping Hierarchical Clustering

The normalized time series of each compound was smoothed by 3 point moving average. This step was necessary to emphasize lasting trends over diurnal variability in clustering, in service of the research question investigating the lasting dynamics governing transformations of the sea spray aerosol carbon pool. Next, each compound's smoothed time series was Z-scored, meaning each time point was subtracted by its own mean concentration over the campaign and divided by the standard deviation of smoothed time points. The top 100 compounds (by cumulative mass concentration over the entire bloom) were then separated from the rest and grouped by hierarchical clustering. Clustering was confined to these top abundance compounds for two reasons: first, to prioritize capturing the temporal dynamics of the compounds that dominate the submicron sea spray aerosol organic carbon, and second, because the other compounds are present at levels increasingly close to instrument detection limits, meaning their temporal variabilities are less precise due to decreasing signal to noise ratios. 100 was selected as the ideal split point, because the top 100 compounds accounted for ~90% of the mass and therefore achieved a balance of representing the majority of the mass distribution while avoiding potential biases due to signal to noise ratio issues with the least abundant compounds. Clustering was optimized at 7 clusters, as indicated by local minimization of the Modified Davies Bouldin Index (Figure S4).[7] Below 4 clusters, all

compounds exhibiting rising and falling temporal variability were grouped together regardless of peak timing, and as coordination between concentration increases and specific events was an area of specific interest analysis for optimal clustering was restricted to 5 or more clusters. The Z-scored temporal profiles of the constituents in each factor are visualized in Figure S5. Next, compounds detected with sufficient frequency to enable timeline construction (defined as compounds identified above detection limits in greater than 20% of samples) were assigned to the cluster with which they correlated most positively through construction of a Pearson correlation matrix between each of these compounds (time series smoothed and Z-scored identically to the cluster profiles) and the mean profile of each cluster. This process accounted for an additional 482 compounds. The remaining compounds were quantified and included in analysis but not assigned to clusters. Their presence is visualized by the "NA" designation in Figure 3 and Figure S7 and does not contribute significantly to total organic mass. Source group assignment is discussed in "Results and Discussion". The mean smoothed profiles of the 7 clusters in comparison to indicators of biological activity are illustrated in Figure 1.

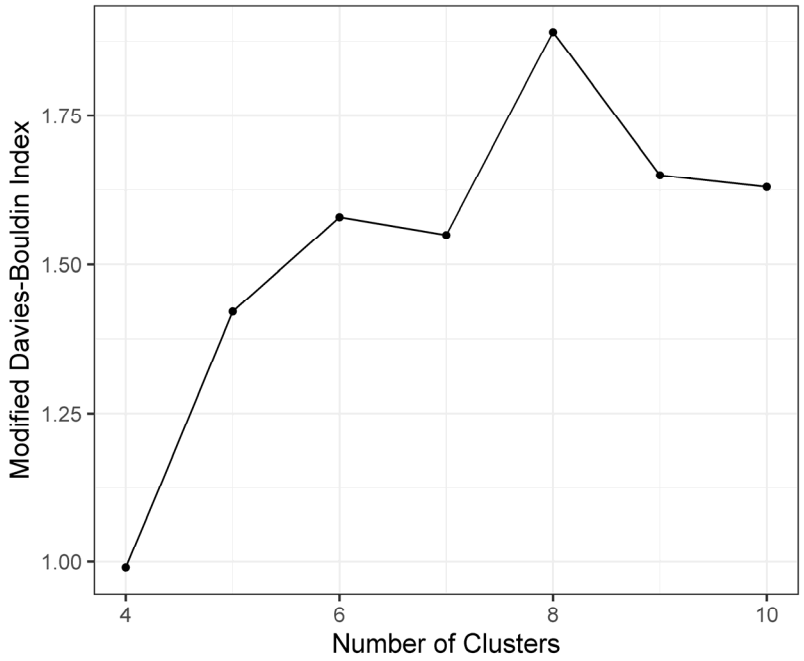


Figure 3.S4. Silhouette indices of hierarchical clustering of 100 most abundant compounds from SeaSCAPE experimental bloom by number of clusters selected.

Figure S5. Z-scored time series of sea spray aerosol organic cluster constituents from nascent sea spray aerosol samples collected during the SeaSCAPE mesocosm experiment. Clusters were generated by dynamic time warping hierarchical clustering under optimized 7 group solution.

3.SI.4: Temporal Variability of Species-Specific Biological Concentrations at SeaSCAPE, Additional Figures

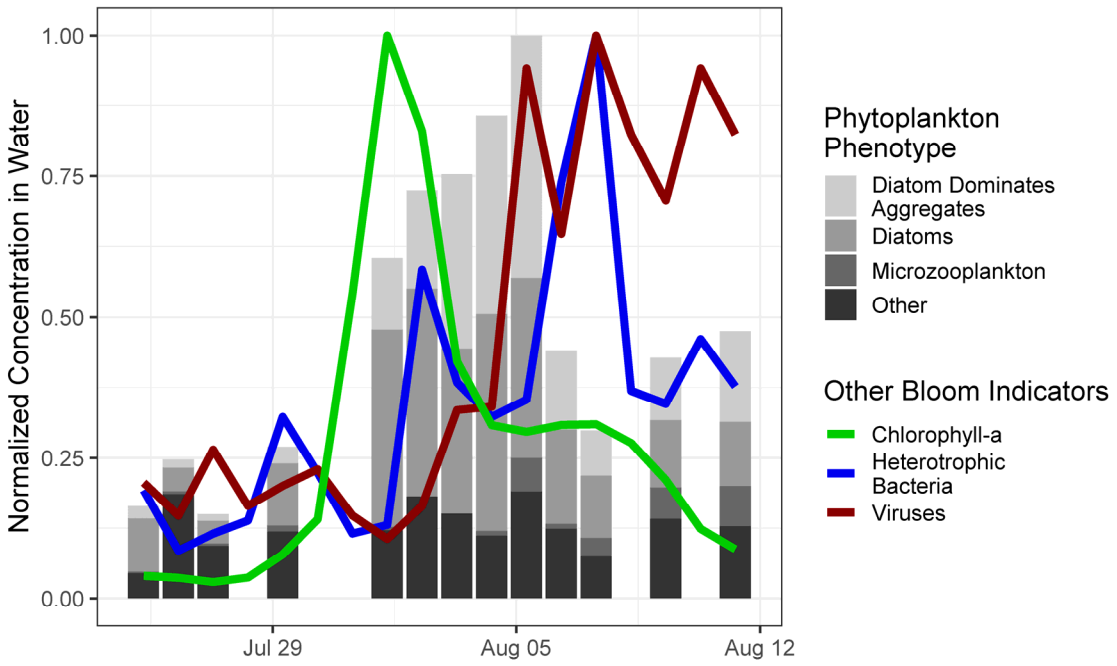


Figure 3.S6. Concentration in bulk water (normalized to time series maximum) of microbiology and biological activity indicators over the SeaSCAPE mesocosm experiment.

3.SI.5 Phytoplankton Enumeration

50mL's of bulk seawater were collected and poured into settlement chambers and settled for 24 hours prior to phytoplankton enumeration using the Utermöhl method under an Olympus IX-71 inverted microscope (Utermöhl, 1931). Cells/L for each phytoplankton species (genus level) were calculated and then binned into functional phytoplankton types (PFTs). Microzooplankton were also counted, and include tintinnids and other cilates to better identify grazing activities across the experiment .

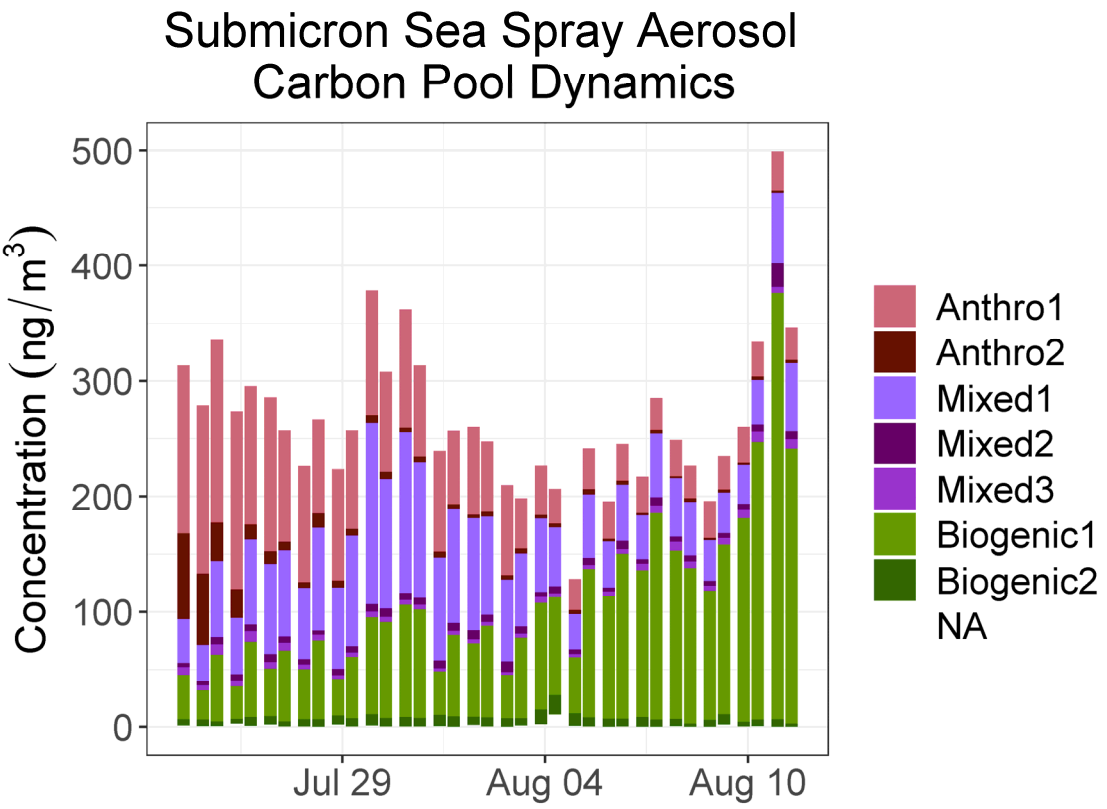


Figure 3.S7. Temporal variability of submicron sea spray aerosol organic mass concentration recovered by GCxGC from samples collected at the SeaSCAPE mesocosm campaign, separated by the temporal variability cluster to which each compound is assigned.

3.SI.6 Chemical Properties Modelling

Two classes of chemical information were predicted for the unidentifiable compounds at SeaSCAPE: average carbon oxidation state ($\overline{OS_c}$) and carbon number (n_c). All predictions were performed using Ch3MS-RF, which is described in Franklin et al., 2022. The external standard was randomly split (maintaining ratios of chemical functional group classes) into a training set consisting of 102 compounds and a test set of 24 compounds. An extrapolation set was curated from the 105 organic compounds detected in the SeaSCAPE aerosol samples that were identifiable by database match. Both mass spectrum and retention or kovats index matches were required for a compound to be included in the validation set. The chemical formulae of all of these compounds (both external standard and extrapolation set) are known and were parsed to create features of n_c and $\overline{OS_c}$, as described in Franklin et al., 2022. The random forest-based model was trained on the retention indices and featurized mass spectra (mass spectral featurization described in Franklin et al., 2022) of the external standard training set and tested on both the external standard test set and the extrapolation set of identifiable constituents in the SeaSCAPE aerosol samples. Performance for n_c prediction is summarized in Table S2 and Figure S8 panels A and B, while performance for $\overline{OS_c}$ prediction is summarized in

Table S3 and Figure S8 panels C and D. Prediction performance is strong and comparable to results described in Franklin et al., 2022, from which we conclude that property prediction accuracy is sufficient for useful visualization of all compound properties. Sample compounds with known identities were assigned their true n_c and $\overline{OS_c}$ values, while all other compounds were assigned their predicted properties as generated by the described models. The results of this property prediction in orienting sea spray aerosol compounds in $\overline{OS_c}$-n_c space are visualized in Figure 4.

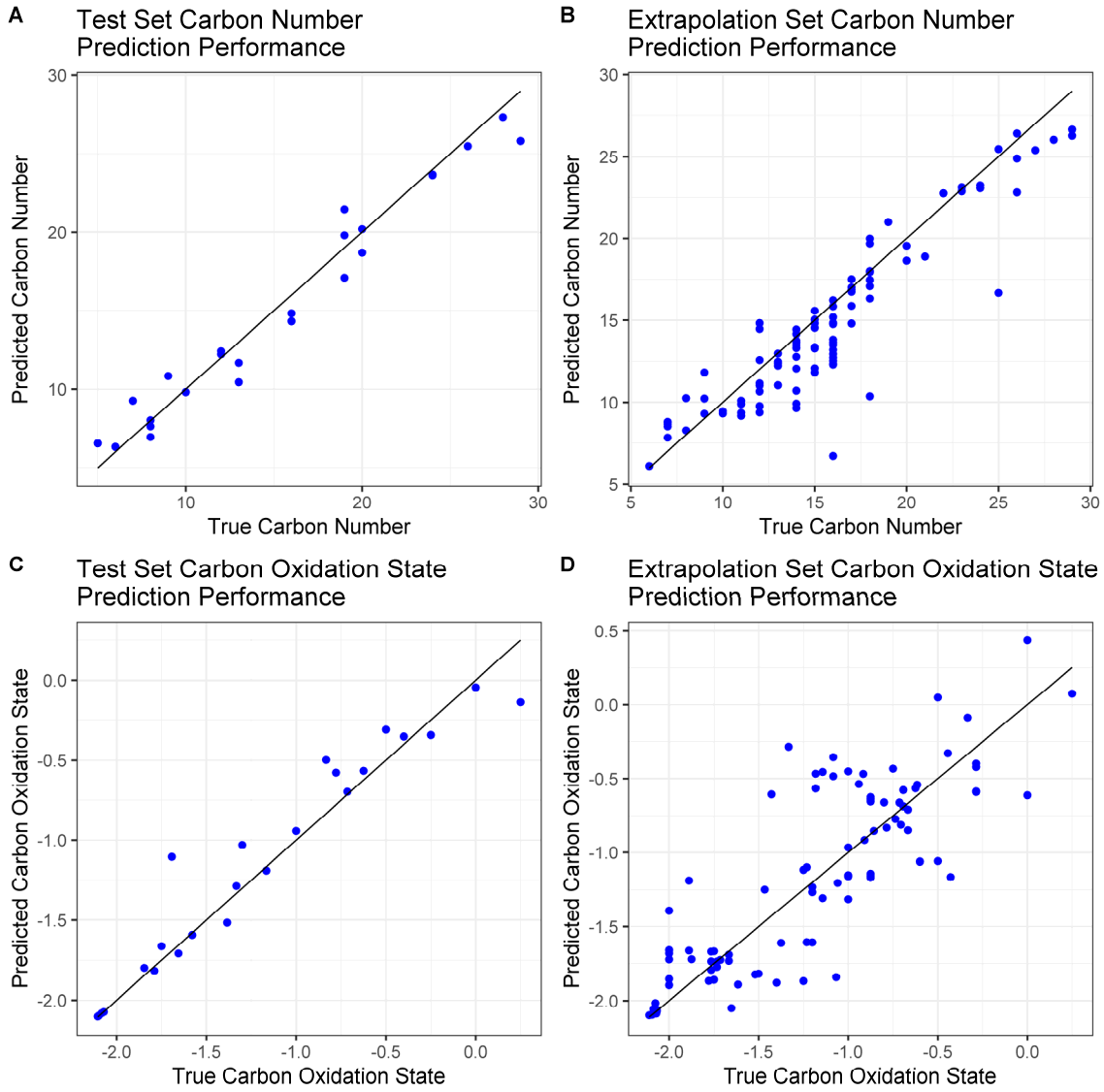


Figure 3.S8. Predicted versus true chemical properties for Ch3MS-RF-enabled chemical property prediction of the following: A) Carbon number, as tested on the external standard test set; B) Carbon number, as tested on the extrapolation set of identifiable organic constituents at SeaSCAPE; C) Average carbon oxidation state, as tested on the external standard test set; D) Average carbon oxidation state, as tested on the extrapolation set of identifiable organic constituents at SeaSCAPE

Data Set	Performance Metric	Performance
External Standard Test Set	OSR^2	.96
	RMSE	1.4
	MAE	1.1
	R^2	.96
SeaSCAPE Identifiable Organics Extrapolation Set	OSR^2	.80
	RMSE	2.3
	MAE	1.6
	R^2	.84

Table 3.S2. Performance metrics for carbon number prediction using Ch3MS-RF, described in Franklin et al., 2022[4]. OSR^2 refers to the out-of-sample r^2, MAE refers to the mean average error, RMSE refers to the root mean square error. All performance metric equations are defined in Franklin et al., 2022.

Data Set	Performance Metric	Performance
External Standard Test Set	OSR^2	.93
	RMSE	.18
	MAE	.11
	R^2	.93
SeaSCAPE Identifiable Organics Extrapolation Set	OSR^2	.69
	RMSE	.34
	MAE	.25
	R^2	.72

Table 3.S3. Performance metrics for average carbon oxidation state prediction using Ch3MS-RF, described in Franklin et al., 2022[4]. OSR^2 refers to the out-of-sample r^2, MAE refers to the mean average error, RMSE refers to the root mean square error. All performance metric equations are defined in Franklin et al., 2022.

3.SI.7 Sea Spray Aerosol PAH observations, supplemental table

Compound	Average Concentration over 1st day of analysis (ng/m^3)
Chrysene	2.0
Benz[a]anthracene, 7-methyl-	.52
Pyrene, 4-methyl-	.50
Phenanthrene, 2,3,5-trimethyl-	.49
Phenanthrene, 1,7-dimethyl-	.35
Anthracene, 1,4-dimethyl	.19
Phenanthrene, 2,3-dimethyl-	.17
Pyrene	.03

Table 3.S4. Speciation of polycyclic aromatic hydrocarbons (PAH's) identified in nascent sea spray aerosol during the SeaSCAPE mesocosm experiment, along with their mass concentrations in sampled air averaged over the first full day of analysis.

3.SI.8 Unidentifiable organic compounds, supplemental figures

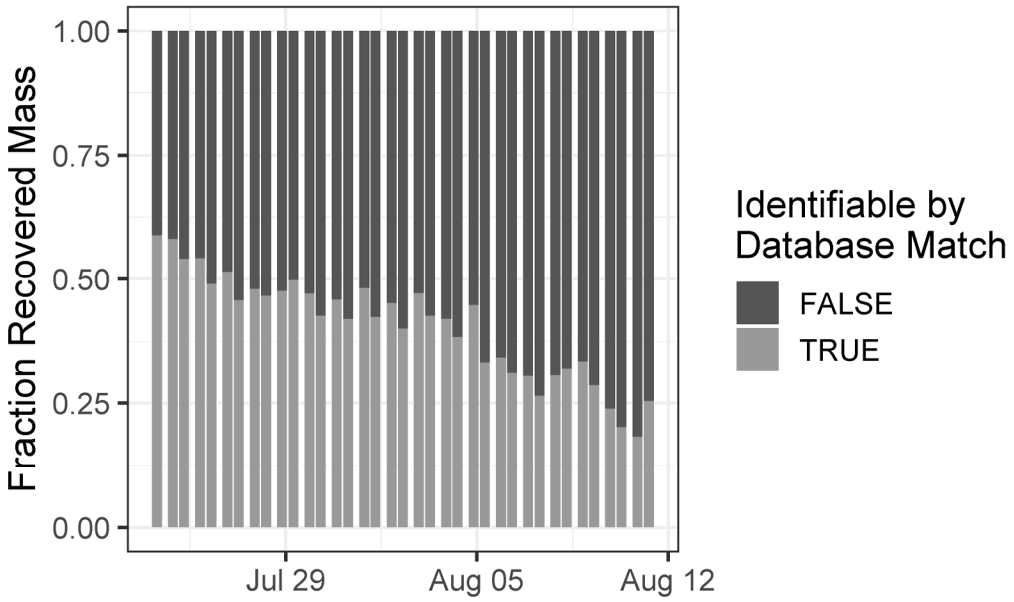


Figure 3.S8. Fraction of total submicron sea spray aerosol organic mass attributable to identifiable compounds over time throughout the SeaSCAPE bloom.

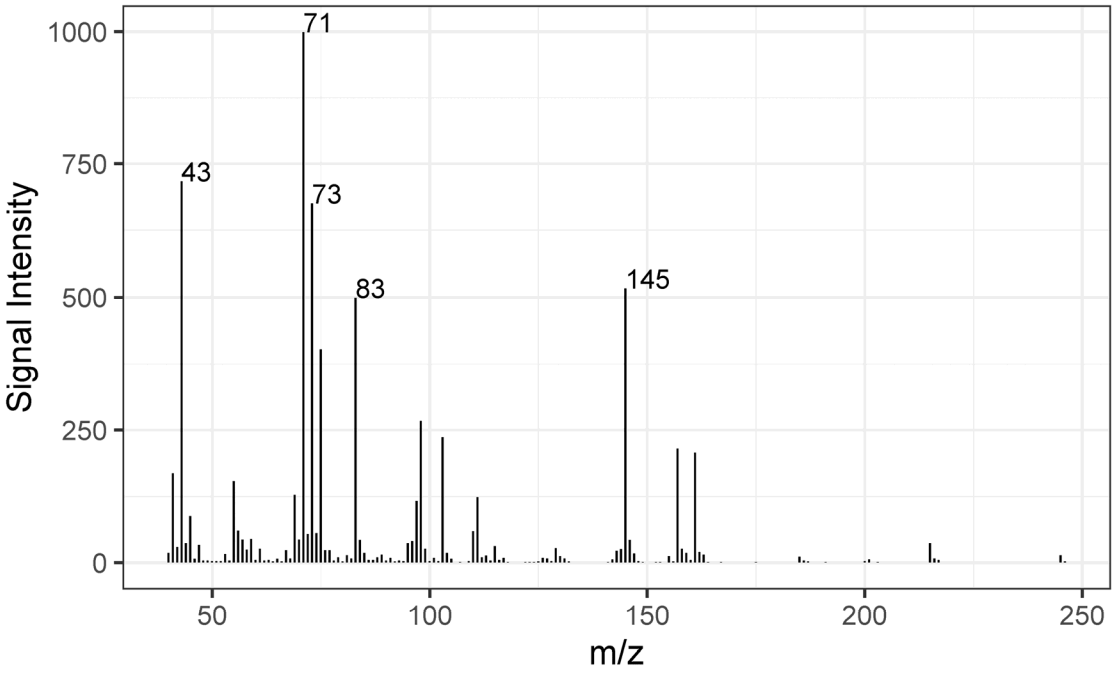


Figure 3.S9. 70 eV EI mass spectrum of unidentifiable biogenic organic compound A, observed in SeaSCAPE experimental bloom. This compound was observed at an equivalent kovats index (n-alkane equivalent) of 1417 using a semistandard nonpolar column. Predicted properties: n_c = 11, $\overline{OS_c}$ = -1.2

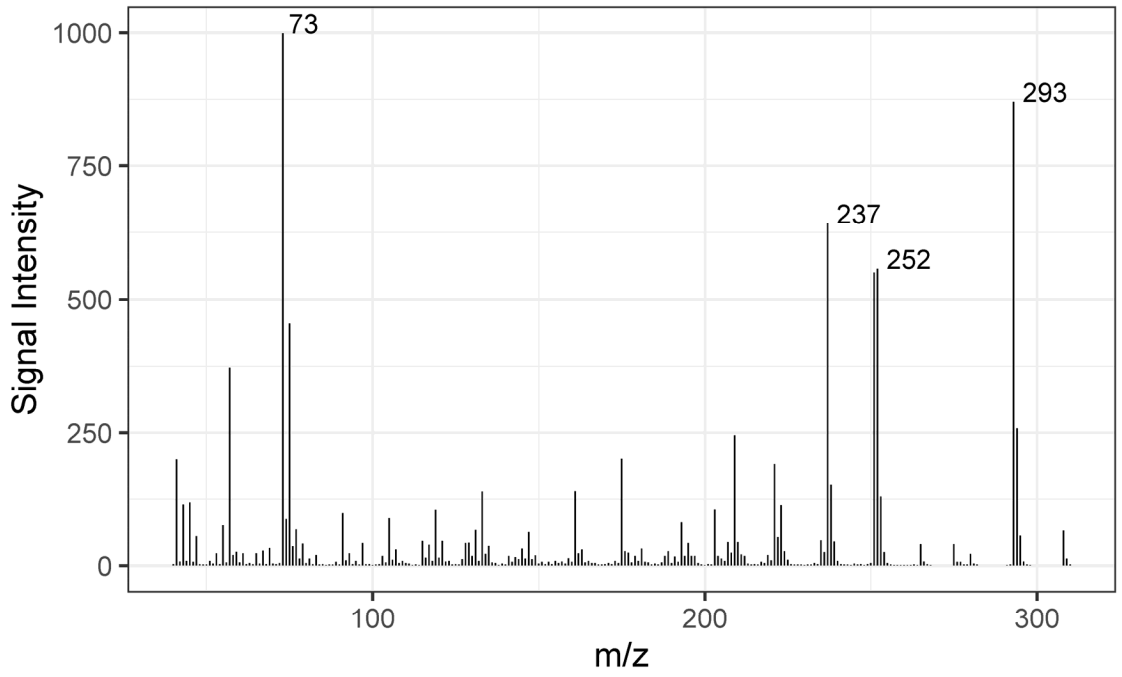


Figure 3.S10. 70 eV EI mass spectrum of unidentifiable biogenic organic compound B, observed in SeaSCAPE experimental bloom. This compound was observed at an equivalent kovats index (n-alkane equivalent) of 1503 using a semistandard nonpolar column. Predicted properties: n_c = 11, $\overline{OS_c}$ = -.93

Supporting Information References

(1) Franklin, E. B.; Alves, M. R.; Moore, A. N.; Kilgour, D. B.; Novak, G. A.; Mayer, K.; Sauer, J. S.; Weber, R. J.; Dang, D.; Winter, M.; Lee, C.; Cappa, C. D.; Bertram, T. H.; Prather, K. A.; Grassian, V. H.; Goldstein, A. H. Atmospheric Benzothiazoles in a Coastal Marine Environment. *Environ. Sci. Technol.* **2021**, *55* (23), 15705–15714. https://doi.org/10.1021/ACS.EST.1C04422/SUPPL_FILE/ES1C04422_SI_001.PDF.

(2) Jen, C. N.; Hatch, L. E.; Selimovic, V.; Yokelson, R. J.; Weber, R.; Fernandez, A. E.; Kreisberg, N. M.; Barsanti, K. C.; Goldstein, A. H. Speciated and Total Emission Factors of Particulate Organics from Burning Western US Wildland Fuels and Their Dependence on Combustion Efficiency. *Atmos. Chem. Phys.* **2019**, *19* (2), 1013–1026. https://doi.org/10.5194/acp-19-1013-2019.

(3) Zhang, H.; Yee, L. D.; Lee, B. H.; Curtis, M. P.; Worton, D. R.; Isaacman-VanWertz, G.; Offenberg, J. H.; Lewandowski, M.; Kleindienst, T. E.; Beaver, M. R.; Holder, A. L.; Lonneman, W. A.; Docherty, K. S.; Jaoui, M.; Pye, H. O. T.; Hu, W.; Day, D. A.; Campuzano-Jost, P.; Jimenez, J. L.; Guo, H.; Weber, R. J.; De Gouw, J.; Koss, A. R.; Edgerton, E. S.; Brune, W.; Mohr, C.; Lopez-Hilfiker, F. D.; Lutz, A.; Kreisberg, N. M.; Spielman, S. R.; Hering, S. V.; Wilson, K. R.; Thornton, J. A.; Goldstein, A. H. Monoterpenes Are the Largest Source of Summertime Organic Aerosol in the Southeastern United States. *Proc. Natl. Acad.*

Sci. U. S. A. **2018**, *115* (9), 2038–2043. https://doi.org/10.1073/PNAS.1717513115/SUPPL_FILE/PNAS.1717513115.SD01.RTF.

(4) Franklin, E. B.; Yee, L. D.; Aumont, B.; Weber, R. J.; Grigas, P.; Goldstein, A. Ch3MS-RF: A Random Forest Model for Chemical Characterization and Improved Quantification of Unidentified Atmospheric Organics Detected by Chromatography-Mass Spectrometry Techniques. *Atmos. Meas. Tech- Discuss.* **2022**. https://doi.org/10.5194/amt-2022-99.

(5) Wu, C.; Hilda Huang, X. H.; Man Ng, W.; Griffith, S. M.; Zhen Yu, J. Inter-Comparison of NIOSH and IMPROVE Protocols for OC and EC Determination: Implications for Inter-Protocol Data Conversion. *Atmos. Meas. Tech.* **2016**, *9* (9), 4547–4560. https://doi.org/10.5194/AMT-9-4547-2016.

(6) Russell, L. M. Aerosol Organic-Mass-to-Organic-Carbon Ratio Measurements. *Environ. Sci. Technol.* **2003**, *37* (13), 2982–2987. https://doi.org/10.1021/ES026123W.

(7) Kim, M.; Ramakrishna, R. S. New Indices for Cluster Validity Assessment. *Pattern Recognit. Lett.* **2005**, *26* (15), 2353–2363. https://doi.org/10.1016/J.PATREC.2005.04.007.

4 Atmospheric Benzothiazoles in a Coastal Marine Environment

This work is adapted from:
Emily B. Franklin, Michael R. Alves, Alexia N. Moore, Delaney B. Kilgour, Gordon A. Novak, Kathryn Mayer, Jon Sauer, Robert J. Weber, Duyen Dang, Margaux Winter, Christopher Lee, Christopher D. Cappa, Timothy H. Bertram, Kimberly A. Prather, Vicki H. Grassian, Allen H. Goldstein, "Atmospheric Benzothiazoles in a Coastal Marine Environment" *Environmental Science &Technology* (**2021**)

4.1 Abstract

Organic emissions from coastal waters play an important but poorly understood role in atmospheric chemistry in coastal regions. A mesocosm experiment focusing on facilitated biological blooms in coastal seawater, SeaSCAPE (Sea Spray Chemistry and Particle Evolution), was performed to study emission of volatile gases, primary sea spray aerosol, and formation of secondary marine aerosol as a function of ocean biological and chemical processes. Here we report observations of aerosol-phase benzothiazoles in a marine atmospheric context with complementary measurements of dissolved-phase benzothiazoles. Though previously reported dissolved in polluted coastal waters, we report the first direct evidence of the transfer of these molecules from seawater into the atmosphere. We also report the first gas-phase observations of benzothiazole in the environment absent a direct industrial, urban, or rubber-based source. From the identities and temporal dynamics of the dissolved and aerosol species, we conclude that the presence of benzothiazoles in the coastal water (and thereby their emissions into the atmosphere) is primarily attributable to anthropogenic sources. Oxidation experiments to explore the atmospheric fate of gas-phase benzothiazole show that it produces secondary aerosol and gas-phase SO_2, making it a potential contributor to secondary marine aerosol formation in coastal regions and participant in atmospheric sulfur chemistry.

4.2 Introduction

Coastal oceans are often enriched with organic species from both biogenic and anthropogenic sources. Biogenic sources include phytoplankton, which convert CO_2 to ocean biomass that is then transformed by the microbial loop.[1,2] Direct anthropogenic sources of marine organics in seawater include wastewater discharge and urban runoff (often enriched with personal care products), trash, and shipping pollution.[3,4,5] Phytoplankton blooms occur naturally, but also can be induced by anthropogenic discharges of fertilizer enriched runoff, and can be enhanced by climate change induced perturbations to ocean temperatures and chemistry.[6,7,8,9] Ocean-derived organic species can be transferred from the ocean to the atmosphere through two major mechanisms, both of which can influence

atmospheric chemistry over the ocean. The first of these produces sea spray aerosol (SSA) from bubble bursting and wind shear at the surface.[10] The second method involves the emission of volatile organic compounds (VOC)s from the ocean surface, either through volatilization or interfacial reactions producing volatile products.[11,12,13]

The composition and properties of organic material emitted from the ocean are important areas of atmospheric research, as they influence marine atmospheric chemistry and public health. Salt and organics have different cloud and ice nucleation properties, meaning that organic enrichment of SSA is important for climate.[14,15,16,17] Marine VOCs oxidize in the atmosphere, where they can form secondary marine aerosols (SMA).[18] Recent laboratory and field studies have suggested that SMA may play a significant role in cloud formation over the ocean,[19,20] making the emission and oxidation processes of marine volatile organics a critical area of atmospheric chemistry research. Organic enriched marine aerosols are also a growing area of concern in public health, as they can expose coastal communities to marine toxins and pollutants, including biogenic toxins from harmful algal blooms, pesticides, and phthalates.[21,22,23]

Production of plasticizers, pharmaceuticals, pesticides, and more, over the course of the last century, has led to a portion of dissolved organic matter (DOM) in marine ecosystems being classified as anthropogenic dissolved organic carbon (ADOC).[24] The 100,000+ commercialized synthetic compounds produced by humans have been continually increasing in both number and concentration in many environments.[25,26] These human-produced compounds tend to be hydrophobic, and for the most part, have completely unknown breakdown by-products, toxicities, and fates in the environment.[27,28] Though there is no study to our knowledge that has reported total anthropogenic DOM concentrations in coastal waters near populated areas, it is expected that this number will be greater than the < 1 uM, or 0.05 to 1% of total DOM, reported for water collected from the open ocean due to the more concentrated coastal inputs and sources.[24,29] The hydrophobic fraction of these species will be highly concentrated at the surface of the ocean, joining biogenically formed molecules in a 1 to 100 μm thick region known as the sea surface microlayer (SSML).[30] Many studies have shown that the ocean surface is a source of organic material for both SSA and SMA, significantly contributing to the overall chemistry and properties of these aerosols.[31,32,15,33,13,34]

Benzothiazoles are a class of anthropogenic pollutants that have been previously reported in a variety of freshwater and coastal aquatic environments, including the southern California coast,[35,36,37,38,39,40] but they are also produced biogenically from select marine microbiological species.[41,42] Benzothiazole (hereafter BT) (C_7H_5NS) is an aromatic heterocyclic organic compound containing both sulfur and nitrogen. Benzothiazoles are here defined as compounds containing a benzothiazole moiety. Benzothiazoles are high production volume chemicals used in a wide range of industrial and consumer products, with significant sources including rubber production, leather and paper production, antifreeze, herbicides, textiles, and plastics.[35] They are found in both urban runoff and

wastewater, with urban runoff frequently the more concentrated of the two source groups.[43] A review of the ranges of measured concentrations of dissolved BT and 2-(Methylthio)benzothiazole in stormwater runoff and wastewater effluent at sites along the southern California coast (ranging from 0.05 to 0.5 μg/L for BT and 0.04 to 0.3 μg/L for 2-(Methylthio)benzothiazole) can be found in Zeng et al., 2004.[40] Benzothiazoles have been observed in the dissolved-phase in a variety of coastal marine settings and used as tracers of wastewater discharge.[38,40,37,36] This compound class is a growing area of concern in both marine ecosystem and public health; benzothiazole derivatives from rubber leachates are toxins hazardous to marine microbiology, and various benzothiazoles are human dermal sensitizers, endocrine disruptors, carcinogens, and genotoxins.[35,44,45,46]

In this study, we investigate the transfer of benzothiazoles from coastal water into the atmosphere through a controlled mesocosm study. While dissolved benzothiazoles have been reported in the water in a variety of coastal regions, we report observations of benzothiazoles in sea spray aerosols and emitted VOCs, showing for the first time that polluted coastal oceans likely emit these chemicals to the atmosphere. Furthermore, we explore the atmospheric oxidation and aerosol formation potential of gas-phase BT, which has implications for air quality in both polluted coastal regions and urban or industrial centers in which benzothiazoles are most concentrated (Figure 4.S1).

4.3 Materials and Methods

Experimental Campaign, Sample Collection, and Online Analysis

The results from this study derive primarily from Sea Spray Chemistry And Particle Evolution (SeaSCAPE) 2019, a collaborative mesocosm experiment described in Sauer, Mayer, and Lee et al., 2021.[47] Coastal water (12,000 L) from Ellen Browning Scripps Memorial Pier (hereafter Scripps Pier) in La Jolla, California was transported into an indoor wave chamber facility (hereafter referred to as the wave channel), in which mechanically generated waves break on an artificial beach to create realistic primary sea spray aerosols. The natural coastal water was amended with nutrients to initiate an algal bloom of the naturally existing phytoplankton species, including three bloom experiments, replicating methodology described in Wang et al., 2015.[48] Water from the wave channel was diverted through an inert glass and Teflon chamber forming an isolated sampling vessel (ISV), allowing analysis of gas-phase marine emissions. This analysis focuses on the third algal bloom experiment, which lasted 20 days from July 24th to August 12th, 2019 and is described in detail in Sauer, Mayer, and Lee et al., 2021.[47]

Periodically over the course of the bloom, water from the SeaSCAPE channel was collected for offline analysis of headspace gasses (described below) and of extracted DOM chemical composition (also described below). One water sample directly from Scripps Pier was collected on July 23rd for DOM extraction and analysis of dissolved analytes under ambient conditions. Submicron aerosol samples from the channel were collected on quartz fiber filters (Pallflex Tissuequartz) using a custom designed automated sequential sampler. Additional sample collection and storage details can be found in Supporting Information

section 3 (4.SI.3). ISV VOCs were collected on custom triple bed sorbent tubes (Camsco). 2 L samples were collected every 2-3 days during the early stages of the bloom and 1-2 times per day during peak biological activity. Sorbent tube material and sampling details can be found in 4.SI.1 and 4.SI.4. VOCs both within the ISV and in the channel were additionally measured by Vocus proton-transfer-reaction time-of-flight mass spectrometer (Vocus PTR-TOF by Aerodyne/Tofwerk).[47] Following the conclusion of SeaSCAPE, the Vocus was relocated to the end of the Scripps Pier for ambient coastal VOC sampling. A full description of Scripps Pier VOC sampling can be found in 4.SI.5.

Collection and extraction of marine DOM was performed using solid-phase extraction (SPE) by PPL resin (Bond Elut, Agilent), following methods previously characterized in Dittmar et al., 2008 and described in detail in 4.SI.6.[49]

Offline Analysis

A high resolution Orbitrap spectrometer equipped with a modified Atmospheric Pressure Chemical Ionization source (APCI-Orbitrap, ThermoFisher) was used to detect VOCs in the headspace of collected water during the SeaSCAPE campaign in a method adapted from Roveretto and coworkers in 2019.[50] Operational details can be found in 4.SI.7.

Aerosol filters, DOM, and VOC sorbent tubes were all analyzed by thermal desorption two-dimensional gas chromatography coupled with electron ionization time of flight mass spectrometry (TD-GCxGC-ToF-MS) on two separate instruments covering differing volatility ranges. DOM and aerosol samples were analyzed on GCxGC A, while VOC sorbent tubes were analyzed on GCxGC B (details in 4.SI.2). DOM and aerosol samples were normalized by internal standard and derivatized to enhance recovery of polar organics. TD-GCxGC methodology is described in detail in Worton et al. 2017, and details specific to this analysis can be found in 4.SI.9.[51]

From SeaSCAPE samples, 754 unique aerosol organics and 991 unique DOM organics were compiled into libraries of mass spectra and retention indices (internal standard normalized position in the volatility dimension) were catalogued. Of these unique organic species, 12 SSA organics and 6 DOM organics were identified as benzothiazoles based on matches and similarities to authentic standards and species catalogues in the NIST-14 mass spectral database according to 4 classes of identification certainty (see Table 3.1). Each observed benzothiazole was assigned to a benzothiazole external standard chemical proxy for quantification based on exact match or chemical similarity and proximity in GCxGC space. Proxy assignments and additional details can be found in 4.SI.11.

Benzothiazole Oxidation Study

To complement the mesocosm study and better understand the atmospheric fate of gas-phase BT, we used a Potential Aerosol Mass Oxidation Flow Reactor (PAM-OFR) in a separate laboratory oxidation study to produce BT oxidation products. PAM-OFR operation and calibration details can be found in 4.SI.16. BT was introduced into the PAM-OFR through two methods; in the first, liquid BT dissolved in a methanol carrier was introduced in a plug injection experiment at a constant exposure setting. In the second, a

BT permeation tube continuously introduced gas-phase BT at different OH concentrations (2.9, 3.5, and 4.9 days equivalent ageing). Particle formation and size distributions, SO_2 production, and HSO_4^- production from the BT oxidation experiments were all monitored (4.SI.16).

Reported herein we show that BT and benzothiazole-moiety compounds are not just present in sea spray aerosols but also show an ability to form secondary aerosols using an oxidative flow reactor to simulate photochemical aging.

4.4 Results and Discussion

Gas-Phase Benzothiazole Observations

In the early stages of the SeaSCAPE experiment, a significant and unexpected peak was observed to correspond to $C_7H_5NSH^+$ (BT) by APCI-Orbitrap headspace VOC analysis. Gas-phase BT and its isomeric identity was confirmed using the Orbitrap's tandem mass spectrometry capabilities and the clear observation of the benzene and cyclic ring (containing the nitrogen and sulfur) fragments. This peak's identity was later confirmed and quantified by offline TD-GCxGC-ToF-MS analysis of sorbent tube samples collected from the SeaSCAPE ISV (14 eV and 70 eV EI spectra illustrated in SI Figure 4.S2). Time resolved measurements of the ion corresponding to gas-phase benzothiazole were obtained by the Vocus throughout SeaSCAPE. A quantitative comparison of GCxGC and Vocus benzothiazole and summed monoterpene ISV measurements is in SI (Figure 4.S4); while absolute quantities measured by each instrument differ, variability over the SeaSCAPE bloom and relative abundances of significant observed gases agree, with R^2 of 0.91 and 0.98 for benzothiazole and summed monoterpenes respectively.

Gas-phase benzothiazole has been reported in atmospheric measurements in the contexts of freshly shredded rubber,[35] urban traffic, and coal burning in a region with plastic production,[52] but emissions of gas-phase benzothiazole from coastal waters are, to our knowledge, a novel finding. As benzothiazole is a common anthropogenic contaminant that could potentially be added if the water were mishandled, we confirmed the presence of benzothiazoles in the collected coastal water absent perturbations from transport into the SeaSCAPE wave channel. Benzothiazole was consistently found in all monitored phases (dissolved, aerosol, and gas) (Table 4.1; Figure 4.1), and dissolved-phase concentrations did not increase between the first sample (collected directly at Scripps Pier) and the next DOM sample, collected from the channel two days following transport. Additionally, the ion corresponding to gas-phase BT was observed by the Vocus PTR-MS at the end of Scripps Pier directly following the SeaSCAPE experimental campaign, confirming that gas-phase BT is present in the local coastal marine atmosphere. End-of-pier BT concentrations averaged 2.5 ppt (compared to a detection limit of 0.5 ppt for a 10 s averaging time, averaged over the entire ~1 month measurement period), and SeaSCAPE ISV concentrations (as quantified by GCxGC) averaged 53 ppt. Given the elevated water temperature, enclosed conditions, and low gas-phase flow rate of the SeaSCAPE ISV (described in Sauer, Mayer, and Lee et al., 2021), absolute gas-phase ISV concentrations

exceeding ambient concentrations at the end of the pier is indicative of emissions to the gas-phase.[47]

Previously reported Henry's law constants for BT ($K_{H'}$, estimated from vapor pressure and solubility in Reddy et al 1997, predicted from structure through various methods as reported in Sander 2015, and predicted from structure based on HENRYWIN v. 3.10, EPA EPI Suite) range from 1.5×10^{-5} to 2.6×10^{-4} (dimensionless gas over liquid, converted assuming 25 °C).[53,54] Under the aqueous concentration (0.29 µg/L at day 1 sampling time, as shown in Table 1) and temperature conditions observed at SeaSCAPE, this would imply a maximum equilibrium gas concentration of BT between 0.7 and 12 ppt, a factor of 75-4.4 below the average concentrations observed in the ISV (53 ppt). A simplified partitioning experiment using benzothiazole and reference species dissolved in simplified simulated sea water in concentrations mirroring those identified in the SeaSCAPE DOM (described in SI.10) did not detect benzothiazole above a detection limit of 10 ppt, supporting the conclusion that observed gas-phase concentrations of benzothiazole in the SeaSCAPE ISV cannot be adequately explained by ideal air water partitioning governed by Henry's law. This implies that some ocean-atmosphere transfer mechanism other than idealized aqueous partitioning plays an important role in gas-phase BT emissions. It should also be noted that the lowest reported Henry's law constant of BT, that reported by HENRYWIN v. 3.10, EPA EPI Suite, is the most commonly referenced value across popular chemical information repositories (such as PubChem and ChemSpider) and has been used to assume negligible volatilization of benzothiazole from organic rich aquatic systems (wastewater), rendering this finding of elevated benzothiazole emissions from real organic-rich aquatic systems particularly important.[55]

Aerosol and DOM Benzothiazole Observations and Comparisons

In addition to BT, a suite of larger compounds containing a benzothiazole moiety were observed in both the aerosol and dissolved phases (Table 1). BT accounts for 50% of the total mean benzothiazole carbon pool in the aerosol phase and is the second most significant contributor to dissolved benzothiazoles at 43% of the mean DOM benzothiazole carbon pool. DOM BT concentrations measured directly at the pier (0.29 ± 0.15 ug/L), reported in Table 1, fall within reported ranges of observed BT in freshwater aquatic environments along the southern California coast, which range from 0.05 to 0.5 µg/L.[40] 2-(Methylthio)benzothiazole, a commonly reported tracer species for benzothiazoles originating from waste water or runoff,[40,43] is the most abundant benzothiazole species observed in the dissolved phase at 47% of mean DOM benzothiazole carbon pool. Butyl-benzothiazole, a species not previously reported in environmental measurements and present at near detection limit levels in the dissolved phase, contributes the second highest fraction of the aerosol benzothiazole carbon pool (37%), while 2-(Methylthio)benzothiazole contributes the third highest fraction at 6%.

The time series of gas and aerosol-phase benzothiazoles provide valuable insights into likely sources and transfer processes governing marine benzothiazole emissions (Figure

4.1). Prior to August 5th, a key biological transition period between peak chlorophyll and peak heterotrophic bacteria during which many key biogenic gases peak (as illustrated in SI Figure 4.S3), gas and aerosol-phase BT are highly correlated ($R^2 = 0.78$). In both phases, BT begins relatively low before increasing (by a factor of 5 in the gas phase and a factor of 3 in the aerosol phase) to a peak in the afternoon through night of July 30th, after which it declines, back to initial concentrations for the aerosol phase and to approximately double initial concentrations in the gas phase. There is no clear temperature, biological, or perturbation-based explanation for the July 30th peak, but in the gas phase it is observed for BT across all instruments and sampling strategies (GCxGC, Vocus PTR-ToF, and APCI-Orbitrap) and is common to an array of anthropogenic gases, as illustrated in SI Figure 4.S3. After August 5th however, the BT traces of the two phases are loosely anticorrelated with an R^2 of 0.11. As previously noted, the composition and thickness of the sea surface microlayer has been observed, modeled, and demonstrated to play an important role in the transfer of marine gases.[11,34] Changes in SSML composition and characteristics during the second half of the bloom, described in Crocker et al., 2021, could therefore at least partially explain the lack of correlation in this period.[56]

Although there are some similarities between the temporal variability of benzothiazoles and biological indicators (specifically chlorophyll-a concentrations in the bulk water), the time series of both gas and aerosol-phase benzothiazoles are much more similar to anthropogenic species than to any known biogenic products. In the aerosol phase, the total benzothiazole carbon pool time series is strongly correlated across the entire bloom ($R^2 = 0.85$, see SI Figure 4.S8) with that of tetradecamethyl-cycloheptasiloxane (more commonly known as D7), an anthropogenic species attributable to personal care products,[57] wastewater,[58] and sewage, and also observed to bioaccumulate in marine ecosystems.[59] In the gas phase, BT also displays similar temporal dynamics to anthropogenic species (SI Figure 4.S3), specifically benzophenone (a common sunscreen and personal care product component)[60] and naphthalene (a polycyclic aromatic hydrocarbon (PAH), previously identified as a coastal marine contaminant off the Southern California coast and likely originating from some combination of petroleum and combustion sources).[61] While the correlations between BT and benzophenone and naphthalene are not strong ($R^2 = 0.25$ and 0.35 respectively), they are far stronger than the correlations with any of the known biogenic gases; the R^2 of the correlations between BT and isoprene, DMS, and beta-cyclocitral are all below 0.005. While there may be some degree of biogenic contribution to observed atmospheric benzothiazoles, there is no compelling evidence for such a source in the temporal variability. A full discussion of potential sources and justification for a conclusion of a dominant anthropogenic origin based on additional chemical indicators can be found in the following section.

One observed aerosol-phase species deserving particular attention is butyl-benzothiazole. Although butyl-benzothiazole sulphenamides are broadly characterized rubber vulcanization agents[62] and butyl-benzothiazole has been synthesized in laboratory

environments, it has never to our knowledge been reported in the environment. During the bloom, butyl-benzothiazole decreases both in terms of absolute mass concentration and as a fraction of the total benzothiazole carbon pool and drives the majority of the decrease in the total aerosol-phase benzothiazole carbon pool. To our knowledge there are no studies investigating the biodegradation of butyl-benzothiazole in marine settings, but this finding in conjunction with the significant observed aerosol levels indicates that butyl-benzothiazole biodegradation may be an important environmental process. No identified benzothiazoles are observed to increase significantly as the aerosol benzothiazole carbon pool shrinks (see SI Figure 4.S7). There are several probable contributions to this phenomenon. The biodegradation of butyl-benzothiazole has not been studied in marine contexts and its products may not have published mass spectra allowing them to be confirmed as benzothiazole biodegradation products. Additionally, benzothiazole degradation products with published mass spectra may fall outside the TD-GCxGC's sensitivity range, as biological degradation often produces highly oxygenated species not amenable to GC analysis, as described in Nowak et al., 2018.[63] Benzothiazoles may also bioaccumulate in the biological species within the channel, partition into suspended organic matter or onto the organic film on the wave channel surfaces, or decrease as a relative fraction of submicron organics emitted from bubble bursting processes due to changes in the structure and composition of the SSML.

While the most abundant benzothiazole species are common to both dissolved and aerosol phases, unique isomers are observed in each and the relative distribution of common species differ (Figure 4.2), leading to a significant discrepancy between the solubility distributions of the benzothiazole carbon pools in the bulk and aerosol phases; bulk-phase benzothiazoles are more water-soluble, and less diverse in low solubility species, compared to aerosol-phase benzothiazoles (Figure 4.3). Of the 5 species observed in both phases, the low solubility (<15 mg/L) species are significantly enhanced as a fraction of the total benzothiazole carbon pool in SSA (Figure 4.S10). Solubilities in water are estimated from log K_{ow} (WSKOW v. 1.41, EPA EPI Suite) with all alkyl benzothiazoles parameterized as equal in solubility to butyl-benzothiazole. The solubility distribution discrepancy is likely due to the concentration of low-solubility organics in the SSML[64] and previously reported film-jet sea spray aerosol formation dynamics, which cause SSML organics to preferentially aerosolize into smaller aerosol particles from film drops.[65,66,67] The different benzothiazoles' implications for climate and public health relevant properties of sea spray aerosol lie beyond the scope of this publication. However, the differing chemical distributions between dissolved and aerosol phases, in particular the abundance of low solubility benzothiazoles not observed in the bulk water, highlight that measurements of bulk-phase organics in ocean waters are imperfect indicators of which organic species enrich the submicron aerosol particles. This size population has the longest atmospheric lifetime, dominates the marine aerosol surface area distribution, and has the potential to be transported farthest inland. This finding therefore has implications for

marine pollutant human exposure in coastal regions; marine pollutant monitoring focuses nearly exclusively on the bulk water, but as this study demonstrates, bulk water and aerosol-phase organic distributions and concentrations differ significantly, meaning that bulk water toxin measurements may not accurately reflect the sea spray aerosol concentrations of hazardous marine pollutants including carcinogens and respiratory irritants not reported in this work.

Evidence for Anthropogenic Origin

While benzothiazoles are high-volume industrial chemicals and are commonly studied as wastewater contaminants in coastal areas, there are also marine biogenic sources of benzothiazoles.[41,42] Coupled with the early bloom increase in gas and aerosol-phase benzothiazole levels that cannot be explained by (do not correlate with) either the rising temperatures or indicators of biological activity (including chlorophyll-a concentrations, heterotrophic bacteria concentrations, and peaks in typical biogenic gases such as DMS and isoprene) (Figure 4.1), this necessitates a more nuanced investigation of the most probable origins of the benzothiazoles observed at the SeaSCAPE campaign and in ambient air at Scripps Pier. Prominent among benzothiazoles in both dissolved and aerosol phases and observed in particularly high abundance in the dissolved phase was 2-(Methylthio)benzothiazole, a known and commonly reported tracer of anthropogenic benzothiazoles originating from runoff or wastewater.[40,68,35] Compared to other anthropogenic benzothiazole tracers, 2-(Methylthio)benzothiazole is relatively resistant to both photochemical and biological degradation, and is itself a biodegradation product of another commonly reported anthropogenic benzothiazole, 2-Mercaptobenzothiazole.[69] While multiple naturally occurring benzothiazoles have been characterized and reported, notably including several originating from the marine bacterium species Micrococcus sp.[41], to our knowledge none (with the exception of BT) overlap with those observed in this study. Furthermore, the majority of reported biogenic benzothiazoles contain hydroxy groups, a functional group that is notably absent from all benzothiazole species observed in the dissolved and aerosol phases.[41,70] On the water collection day for SeaSCAPE bloom 3 (July 23 2019), the coastal current near San Diego ran from north to south at ~0.3 m/s (SCCOOS HR radar online mapping, documented in Harlan et al. 2010)[71], which would have transported the wastewater from the nearest up-current wastewater discharge point, Oceanside Outfall, to Scripps Pier on a timescale of 36 hours (see SI Figure 4.S5). While enrichment of benzothiazoles from the Oceanside Outfall is certainly a potential contributing factor, given distance and dilution more local sources also merit consideration. As illustrated in SI Figure 4.S6, there are multiple runoff, storm drain, and residential use discharge points along the beach surrounding the Scripps Pier sampling location, all of which could have washed road residues enriched in benzothiazoles into the coastal waters.

Finally, when compared to confidently identified biogenic and anthropogenic gas species, the temporal profile of gas-phase BT is more similar to those of multiple positively identified anthropogenic species than to any positively identified biogenic species or

biological indicators, as illustrated in SI Figure 4.S3 and previously noted in greater detail. The temporal evolution of a suite of anthropogenic gases suggests that some process related to the establishment of thermal, physical, and chemical equilibrium within the wave channel and ISV caused a lagged peak in some anthropogenic VOCs compared to bulk water temperatures and is not indicative of a significant source of benzothiazole within the wave channel. The identification of known tracers of anthropogenic benzothiazole pollution, multiple logical local sources of anthropogenic benzothiazole runoff, similarity in benzothiazole and some other known anthropogenic VOC temporal profiles, and absence of known biogenic benzothiazoles all lead to the conclusion that the source of benzothiazoles in the SeaSCAPE and Scripps Pier studies is primarily anthropogenic in nature. Scripps Pier lies within the San Diego Scripps State Marine Conservation Area, a relatively clean and protected area of coastline. Given this, the results from this study may be considered a relatively conservative lens into the extent to which these anthropogenic marine contaminants may influence the composition of aerosol and gas-phase emissions in coastal areas.

Benzothiazole Oxidation and Secondary Aerosol Formation

In order to better understand the eventual fate of atmospheric benzothiazole in both coastal marine and urban settings, we investigated the atmospheric oxidation and aerosol formation potential of gas-phase BT in a controlled laboratory oxidation experiment. When a plug of BT (62 ug in 5 uL of MeOH or 1%) was oxidized in the PAM-OFR to an equivalent 5 days of atmospheric aging, secondary aerosol formed, as illustrated by the new particle formation event in Figure 4.4. From this experiment, we produced a cumulative mass of 7.2 μg of aerosol during the plug injection experiment. Further experiments are necessary to determine the aerosol yield of BT under typical atmospheric conditions, but these results suggest it could be significant. Products formed from this aging, analyzed using previously described HESI Orbitrap mass spectrometry, primarily include reduced nitrogen and CHON species, with aerosol sulfur primarily in the form of sulfuric acid as shown in the identified molecular species listed in SI Table S4. In the second previously described benzothiazole oxidation experiment, in which a constant source of gas-phase benzothiazole from a perm tube diluted to 12.8 ppb was oxidized at three aging equivalents ranging from 2.9 to 4.7 days, both aerosol and SO_2 production were observed to increase with oxidative aging over this range, as illustrated in SI Figure 4.S11. Aerosol mass concentrations ranged from 9 ± 2 μg/m^3 at 2.9 days to 19 ± 2 μg/m^3 at 4.7 days, and produced SO_2 concentrations ranged from $.67 \pm .08$ ppb at 2.9 days to $1.3 \pm .12$ ppb at 4.7 days. The aerosol size distributions peak near 20 nm particle diameter for all exposure experiments, indicating new particle formation from benzothiazole oxidation (Figure 4.S11). While conditions in the PAM-OFR do not directly mimic those of the ambient atmosphere, the results of the benzothiazole oxidation study indicate two things: first, gas-phase BT has the capacity to contribute to SMA (or more generally secondary aerosol) formation and second, BT oxidation has the capacity to form sulfur dioxide and

sulfuric acid in the atmosphere, possibly suggesting the presence of sulfate aerosol produced during BT oxidation. This finding has relevance for polluted coastal marine environments, in which both benzothiazole and other anthropogenic marine pollutants should be evaluated for their potential influence on the abundance and characteristics of secondary marine aerosol. It also suggests implications for urban environments, in which benzothiazoles are highly abundant and may contribute to secondary organic aerosol formation and urban smog.

4.5 Acknowledgements

We would like to thank the NSF Center for Aerosol Impacts on Chemistry of the Environment (an NSF Chemical Innovation Center, (CHE-1801971)) and the entire experimental campaign team which participated in the SeaSCAPE for their support of this work. We specifically thank Julie Dinasquet, Francesca Malfatti, Clare Morris, and Daniel Crocker for their contribution of biological indicator data in support of this publication. The authors gratefully acknowledge the support of the National Science Foundation Graduate Research Fellowship Program (DGE-1650112) as well as Neal Arakawa and the use of facilities at the Environment Complex Analysis Laboratory (ECAL) at the University of California San Diego.

4.6 References

(1) Azam, F.; Fenchel, T.; Field, J. G.; Prog, S.; Gray, J. S.; Meyer-Reil L.A.; Thingstad F. The Ecological Role of Water-Column Microbes in the Sea Placing Marine Mixoplankton in Context View Project DOMAINE View Project MARINE ECOLOGY-PROGRESS SERIES The Ecological Role of Water-Column Microbes in the Sea*. *Mar. Ecol. Prog.* **1983**, *10*, 257–263. https://doi.org/10.3354/meps010257.

(2) Cochran, R. E.; Laskina, O.; Trueblood, J. V.; Estillore, A. D.; Morris, H. S.; Jayarathne, T.; Sultana, C. M.; Lee, C.; Lin, P.; Laskin, J.; Laskin, A.; Dowling, J. A.; Qin, Z.; Cappa, C. D.; Bertram, T. H.; Tivanski, A. V.; Stone, E. A.; Prather, K. A.; Grassian, V. H. Molecular Diversity of Sea Spray Aerosol Particles: Impact of Ocean Biology on Particle Composition and Hygroscopicity. *Chem* **2017**, *2* (5), 655–667. https://doi.org/10.1016/j.chempr.2017.03.007.

(3) Tovar-Sánchez, A.; Sánchez-Quiles, D.; Basterretxea, G.; Benedé, J. L.; Chisvert, A.; Salvador, A.; Moreno-Garrido, I.; Blasco, J. Sunscreen Products as Emerging Pollutants to Coastal Waters. *PLoS One* **2013**, *8* (6), e65451. https://doi.org/10.1371/journal.pone.0065451.

(4) Zeng, E. Y.; Vista, C. L. Organic Pollutants in the Coastal Environment off San Diego, California. 1. Source Identification and Assessment by Compositional Indices of Polycyclic Aromatic Hydrocarbons. *Environ. Toxicol. Chem.* **1997**, *16*

(2), 179–188. https://doi.org/10.1002/etc.5620160212.

(5) Elliott, J. E.; Elliott, K. H. Tracking Marine Pollution. *Science*. American Association for the Advancement of Science May 2013, pp 556–558. https://doi.org/10.1126/science.1235197.

(6) Glibert, P. M.; Maranger, R.; Sobota, D. J.; Bouwman, L. The Haber Bosch-Harmful Algal Bloom (HB-HAB) Link. *Environ. Res. Lett.* **2014**, *9* (10), 105001. https://doi.org/10.1088/1748-9326/9/10/105001.

(7) Kahru, M.; Mitchell, B. G. Ocean Color Reveals Increased Blooms in Various Parts of the World. *Eos (Washington. DC).* **2008**, *89* (18), 170. https://doi.org/10.1029/2008EO180002.

(8) Dale, B.; Edwards, M.; Reid, P. C. Climate Change and Harmful Algal Blooms. *Ecol. Stud.* **2006**, *189*, 367–378.

(9) Paerl, H.; Huisman, J. Blooms like It Hot. *Science (80-.).* **2008**. https://doi.org/10.1126/science.1156721.

(10) BLANCHARD, D. C. Bubble Scavenging and the Water-to-Air Transfer of Organic Material in the Sea. *Adv. Chem.* **1975**, *145*, 360–387. https://doi.org/10.1021/ba-1975-0145.ch018.

(11) Carpenter, L. J.; Archer, S. D.; Beale, R. Ocean-Atmosphere Trace Gas Exchange. *Chem. Soc. Rev.* **2012**, *41* (19), 6473–6506. https://doi.org/10.1039/c2cs35121h.

(12) O'Dowd, C. D.; De Leeuw, G. Marine Aerosol Production: A Review of the Current Knowledge. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. Royal Society July 2007, pp 1753–1774. https://doi.org/10.1098/rsta.2007.2043.

(13) Trueblood, J. V.; Wang, X.; Or, V. W.; Alves, M. R.; Santander, M. V.; Prather, K. A.; Grassian, V. H. The Old and the New: Aging of Sea Spray Aerosol and Formation of Secondary Marine Aerosol through OH Oxidation Reactions. *ACS Earth Sp. Chem.* **2019**, *3* (10), 2307–2314. https://doi.org/10.1021/acsearthspacechem.9b00087.

(14) O'Dowd, C. D.; Facchini, M. C.; Cavalli, F.; Ceburnis, D.; Mircea, M.; Decesari, S.; Fuzzi, S.; Young, J. Y.; Putaud, J. P. Biogenically Driven Organic Contribution to Marine Aerosol. *Nature* **2004**, *431* (7009), 676–680. https://doi.org/10.1038/nature02959.

(15) Lee, H. D.; Morris, H. S.; Laskina, O.; Sultana, C. M.; Lee, C.; Jayarathne, T.; Cox, J. L.; Wang, X.; Hasenecz, E. S.; Demott, P. J.; Bertram, T. H.; Cappa, C.

D.; Stone, E. A.; Prather, K. A.; Grassian, V. H.; Tivanski, A. V. Organic Enrichment, Physical Phase State, and Surface Tension Depression of Nascent Core-Shell Sea Spray Aerosols during Two Phytoplankton Blooms. *ACS Earth Sp. Chem.* **2020**, *4* (4), 650–660. https://doi.org/10.1021/acsearthspacechem.0c00032.

(16) Hodas, N.; Zuend, A.; Mui, W.; Flagan, R. C.; Seinfeld, J. H. Influence of Particle-Phase State on the Hygroscopic Behavior of Mixed Organic–Inorganic Aerosols. *Atmos. Chem. Phys.* **2015**, *15* (9), 5027–5045. https://doi.org/10.5194/acp-15-5027-2015.

(17) Fuentes, E.; Coe, H.; Green, D.; McFiggans, G. On the Impacts of Phytoplankton-Derived Organic Matter on the Properties of the Primary Marine Aerosol - Part 2: Composition, Hygroscopicity and Cloud Condensation Activity. *Atmos. Chem. Phys.* **2011**, *11* (6), 2585–2602. https://doi.org/10.5194/acp-11-2585-2011.

(18) Myriokefalitakis, S.; Vignati, E.; Tsigaridis, K.; Papadimas, C.; Sciare, J.; Mihalopoulos, N.; Facchini, M. C.; Rinaldi, M.; Dentener, F. J.; Ceburnis, D.; Hatzianastasiou, N.; O'Dowd, C. D.; van Weele, M.; Kanakidou, M. Global Modeling of the Oceanic Source of Organic Aerosols. *Adv. Meteorol.* **2010**, *2010*, 1–16. https://doi.org/10.1155/2010/939171.

(19) Mayer, K. J.; Wang, X.; Santander, M. V.; Mitts, B. A.; Sauer, J. S.; Sultana, C. M.; Cappa, C. D.; Prather, K. A. Secondary Marine Aerosol Plays a Dominant Role over Primary Sea Spray Aerosol in Cloud Formation. *ACS Cent. Sci.* **2020**. https://doi.org/10.1021/acscentsci.0c00793.

(20) Zheng, G.; Kuang, C.; Uin, J.; Watson, T.; Wang, J. Large Contribution of Organics to Condensational Growth and Formation of Cloud Condensation Nuclei (CCN) in the Remote Marine Boundary Layer. *Atmos. Chem. Phys.* **2020**, *20* (21), 12515–12525. https://doi.org/10.5194/acp-20-12515-2020.

(21) Kirkpatrick, B.; Fleming, L. E.; Squicciarini, D.; Backer, L. C.; Clark, R.; Abraham, W.; Benson, J.; Cheng, Y. S.; Johnson, D.; Pierce, R.; Zaias, J.; Bossart, G. D.; Baden, D. G. Literature Review of Florida Red Tide: Implications for Human Health Effects. *Harmful Algae*. Elsevier April 1, 2004, pp 99–115. https://doi.org/10.1016/j.hal.2003.08.005.

(22) Grattan, L. M.; Holobaugh, S.; Morris, J. G. Harmful Algal Blooms and Public Health. *Harmful Algae*. Elsevier B.V. July 1, 2016, pp 2–8. https://doi.org/10.1016/j.hal.2016.05.003.

(23) Walsh, J. J.; Lenes, J. M.; Weisberg, R. H.; Zheng, L.; Hu, C.; Fanning, K. A.; Snyder, R.; Smith, J. More Surprises in the Global Greenhouse: Human Health

Impacts from Recent Toxic Marine Aerosol Formations, Due to Centennial Alterations of World-Wide Coastal Food Webs. *Marine Pollution Bulletin*. Elsevier Ltd March 15, 2017, pp 9–40. https://doi.org/10.1016/j.marpolbul.2016.12.053.

(24) Vila-Costa, M.; Cerro-Gálvez, E.; Martínez-Varela, A.; Casas, G.; Dachs, J. Anthropogenic Dissolved Organic Carbon and Marine Microbiomes. *ISME J.* **2020**, *14* (10), 2646–2648. https://doi.org/10.1038/s41396-020-0712-5.

(25) Wang, Z.; Walker, G. W.; Muir, D. C. G.; Nagatani-Yoshida, K. Toward a Global Understanding of Chemical Pollution: A First Comprehensive Analysis of National and Regional Chemical Inventories. *Environ. Sci. Technol.* **2020**, *54* (5), 2575–2584. https://doi.org/10.1021/acs.est.9b06379.

(26) Bernhardt, E. S.; Rosi, E. J.; Gessner, M. O. Synthetic Chemicals as Agents of Global Change. *Front. Ecol. Environ.* **2017**, *15* (2), 84–90. https://doi.org/10.1002/fee.1450.

(27) Ott, A.; Martin, T. J.; Whale, G. F.; Snape, J. R.; Rowles, B.; Galay-Burgos, M.; Davenport, R. J. Improving the Biodegradability in Seawater Test (OECD 306). *Sci. Total Environ.* **2019**, *666*, 399–404. https://doi.org/10.1016/j.scitotenv.2019.02.167.

(28) Cincinelli, A.; Stortini, A. M.; Perugini, M.; Checchini, L.; Lepri, L. Organic Pollutants in Sea-Surface Microlayer and Aerosol in the Coastal Environment of Leghorn - (Tyrrhenian Sea). *Mar. Chem.* **2001**, *76* (1–2), 77–98. https://doi.org/10.1016/S0304-4203(01)00049-4.

(29) Cerro-Gálvez, E.; Dachs, J.; Lundin, D.; Fernández-Pinos, M. C.; Sebastián, M.; Vila-Costa, M. Responses of Coastal Marine Microbiomes Exposed to Anthropogenic Dissolved Organic Carbon. *Environ. Sci. Technol.* **2021**. https://doi.org/10.1021/acs.est.0c07262.

(30) Cunliffe, M.; Engel, A.; Frka, S.; Gašparović, B. Ž.; Guitart, C.; Murrell, J. C.; Salter, M.; Stolle, C.; Upstill-Goddard, R.; Wurl, O. Sea Surface Microlayers: A Unified Physicochemical and Biological Perspective of the Air-Ocean Interface. *Progress in Oceanography*. Pergamon February 1, 2013, pp 104–116. https://doi.org/10.1016/j.pocean.2012.08.004.

(31) Tervahattu, H.; Juhanoja, J.; Kupiainen, K. Identification of an Organic Coating on Marine Aerosol Particles by TOF-SIMS. *J. Geophys. Res. Atmos.* **2002**, *107* (D16), ACH 18-1. https://doi.org/10.1029/2001JD001403.

(32) Vignati, E.; Facchini, M. C.; Rinaldi, M.; Scannell, C.; Ceburnis, D.; Sciare, J.; Kanakidou, M.; Myriokefalitakis, S.; Dentener, F.; O'Dowd, C. D. Global Scale

Emission and Distribution of Sea-Spray Aerosol: Sea-Salt and Organic Enrichment. *Atmos. Environ.* **2010**, *44* (5), 670–677. https://doi.org/10.1016/J.ATMOSENV.2009.11.013.

(33) Prather, K. A.; Bertram, T. H.; Grassian, V. H.; Deane, G. B.; Stokes, M. D.; DeMott, P. J.; Aluwihare, L. I.; Palenik, B. P.; Azam, F.; Seinfeld, J. H.; Moffet, R. C.; Molina, M. J.; Cappa, C. D.; Geiger, F. M.; Roberts, G. C.; Russell, L. M.; Ault, A. P.; Baltrusaitis, J.; Collins, D. B.; Corrigan, C. E.; Cuadra-Rodriguez, L. A.; Ebben, C. J.; Forestieri, S. D.; Guasco, T. L.; Hersey, S. P.; Kim, M. J.; Lambert, W. F.; Modini, R. L.; Mui, W.; Pedler, B. E.; Ruppel, M. J.; Ryder, O. S.; Schoepp, N. G.; Sullivan, R. C.; Zhao, D. Bringing the Ocean into the Laboratory to Probe the Chemical Complexity of Sea Spray Aerosol. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (19), 7550–7555. https://doi.org/10.1073/pnas.1300262110.

(34) Engel, A.; Bange, H. W.; Cunliffe, M.; Burrows, S. M.; Friedrichs, G.; Galgani, L.; Herrmann, H.; Hertkorn, N.; Johnson, M.; Liss, P. S.; Quinn, P. K.; Schartau, M.; Soloviev, A.; Stolle, C.; Upstill-Goddard, R. C.; van Pinxteren, M.; Zäncker, B. The Ocean's Vital Skin: Toward an Integrated Understanding of the Sea Surface Microlayer. *Frontiers in Marine Science*. Frontiers Media S. A May 30, 2017, p 165. https://doi.org/10.3389/fmars.2017.00165.

(35) Liao, C.; Kim, U. J.; Kannan, K. A Review of Environmental Occurrence, Fate, Exposure, and Toxicity of Benzothiazoles. *Environ. Sci. Technol.* **2018**, *52* (9), 5007–5026. https://doi.org/10.1021/acs.est.7b05493.

(36) Halsband, C.; Sørensen, L.; Booth, A. M.; Herzke, D. Car Tire Crumb Rubber: Does Leaching Produce a Toxic Chemical Cocktail in Coastal Marine Systems? *Front. Environ. Sci.* **2020**, *8*, 125. https://doi.org/10.3389/fenvs.2020.00125.

(37) Wever, H. De; Research, H. V. Biodegradation and Toxicity of Benzothiazoles. *Water Res.* **1997**.

(38) Fries, E.; Gocht, T.; Klasmeier, J. Occurrence and Distribution of Benzothiazole in the Schwarzbach Watershed (Germany). *J. Environ. Monit.* **2011**, *13* (10), 2838–2843. https://doi.org/10.1039/C1EM10474H.

(39) Brownlee, B. G.; Carey, J. H.; Fox, M. E. A Review of Benzothiazoles in the Aquatic Environment, Scientific Series No. 126. Inland Waters Directorate 1981.

(40) Zeng, E. Y.; Tran, K.; Young, D. Evaluation of Potential Molecular Markers for Urban Stormwater Runoff. *Environ. Monit. Assess.* **2004**, *90* (1–3), 23–43. https://doi.org/10.1023/B:EMAS.0000003564.24169.86.

(41) Stierle, A. A.; Cardellina, J. H.; Singleton, F. L. Benzothiazoles from a Putatitve

Bacterial Symbiont of the Marine Sponge Tedania Ignis. *Tetrahedron Lett.* **1991**, *32* (37), 4847–4848. https://doi.org/10.1016/S0040-4039(00)93476-2.

(42) Chhalodia, A. K.; Rinkel, J.; Konvalinkova, D.; Petersen, J.; Dickschat, J. S. Identification of Volatiles from Six Marine Celeribacter Strains. *Beilstein J. Org. Chem.* **2021**, *17* (1), 420–430. https://doi.org/10.3762/BJOC.17.38.

(43) Kloepfer, A.; Jekel, M.; Reemtsma, T. Occurrence, Sources, and Fate of Benzothiazoles in Municipal Wastewater Treatment Plants. *Environ. Sci. Technol.* **2005**, *39* (10), 3792–3798. https://doi.org/10.1021/es048141e.

(44) Sorahan, T. Bladder Cancer Risks in Workers Manufacturing Chemicals for the Rubber Industry. *Occup. Med. (Chic. Ill).* **2008**, *58* (7), 496–501. https://doi.org/10.1093/occmed/kqn104.

(45) Sorahan, T. Cancer Risks in Chemical Production Workers Exposed to 2-Mercaptobenzothiazole. *Occup. Environ. Med.* **2009**, *66* (4), 269–273. https://doi.org/10.1136/oem.2008.041400.

(46) Liao, X.; Zou, T.; Chen, M.; Song, Y.; Yang, C.; Qiu, B.; Chen, Z. F.; Tsang, S. Y.; Qi, Z.; Cai, Z. Contamination Profiles and Health Impact of Benzothiazole and Its Derivatives in PM2.5 in Typical Chinese Cities. *Sci. Total Environ.* **2021**, *755*, 142617. https://doi.org/10.1016/j.scitotenv.2020.142617.

(47) Sauer, J. S.; Mayer, K. J.; Lee, C.; Alves, M. R.; Amiri, S.; Bahaveolos, C.; Barnes, E. B.; Crocker, D. R.; Dinasquet, J.; Garofalo, L. A.; Kaluarachchi, C. P.; Kilgour, D.; Mael, L.; Mitts, B. A.; Moon, D. R.; Morris, C. K.; Moore, A. N.; Ni, C.-M.; Pendergraft, M. A.; Petras, D.; Simpson, R.; Smith, S.; Tumminello, P. R.; Walker, J. L.; DeMott, P. J.; Farmer, D. K.; Goldstein, A. H.; Grassian, V. H.; Jaffe, J. S.; Malfatti, F.; Martz, T. R.; Slade, J.; Tivanski, A. V; Bertram, T. H.; Cappa, C. D.; Prather, K. A. The Sea Spray Chemistry and Particle Evolution Study (SeaSCAPE): Overview and Experimental Methods . *Environ. Sci. Process. Impacts* **2021**. https://doi.org/10.33774/CHEMRXIV-2021-SM7VW.

(48) Wang, X.; Sultana, C. M.; Trueblood, J.; Hill, T. C. J.; Malfatti, F.; Lee, C.; Laskina, O.; Moore, K. A.; Beall, C. M.; McCluskey, C. S.; Cornwell, G. C.; Zhou, Y.; Cox, J. L.; Pendergraft, M. A.; Santander, M. V.; Bertram, T. H.; Cappa, C. D.; Azam, F.; DeMott, P. J.; Grassian, V. H.; Prather, K. A. Microbial Control of Sea Spray Aerosol Composition: A Tale of Two Blooms. *ACS Cent. Sci.* **2015**, *1* (3), 124–131. https://doi.org/10.1021/acscentsci.5b00148.

(49) Dittmar, T.; Koch, B.; Hertkorn, N.; Kattner, G. A Simple and Efficient Method for the Solid-Phase Extraction of Dissolved Organic Matter (SPE-DOM) from Seawater. *Limnol. Oceanogr. Methods* **2008**, *6* (6), 230–235.

https://doi.org/10.4319/lom.2008.6.230.

(50) Roveretto, M.; Li, M.; Hayeck, N.; Brüggemann, M.; Emmelin, C.; Perrier, S.; George, C. Real-Time Detection of Gas-Phase Organohalogens from Aqueous Photochemistry Using Orbitrap Mass Spectrometry. *ACS Earth Sp. Chem.* **2019**, *3* (3), 329–334. https://doi.org/10.1021/acsearthspacechem.8b00209.

(51) Worton, D. R.; Decker, M.; Isaacman-VanWertz, G.; Chan, A. W. H.; Wilson, K. R.; Goldstein, A. H. Improved Molecular Level Identification of Organic Compounds Using Comprehensive Two-Dimensional Chromatography, Dual Ionization Energies and High Resolution Mass Spectrometry. *Analyst* **2017**, *142* (13), 2395–2403. https://doi.org/10.1039/c7an00625j.

(52) Dörter, M.; Odabasi, M.; Yenisoy-Karakaş, S. Source Apportionment of Biogenic and Anthropogenic VOCs in Bolu Plateau. *Sci. Total Environ.* **2020**, *731*, 139201. https://doi.org/10.1016/j.scitotenv.2020.139201.

(53) Reddy, C. M.; Quinn, J. G. Environmental Chemistry of Benzothiazeles Derived from Rubber. *Environ. Sci. Technol.* **1997**, *31* (10), 2847–2853. https://doi.org/10.1021/es970078o.

(54) Sander, R. Compilation of Henry's Law Constants (Version 4.0) for Water as Solvent. *Atmos. Chem. Phys.* **2015**, *15* (8), 4399–4981. https://doi.org/10.5194/acp-15-4399-2015.

(55) Kowalska, K.; Felis, E.; Sochacki, A.; Bajkacz, S. Removal and Transformation Pathways of Benzothiazole and Benzotriazole in Membrane Bioreactors Treating Synthetic Municipal Wastewater. *Chemosphere* **2019**, *227*, 162–171. https://doi.org/10.1016/j.chemosphere.2019.04.037.

(56) Crocker, D.; Kaluarachchi, C.; Cao, R.; Dinasquet, J.; Franklin, E.; Morris, C.; Martz, T.; Malfatti, F.; Goldstein, A.; Tivanski, A.; Prather, K.; Thiemens, M. Isotopic Insights into Organic Composition Differences between Supermicron and Submicron Sea Spray Aerosol. *Environ. Sci. Technol.* **2021**.

(57) Tran, T. M.; Kannan, K. Occurrence of Cyclic and Linear Siloxanes in Indoor Air from Albany, New York, USA, and Its Implications for Inhalation Exposure. *Sci. Total Environ.* **2015**, *511*, 138–144. https://doi.org/10.1016/j.scitotenv.2014.12.022.

(58) Bletsou, A. A.; Asimakopoulos, A. G.; Stasinakis, A. S.; Thomaidis, N. S.; Kannan, K. Mass Loading and Fate of Linear and Cyclic Siloxanes in a Wastewater Treatment Plant in Greece. *Environ. Sci. Technol.* **2013**, *47* (4), 1824–1832. https://doi.org/10.1021/es304369b.

(59) Hong, W. J.; Jia, H.; Liu, C.; Zhang, Z.; Sun, Y.; Li, Y. F. Distribution, Source, Fate and Bioaccumulation of Methyl Siloxanes in Marine Environment. *Environ. Pollut.* **2014**, *191*, 175–181. https://doi.org/10.1016/j.envpol.2014.04.033.

(60) Mao, F.; He, Y.; Gin, K. Y. H. Occurrence and Fate of Benzophenone-Type UV Filters in Aquatic Environments: A Review. *Environmental Science: Water Research and Technology*. Royal Society of Chemistry February 1, 2019, pp 209–223. https://doi.org/10.1039/c8ew00539g.

(61) Zeng, E. Y.; Vista, C. L. Organic Pollutants in the Coastal Environment off San Diego, California. 1. Source Identification and Assessment by Compositional Indices of Polycyclic Aromatic Hydrocarbons. *Environ. Toxicol. Chem.* **1997**, *16* (2), 179–188. https://doi.org/10.1002/etc.5620160212.

(62) Aprem, A. S.; Joseph, K.; Mathew, T.; Altstaedt, V.; Thomas, S. Studies on Accelerated Sulphur Vulcanization of Natural Rubber Using 1-Phenyl-2, 4-Dithiobiuret/Tertiary Butyl Benzothiazole Sulphenamide. *Eur. Polym. J.* **2003**, *39* (7), 1451–1460. https://doi.org/10.1016/S0014-3057(02)00382-8.

(63) Nowak, J. A.; Shrestha, P. M.; Weber, R. J.; McKenna, A. M.; Chen, H.; Coates, J. D.; Goldstein, A. H. Comprehensive Analysis of Changes in Crude Oil Chemical Composition during Biosouring and Treatments. *Environ. Sci. Technol.* **2018**, *52* (3), 1290–1300. https://doi.org/10.1021/acs.est.7b05346.

(64) Hardy, J. T. The Sea Surface Microlayer: Biology, Chemistry and Anthropogenic Enrichment. *Progress in Oceanography*. Pergamon January 1, 1982, pp 307–328. https://doi.org/10.1016/0079-6611(82)90001-5.

(65) SPIEL, D. E.; LEEUW, G. DE. FORMATION AND PRODUCTION OF SEA SPRAY AEROSOL. *J. Aerosol Sci.* **1996**.

(66) Cipriano, R.; Blanchard, D.; Hogan, A.; Lala, G. On the Production of Aitken Nuclei from Breaking Waves and Their Role in the Atmosphere. *J. Atmos. Sci.* **1983**.

(67) Wang, X.; Deane, G. B.; Moore, K. A.; Ryder, O. S.; Stokes, M. D.; Beall, C. M.; Collins, D. B.; Santander, M. V.; Burrows, S. M.; Sultana, C. M.; Prather, K. A. The Role of Jet and Film Drops in Controlling the Mixing State of Submicron Sea Spray Aerosol Particles. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (27), 6978–6983. https://doi.org/10.1073/pnas.1702420114.

(68) Fiehn, O.; Reemtsma, T.; Jekel, M. Extraction and Analysis of Various Benzothiazoles from Industrial Wastewater. *Anal. Chim. Acta* **1994**, *295* (3), 297–305. https://doi.org/10.1016/0003-2670(94)80235-1.

(69) De Wever, H.; Verachtert, H. Biodegradation and Toxicity of Benzothiazoles. *Water Res.* **1997**, *31* (11), 2673–2684. https://doi.org/10.1016/S0043-1354(97)00138-3.

(70) Le Bozec, L.; Moody, C. J. Naturally Occurring Nitrogensulfur Compounds. the Benzothiazole Alkaloids. *Aust. J. Chem.* **2009**, *62* (7), 639–647. https://doi.org/10.1071/CH09126.

(71) Harlan, J.; Terrill, E.; Hazard, L.; Keen, C.; Barrick, D.; Whelan, C.; Howden, S.; Kohut, J. The Integrated Ocean Observing System High-Frequency Radar Network: Status and Local, Regional, and National Applications NOAA IOOS ® Program. *Mar. Technol. Soc. J.* **2010**, *44* (6).

4.7 Tables and Figures

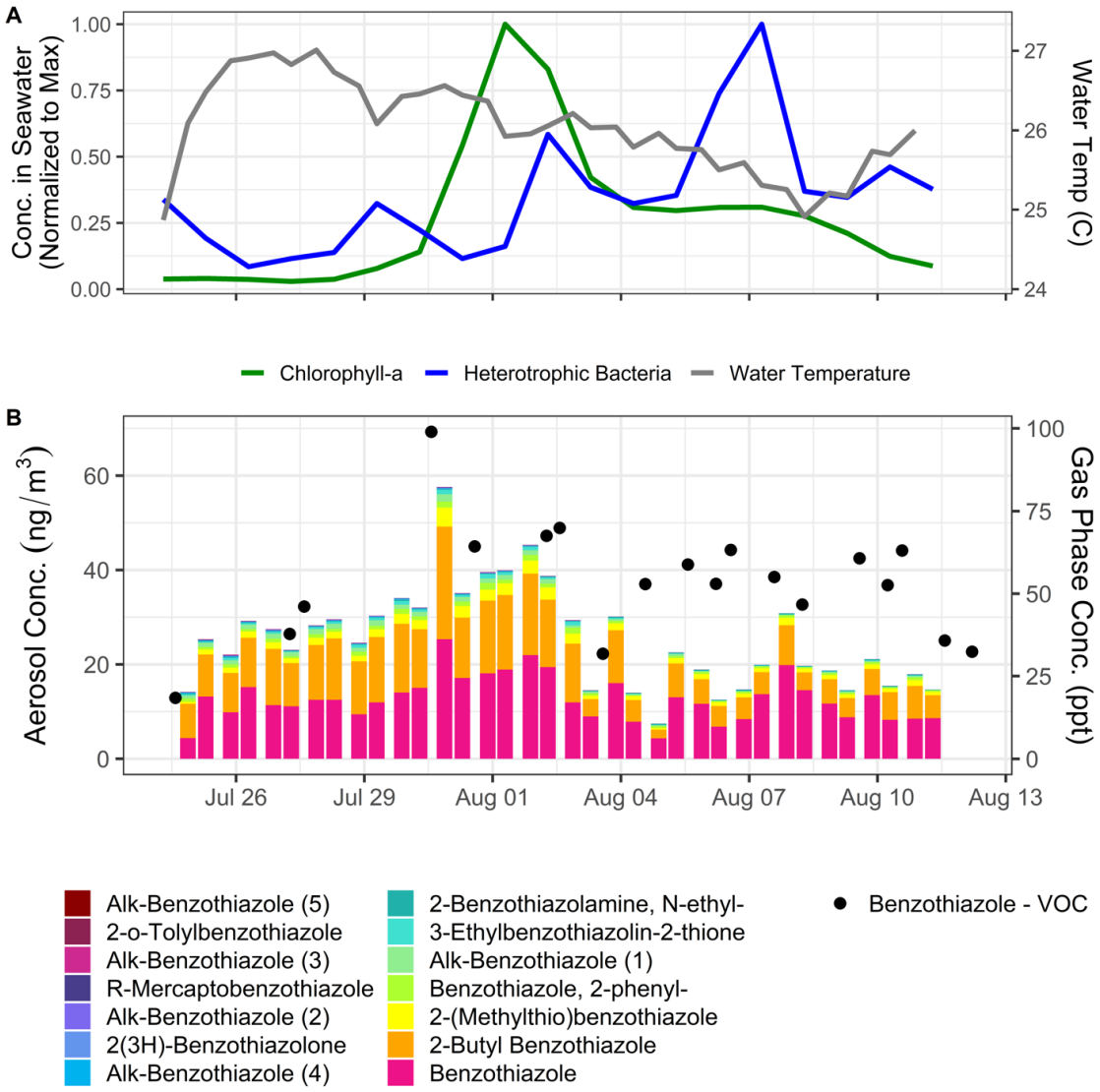


Figure 4.1. Time series of biological activity indicators and water temperature (A) and atmospheric benzothiazoles (B) at SeaSCAPE 2019 bloom 3 experiment. Atmospheric benzothiazole concentrations are differentiated into gas phase BT from the dome (black circles) and submicron nascent sea spray aerosol phase benzothiazoles from the wave chamber (colored bars). Detailed discussion of uncertainties can be found in supplemental information.

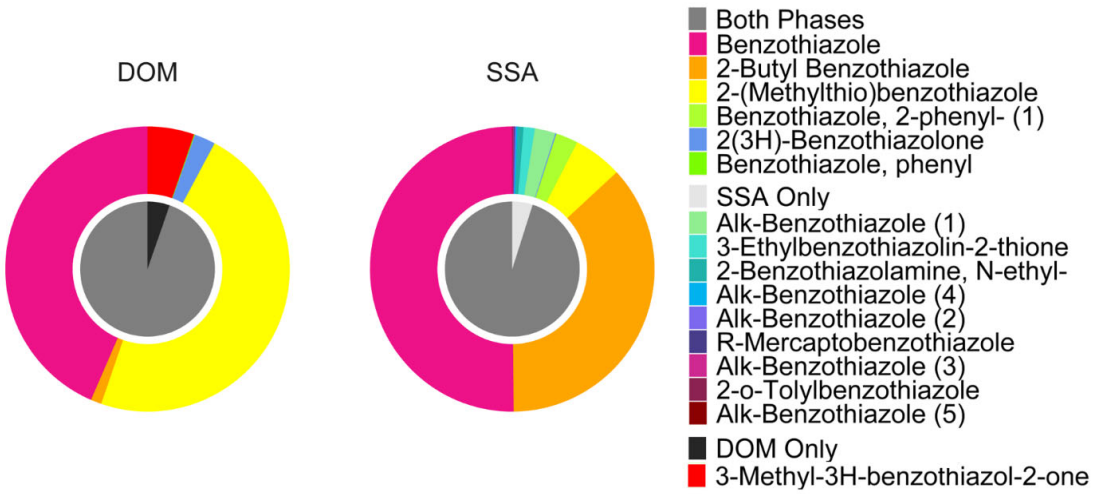


Figure 4.2. Molecular distributions and molecular overlap (by mass) of benzothiazoles observed in DOM (left) and nascent sea spray aerosol (right) at SeaSCAPE 2019

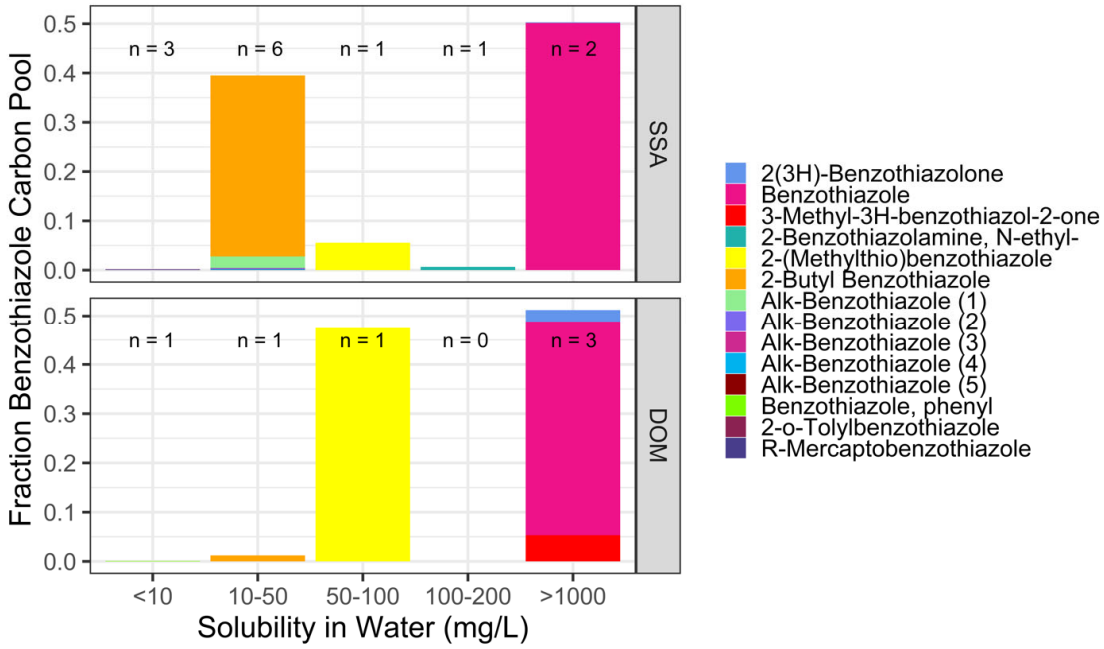


Figure 4.3. Solubility distributions of benzothiazoles weighted by contribution to cumulative observed benzothiazole carbon pools in the dissolved (DOM) and aerosol (SSA) phases. Number of individual species within each solubility bracket in each phase indicated by n. No solubility information for 3-Ethylbenzothiazolin-2-thione is available and it is therefore excluded from visualization.

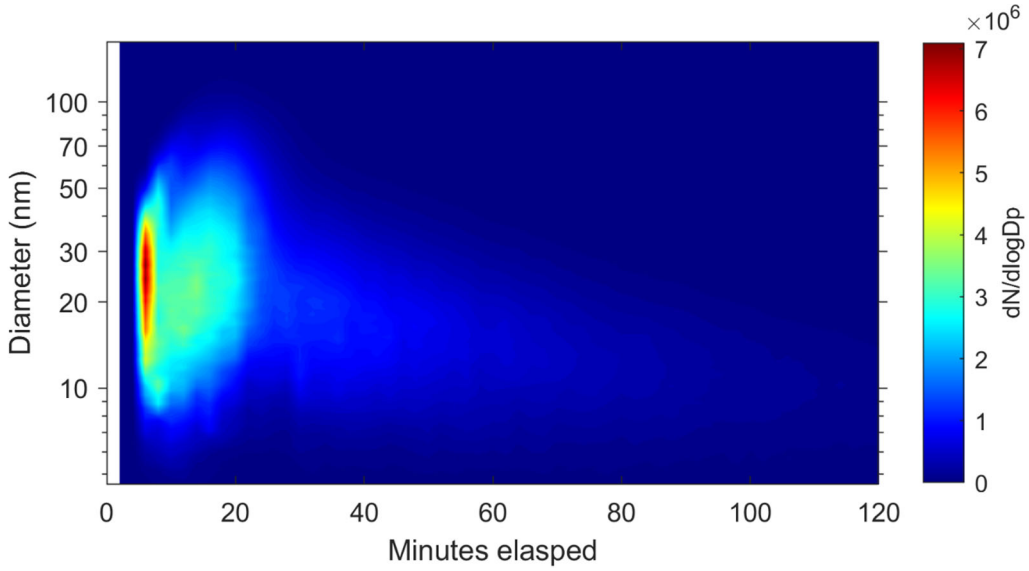


Figure 4.4. Nucleation of new particles from benzothiazole oxidation in PAM-OFR. Liquid benzothiazole dissolved in a methanol carrier (1% BT) is introduced at 5 minutes.

Table 4.1. Identities, solubilities, and concentrations of benzothiazoles observed at SeaSCAPE. Aerosol phase concentrations are reported as an average and standard deviation of concentrations over the full experimental campaign, while dissolved phase concentrations are reported from the initial pre-experiment sample for comparison to other ambient sampling measurements of benzothiazoles. BQL indicates a species detected but present at below quantification limit levels.

Compound Identity	Solubility in Water (mg/L)	Aerosol-Phase Mean Concentration [s.d.] (ng/m^3)	DOM-Phase Concentration- 7/23 Ambient Sample (μg/L)
Benzothiazole[A]	1684	13 [4.7]	.29
Phenyl Benzothiazole[A†]	8.804	.64 [.30]	.0093
2-(Methylthio) benzothiazole[A]	66.61	1.4 [.77]	.96
Butyl-Benzothiazole[C]	12.8	9.2 [4.8]	.017
Alk-Benzothiazole (1)[B]	12.8[1]	.59 [.44]	
Alk-Benzothiazole (2)[B]	12.8[1]	.032 [.033]	
Alk-Benzothiazole (3)[B]	12.8[1]	.028 [.031]	
Alk-Benzothiazole (4)[B]	12.8[1]	.045 [.043]	
Alk-Benzothiazole (5)[B]	12.8[1]	.0067 [.0071]	
2-o-Tolylbenzothiazole[C]	4.81	.012 [.0088]	
3-Methyl-3H-benzothiazol-2-one[C]	1319		.31
3-Ethylbenzothiazolin-2-thione[D]		.32 [.26]	
N-Ethyl-2-Benzothiazoleamine[D]	142.2[2]	.17 [.043]	
R-Mercaptobenzothiazole[E]	0.05096[2]	.030 [.021]	
R- 2(3H)-Benzothiazolone[D]	2354[2]	.032 [.039]	.016

A: Isomer specific identification confirmed via authentic standard

B: Series of alkyl benzothiazoles, identified via high (>750) match factor with spectrum of butyl-benzothiazole and location in GCxGC space. Based on kovats indices, likely C4-C7 straight and branched chain alkyl benzothiazoles

C: Identified via high (>800) match factor with NIST14 mass spectral database entry along with Kovats index matches from previous published isolation where available

D: Classified as benzothiazole due to relatively high (>700, <800) NIST match factor with named benzothiazole or by high match factor but kovats index disagreement; novel compound without published mass spectra, tentatively identified as benzothiazole due to mass spectral indicators.

E. Identity unknown, but based on mass spectrum assigned identity of R-mercaptobenzothiazole, with R an unknown group likely containing heteroatom(s).

† Note that there are two distinct speciations of phenyl benzothiazoles observed between the two phases- 2-phenyl benzothiazole is observed in the aerosol phase, while an undetermined alternate phenyl benzothiazole isomer is observed in the dissolved phase.
1. No predicted or experimental solubility values for C5+ benzothiazoles, therefore assigned solubility of butyl-benzothiazole
2. Note that the exact identities of these species are unknown, but that the solubilities are those of the identified chemically similar benzothiazoles named in column 1

4.8 Supporting Information

4.SI.1: Sorbent Tube Details

Inert coated stainless-steel tubes with PTFE compression caps. Packing materials: Tenax TA 60/80, Carbograph 1 60/80, Carboxen 1003 40/60 (Camsco, custom order). Conditioned for 4 hours at 320 °C, Markes tube conditioner.

4.SI.2: GCxGC Materials and Methods

	GCxGC A	GCxGC B
Thermal Desorption Unit	Gerstel TDS-3 and TDSA2	Gerstel TDU 3.5 and MPS
Thermal Desorption Temp	320 °C	300 °C
Cooled Inlet System Material	Gerstel CIS4, quartz wool	Gerstel CIS glass wool
Cooled Inlet System Temperature	Trapping Temperature: 30°C Release Temperature: 320°C	Trapping Temperature: -150°C (Liquid Nitrogen cooled) Release Temperature: 300 °C
Carrier gas	Helium, 2 ml min^{-1}	Helium 2 ml min^{-1}
GCxGC Oven	Agilent 7890A	Agilent 7890A
GCxGC Column 1	Restek, Rxi-5Sil-MS, 60m × .25mm × .25 µm	Restek, Rtx-624, 30m × .25mm × 1.4 µm
GCxGC Column 2	Restek, Rtx-200MS, 1m × .25mm × .25 µm	Restek, Stabilwax, 1.5m × .25mm × .5 µm
Thermal Modulator	ZOEX cryogenic dual-stage thermal modulator; guard coumn (Restek, Siltek, 1.5m × .25mm × 250 µm)	ZOEX cryogenic dual-stage thermal modulator; guard column (Restek, Siltek, 1.5m × .25mm × 250 µm)
Modulation Period	2.3 s	5 s
GCxGC Primary Oven (Column 1)	Ramped from 40 to 320 °C at 3.5 °C min^{-1}	Ramped from 35 to 215 °C at 4 °C min^{-1}, 6 minute hold time

GCxGC Secondary Oven (Column 2)	Maintained at 15 °C warmer than column 1	Maintained at 20 °C warmer than column 1

Table 4.S1. GCxGC manufacturers and methods

4.SI.3: Filter Sampling Description

The sampled air was passed through a 2 m section of cooled copper tubing to condense and remove excess water and prevent sample saturation, a methodological approach described in Yee et al., 2018.[1] Supermicron aerosols were excluded using a greaseless cyclone (BGI Mesa Labs), operated at 21.8 LPM to achieve a cut point of PM_1. Filter changes were coordinated with the SeaSCAPE grow light schedule to generate one 14 hour "day" and 10 hour "night" sample per experiment day. Post collection, filters were wrapped in baked foil (12 hours at 550 °C), individually sealed in mylar bags, secondarily sealed in plastic bags, and frozen immediately (within 1 hour). Field blanks were collected before, after, and consistently throughout the bloom experiment at a rate of two field blanks every 5 days. Samples and blanks were stored primarily frozen and always below ambient temperatures prior to analysis in early 2020. Two SeaSCAPE aerosol sample filters were analyzed within two months of collection, and comparison to their reanalysis during the full analysis period confirmed sample preservation in storage.

4.SI.4: Sorbent Tube VOC Sampling Methods

The sorbent tubes were inert coated stainless steel, packed with Tenax TA, Carbograph 1, and Carboxen 1003 (SilcoNert™ coated, CAMSCO). These three materials were selected to enable capture of C_3-C_{20} organics while minimizing water retention. All collections were performed at 100 sccm. Sorbent tubes were conditioned for 4 hours at 320°C prior to use, were stored frozen prior to sampling, and were frozen immediately post sample collection, with field blank, room air, zero air, and empty ISV samples collected for background and contamination analysis, consistent with best practices advanced in Sheu et al., 2018.[2] Day (light) samples were collected at approximately 14:00 and night (dark) samples were collected at 04:00. Background BT concentrations observed in blanks collected from the empty ISV and ISV air intake were negligible compared to concentrations observed under experimental conditions.

4.SI.5: Pier VOCUS Sampling Methods

Measurements of benzothiazole gas phase mixing ratios were made from the end of the Ellen Browning Scripps Pier Memorial Pier (hereon SIO Pier) at the Scripps Institution of Oceanography in La Jolla, CA, USA during the month of September 2019. The SIO Pier site has been regularly for studies of coastal ocean trace gas exchange (Kim et al., 2014; Novak et al., 2019; Vermeuel et al., 2020). Benzothiazole was detected using the same Vocus PTR-TOF used in the SeaSCAPE study. The Vocus was housed in a temperature-controlled trailer at the end of the pier and sampled through a 19 m long

PFA inlet (0.625 cm backgrounds were determined by overflowing the full inlet line with dry ultrahigh purity nitrogen at the tip of the ambient sampling point. Instrument sensitivities to DMS were determined during ambient sampling by a two-point standard addition of a DMS gas standard (Praxiar, 5.08 ppm ± 5%) to the full sampling inlet every 2.5 to 4 hours. Benzothiazole detection sensitivity relative to DMS was determined in the laboratory post campaign, and ambient benzothiazole sensitivities were determined by scaling the infield DMS calibration factors by the relative instrument sensitivity of DMS to benzothiazole. The SIO Pier experiences a characteristic sea-breeze circulation pattern during summer where winds are from the ocean at moderate windspeeds (0-6 m s^{-1}) during daytime and are from land at night.

4.SI.6: DOM Collection Methods

In a methodology adapted from Dittmar et al. 2008[3], 20-L samples of water were collected into a cleaned polypropylene jug and immediately transferred to a nearby laboratory for extraction via SPE-PPL. All plasticware was cleaned using 3x methanol followed by a 3x Milli-Q water rinse before usage. The water was pumped through a series of filters: 10 micron, 0.7 micron, and 0.2 micron, using a peristaltic pump with PTFE-lined tubing at a rate of 100 mL/min. Backpressure and flowrate were kept to a minimum to prevent lysing of cells. Finally, the sample water was allowed to gravity filter through a pre-cleaned SPE-PPL cartridge at a rate of 3 drops/sec or less overnight. The sample was then washed and eluted with methanol three times and immediately dried down to solid. Samples were stored in a freezer at –18 °C.

4.SI.7: Headspace Gas Analysis by APCI-Orbitrap-MS/MS

Data was collected in positive mode, where the needle voltage was set to 4 kV, needle current at 5 mA, and vaporizer temperature at 150 °C. Sheath and auxiliary flow were set to zero. From the wave channel, 200 mL of surface water was collected and transferred into a 350 mL jacketed custom glass tube (Ace Glass Inc.) with quartz windows on each end. The surface area of the water sample in the tube was approximately 77 cm2. With a headspace of 150 mL, pure nitrogen gas was used as a carrier at a rate of 200 sccm. Temperature was regulated, at 20°C, and measured constantly to ensure minimal thermal variation (± 1°C) during the analysis.

4.SI.8: Sorbent Tube Analysis – Dry Purging

Sorbent tubes were dry purged in the sampling direction with ultrapure dry nitrogen to remove trapped water prior to offline analysis. Repeated tests in which sorbent tubes were loaded with relevant concentrations of target analyte standards, sampled humid air (produced by bubbled Milli Q water) at flowrates and intervals replicating those at the experimental campaign, and subjected to the dry purge protocol found no significant loss in benzothiazole recovery from dry purging.

4.SI.9: DOM, Aerosol and VOC analysis by GCxGC- Methodological Details

DOM samples were reconstituted in methanol immediately prior to analysis and injected onto quartz fiber filter segments to maximize analytical consistency between aerosol and dissolved phase samples. Both DOM and aerosol filter samples were doped with a custom blend of 23 deuterated internal standard compounds prior to analysis, allowing corrections for instrument condition and matrix effects across samples. Broadly, both instruments thermally desorb samples from the sorbent tube or filter media, focus desorbed samples on a cooled inlet system (CIS, Gerstel), then simultaneously release all desorbed analytes into the GC oven. Instrument A employs online derivatization during thermal desorption with MSTFA (n-methyl-n-trimethylsilyl-trifluoro-acetamide), which replaces OH groups with O-(Si(CH3)3) to enhance recovery of polar organics. Next, analytes are separated by both volatility and polarity by two GC columns in sequence, with the first to second column transition modulated by cryogenic focus and rapid thermal release. Separated analytes from GCxGC A are ionized by 70 eV using EI and detected by HR-ToF-MS (Tofwerk), resolving power 4000 acquired at 100 Hz. Separated analytes from GCxGC B are ionized by both 14 eV and 70 eV EI (alternating in 40 milliseconds allowing analysis of all peaks at dual ionization energies) and detected by ToF-MS (Markes BenchToF) at a resolving power of 1000 acquired at 50 Hz. These methods generate data for individual analytes separated in two chromatographic dimensions, and characterized by 70 eV EI mass spectra from GCxGC A and both 14 eV and 70 eV mass spectra from GCxGC B. Methodological details and documentation for GCxGC A can be found in Worton et al. 2017[4], and thermal desorption unit and gas chromatography methods, component manufacturers, and column materials for both GCxGC instruments may be found in SI table S1. Six-point calibration curves of custom external standard blends containing ~150 representative organic compounds were performed periodically throughout sample analysis for each sample medium class to maximize quantification accuracy.

4.SI.10: Benzothiazole Partitioning Experiment

To assess whether currently reported Henry's constants for BT (all of which are predicted based on structure rather than experimentally measured) could be many orders of magnitude different from true partitioning behavior (thereby explaining the higher than predicted BT observations at SeaSCAPE), the following partitioning experiment was performed. A solution of α-pinene, limonene, and benzothiazole (0.4 ug/L each) in simplified simulated sea water (milli-Q with 35 g/L NaCl and .5 g/L $NaHCO_3$) was mixed to simulate the observed aqueous phase concentration of benzothiazole in sampled water at SeaSCAPE. This concentration was selected as it is similar to the observed dissolved concentration of benzothiazole at the initial water sampling period. 250 ml of this solution was bubbled with ultrapure dry nitrogen at 10 sccm for 200 minutes to generate 2 L samples (the same sample volume collected experimentally at SeaSCAPE) in a temperature monitored water bath as recommended in Lee at al., 2012.[5] Samples were collected on sorbent tubes and analyzed on GCxGC B. Bubble height is ~20 cm and gas residence time in the headspace is estimated at 2 minutes, equal to the residence time in the offline orbitrap VOC analysis at SeaSCAPE. All measurements are made at 25 °C

and 978 mbar, and the bubbler is submerged in a water bath to maintain a constant temperature. Gas phase benzothiazole is not observed above detection limits (of 10 ppt) in any replicates. α-pinene and limonene are observed in the gas phase in significant quantities, though not equal to quantities that would indicate equilibrium partitioning. We therefore have insufficient evidence to propose significant differences between true and structurally predicted Henry's constants for benzothiazole as the driving mechanism behind the unexpectedly high benzothiazole volatilization observed at SeaSCAPE.

4.SI.11: Benzothiazole GCxGC Quantification and Uncertainties- Aerosol and DOM

Aerosol and DOM proxies and uncertainties are described in table S2 below.

Observed Benzothiazole	Quantification Proxy	Proxy Assignment Justification	Aerosol Quantification Uncertainty (%)	DOM Quantification Uncertainly (%)*
Benzothiazole	Benzothiazole	Authentic Match	10	60[A]
2-Phenyl Benzothiazole	2-Phenyl Benzothiazole	Authentic Match	10	30
2-(Methylthio) benzothiazole	2-(Methylthio) benzothiazole	Authentic Match	10	50[B]
Butyl-Benzothiazole	2-(Methylthio) benzothiazole	Closest in GCxGC; most chemically similar	50	60
Alk-Benzothiazole 1	2-(Methylthio) benzothiazole	Closest in GCxGC; most chemically similar	50	
Alk-Benzothiazole 2	2-(Methylthio) benzothiazole	Closest in GCxGC; most chemically similar	50	
Alk-Benzothiazole 3	2-(Methylthio) benzothiazole	Closest in GCxGC; most chemically similar	50	
Alk-Benzothiazole 4	2-(Methylthio) benzothiazole	Closest in GCxGC; most chemically similar	50	

Alk-Benzothiazole 5	2-(Methylthio) benzothiazole	Closest in GCxGC; most chemically similar	50	
2-o-Tolylbenzothiazole	2-Phenyl Benzothiazole	Closest in GCxGC; most chemically similar	50	
2(3H)-Benzothiazolone	2-(Methylthio) benzothiazole	Closest in GCxGC	60	60
3-Ethylbenzothiazolin-2-thione	2-Phenyl Benzothiazole	Closest in GCxGC	60	
3-Methyl-3H-benzothiazol-2-one	2-(Methylthio) benzothiazole	Closest in GCxGC	60	60
2-Benzothiazolamine, N-ethyl-	2-(Methylthio) benzothiazole	Closest in GCxGC	60	

Table S2. Benzothiazole GCxGC quantification proxies and uncertainties for all observed aerosol phase species

A) DOM phase benzothiazole quantification is subject to unusually high uncertainties despite the availability of authenticated standard calibration curves due to a significant difference between the volatility distributions of the DOM phase sample matrix and the standard mix, leading to extreme matrix effects in the volatile region of the GCxGC chromatogram. This is verified by order of magnitude increased recoveries of the most volatile internal standards in DOM sample analysis and does not influence the aerosol phase measurements.

B) Measured 2-(Methylthio)benzothiazole in the pre-transport sample exceeds the maximum point of the calibration curves and is quantified through extrapolation of a quadratic calibration fit to account for observed exponential behavior at the high end of the 2-methylthiobenzothiazole internal standard normalized calibration curve. This increases the quantification uncertainty of this species. All other measured benzothiazole species for which 2-methylthiobenzothiazole serves as a quantification proxy lie within the linear lower bounds of the calibration curve for this species and are normalized by a linear quantification factor.

*Quantification certainties for DOM phase species are generally reduced due to the increased number of preparation steps, and volatility distribution discrepancies between samples and standards coupled with the novelty of TD-GCxGC application to this sample medium. This increased uncertainty applies to absolute quantification accuracy

uncertainties but would be expected to apply comparably across individual compounds, and therefore does not apply to relative distribution precision.

4.SI.12: Benzothiazole GCxGC Quantification and Uncertainties- Gas

Gas phase benzothiazole is positively identified from exact GCxGC position and mass spectral match to an authentic benzothiazole standard (96% purity, Sigma Aldrich). A liquid benzothiazole standard dissolved in methanol is introduced to sorbent tubes through a constant flow of ultrapure dry nitrogen using a Calibration Solution Loading Rig (CSLR, Markes) in five different mass loading levels, and these standard tubes are run in series to establish the ratio between mass of benzothiazole introduced to the VOC sorbent tube and detected analyte volume in GCxGC, also known as the quantification factor. Uncertainty is estimated at 12% from the geometric mean of the percent differences between the quantification factor- predicted instrument volume at each calibration point and the observed instrument volumes at those points.

4.SI.13: Cartoon Visualization of Benzothiazole Dynamics in a Coastal Marine Context

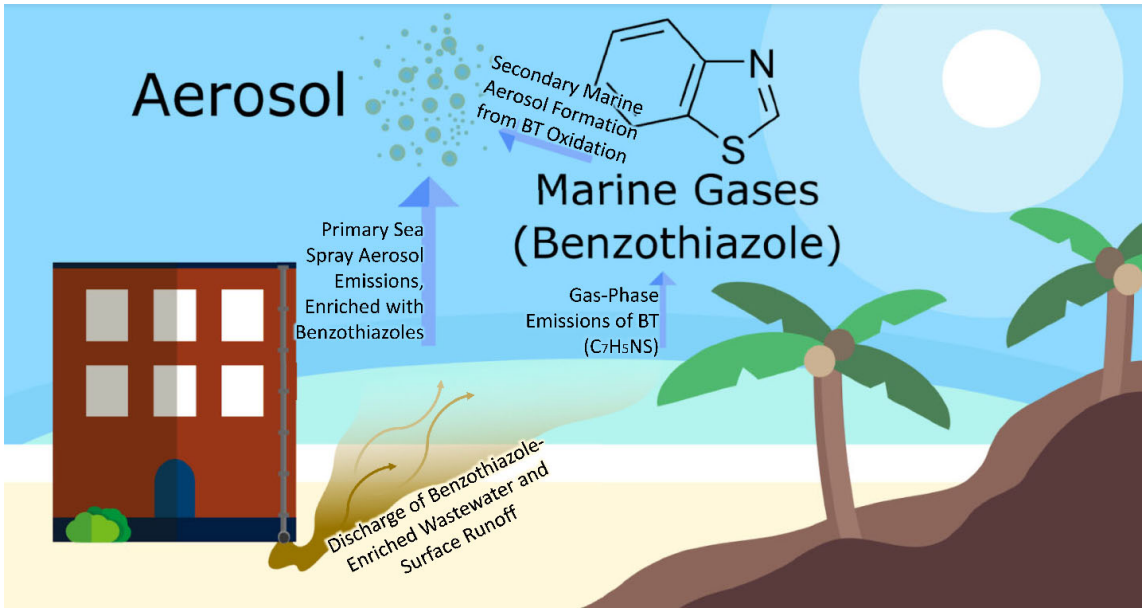
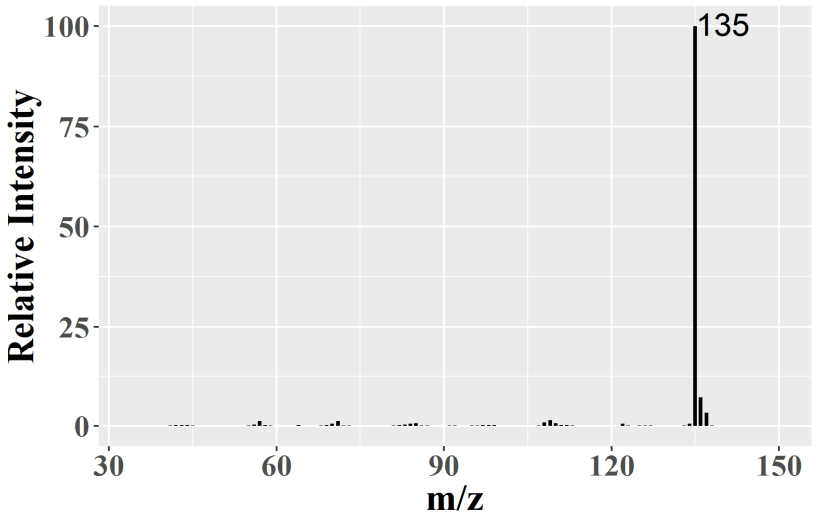


Figure 4.S1. Annotated Table of Contents graphic illustrating transport and transformation processes leading to the presence of anthropogenic benzothiazoles in coastal marine atmospheres. Anthropogenic benzothiazoles are discharged into coastal waters, transported into the atmosphere through primary sea spray aerosol emissions and volatilization, and transformed through atmospheric oxidation.

4.SI.14: Supporting Figures- SeaSCAPE Measurements and Context

A)



B)

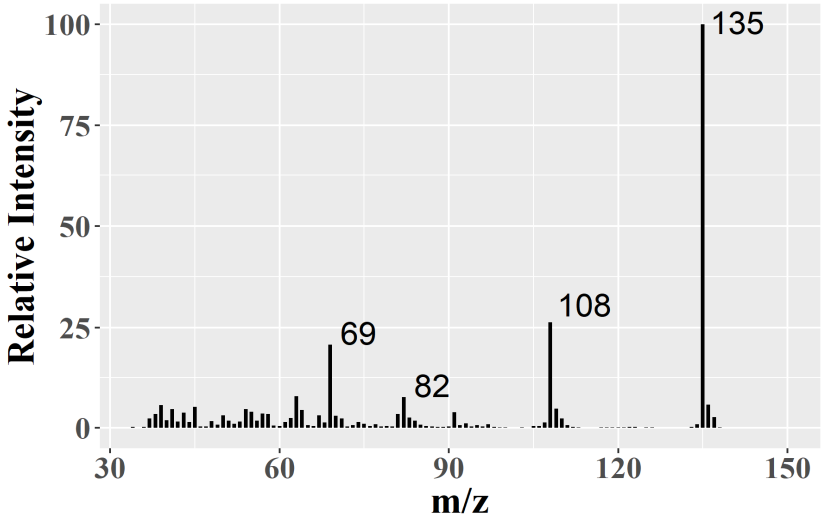


Figure 4.S2. A) 14 eV EI spectrum of benzothiazole as identified by TD-GCxGC in samples at SeaSCAPE; B) 70 eV EI spectrum of benzothiazole as identified by TD-GCxGC in samples at SeaSCAPE

Time Series of Representative Gases at SeaScape

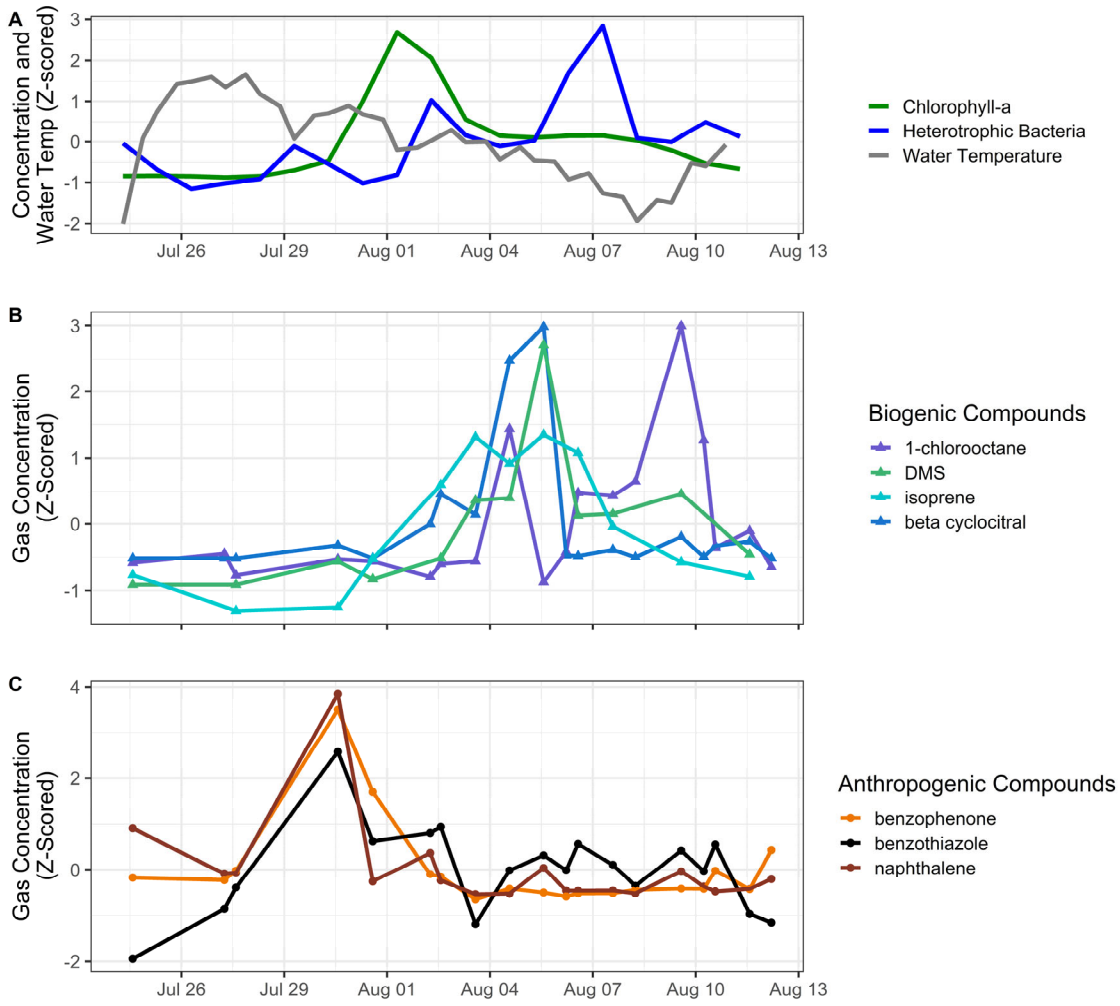


Figure 4.S3. A) Z-scored water concentrations of biological activity indicators (Chlorophyll-a and heterotrophic bacteria) and z-scored wave channel water temperature; B) Z-scored gas phase time series of biogenic VOCs observed in the SeaSCAPE ISV; C) Z-scored gas phase time series of anthropogenic VOCs observed in SeaSCAPE ISV. All gas concentrations from GCxGC measurements, with the exception of DMS (supplied by Vocus) and isoprene (supplied by benzene CIMS, as described in Sauer and Mayer et al. 2021). Note that the increase in this population of anthropogenic compounds after initial water introduction is unexpected but consistently observed across instruments and does not correspond to any perturbation that could have introduced additional anthropogenic organic material into the wave channel.

Quantitative Comparison, VOCUS and GCxGC VOC measurements

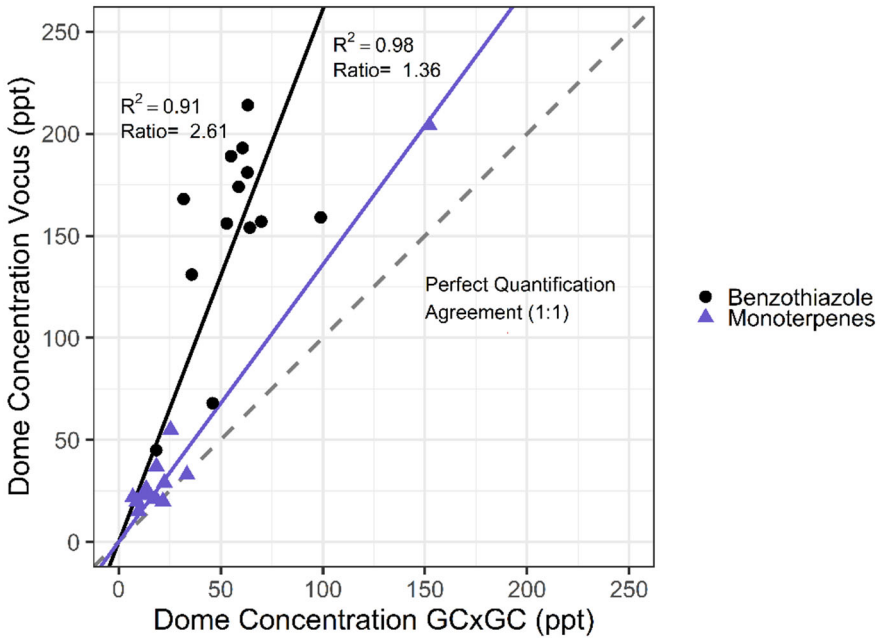


Figure 4.S4. Quantitative comparison of gasses measured by both VOCUS and GCxGC in the ISV at SeaSCAPE bloom 3.

Oceanside Outfall and Currents Near Scripps Pier

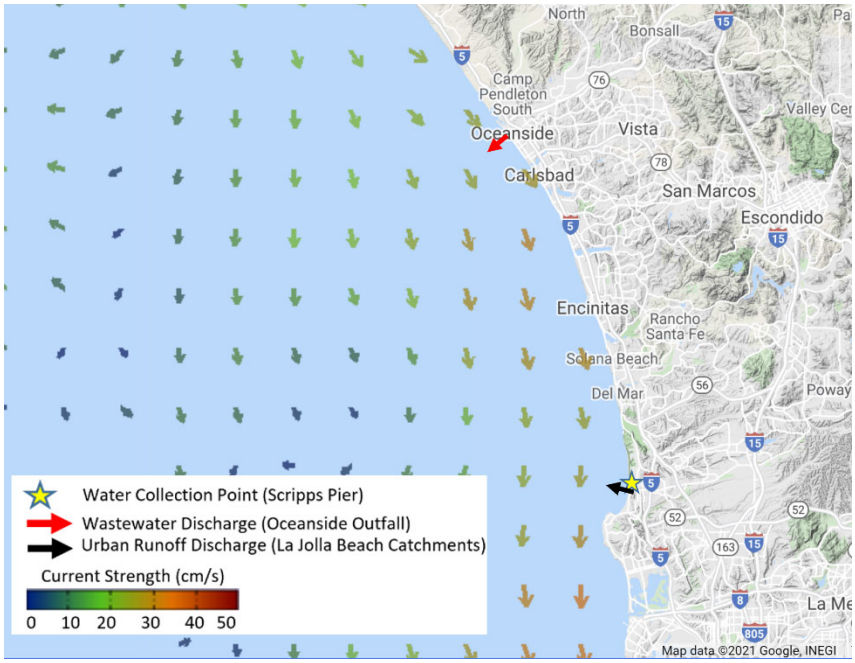


Figure 4.S5. Currents and locations of water sampling and Oceanside outfall on SeaSCAPE bloom 3 water collection day, July 23 2019. Current graphics from Southern California Coastal Ocean Observation System high frequency radar system website (https://sccoos.org/high-frequency-radar/)

Wastewater and Municipal Storm Water Discharges near Scripps Pier

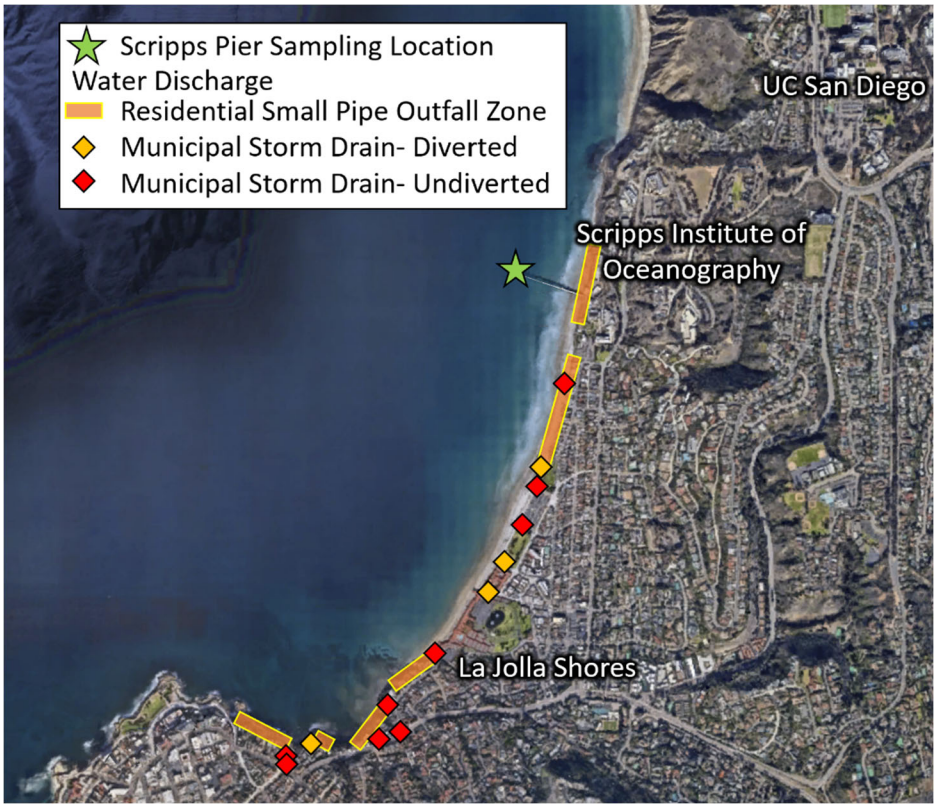


Figure 4.S6. Map of La Jolla coast with locations of SeaSCAPE water sampling and Scripps Pier VOC sampling, residential small pipe outfall zones, and municipal storm drain outlets highlighted. Locations of water discharge points from Collins et al. 2008,[6] underlying mapping from Google Earth.

Individual Time Series of Aerosol Phase Benzothiazoles at SeaSCAPE

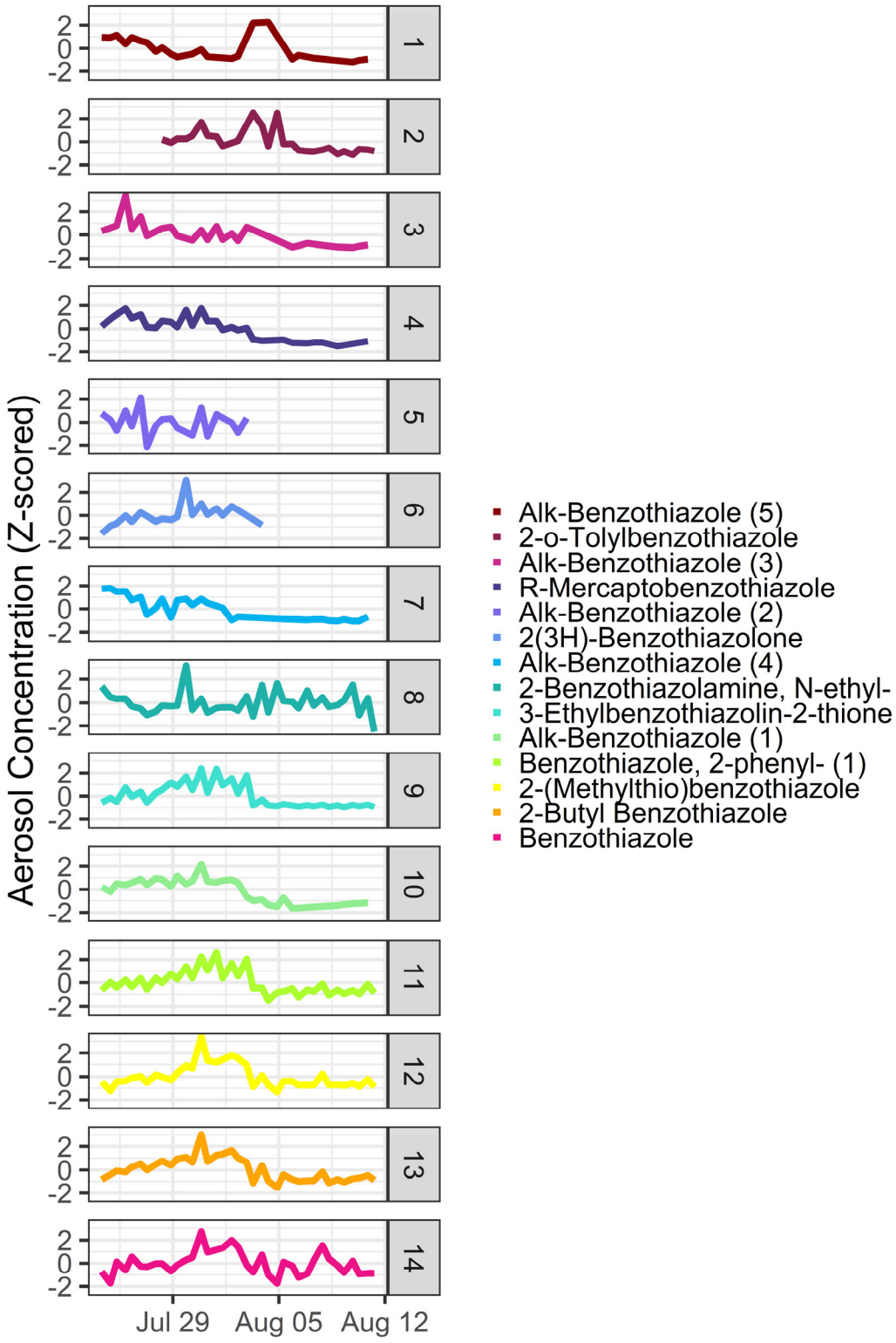


Figure 4.S7. Concentration time series of aerosol-phase benzothiazoles observed at SeaSCAPE, listed in order of increasing contributions to the total benzothiazole carbon pool

Sea Spray Aerosol D7 Siloxane and Benzothiazole Carbon Pool Time Series at SeaSCAPE

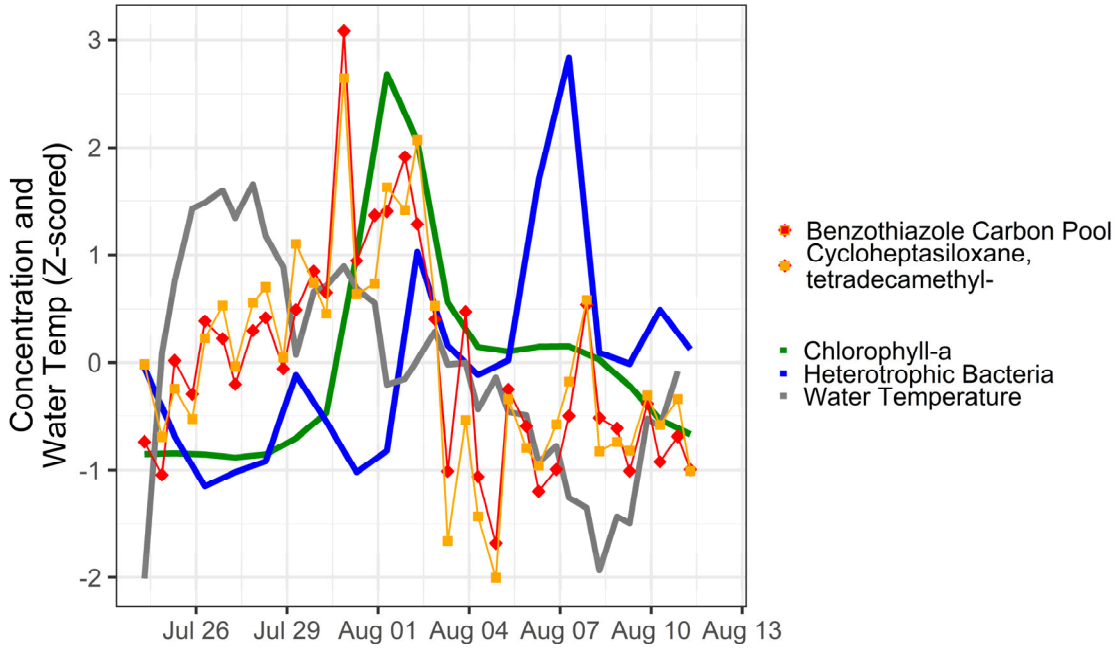


Figure 4.S8. Aerosol-phase benzothiazole carbon pool (atmospheric concentration, red) time series in comparison to biological activity indicators (chlorophyll-a water concentration, green, and heterotrophic bacteria water concentration, blue) and abundant anthropogenic organic compound tetradecamethyl-cycloheptasiloxane

Fractional Changes in the Sea Spray Aerosol Benzothiazole Carbon Pool at SeaSCAPE

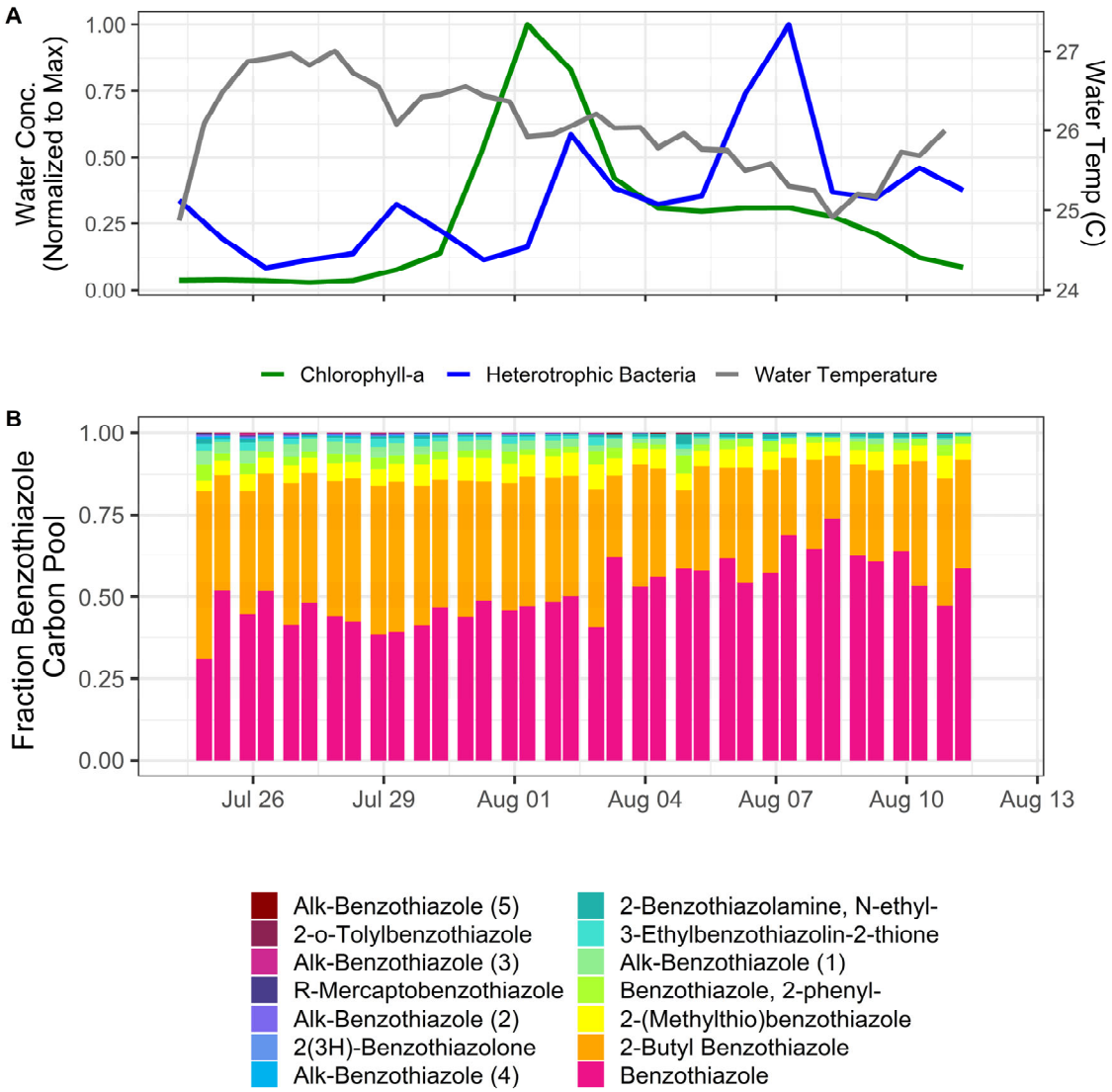


Figure 4.S9. Time series of fractional contributions of observed benzothiazoles to total benzothiazole carbon pool.

Enrichment Analysis of Common Benzothiazoles in SSA and DOM Phases

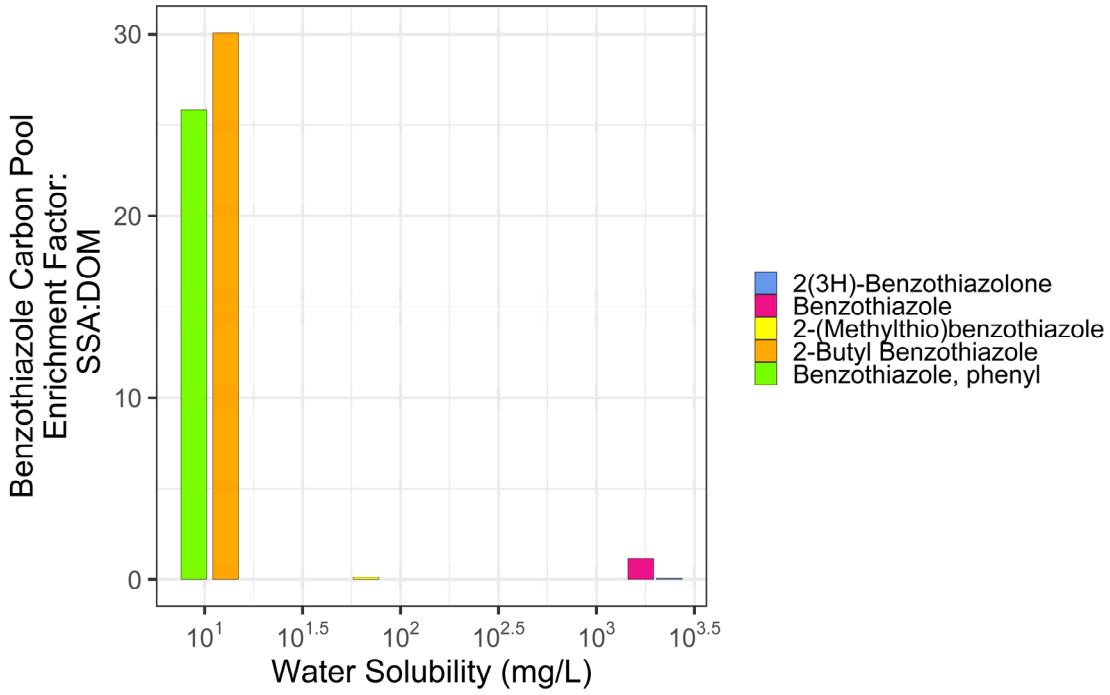


Figure 4.S10. Ratio of fractional contributions to the total benzothiazole carbon pool in the aerosol (SSA) phase compared to the dissolved (DOM) phase for benzothiazole compounds observed in both phases.

4.SI.15. Literature Review of Coastal Benzothiazole Observations

Concentration	Source Type	Reference
BT: .85 ug/L MTBT: .17	Wastewater treatment influent, Berlin 1a	Kloepfer 2005[7]
BT: .55 ug/L MTBT: .44	Wastewater treatment effluent Berlin 1a	
BT: 2.26 ug/L MTBT: .55	Wastewater treatment effluent Beijing	
Summed benzothiazoles: 74-25 ug/L	Street runoff	
Household wastewater summed benzothiazoles (~40% of which is BT): 1.7-2.2 ug/L		
BT: 1.21 ug/L	Moshassuck River (urban runoff)	Reddy 1997[8]
BT: .378 ug/L	Pawtuxet River (urban runoff) (Feb)	
BT:.940 ug/L	Pawtuxet River (urban runoff) (May)	
BT: .819 ug/L	Pawtuxet River (urban runoff) (Sept)	
BT: 10.5 ug/L MTBT: 39 ug/L	Untreated tannery wastewater	Fiehn 1994[9]
BT: 99 ug/L MTBT: 115.5 ug/L	Anaerobic treated tannery (wastewater)	
BT: 5.5 ug/L MTBT: 24.5 ug/L	Aerobic treated tannery (wastewater)	
BT: .396 MTBT: .119	Dissolved, in Ballona creek (stormwater runoff)	Zeng 2004[10]
BT: .06-2500 ug/L MTBT: .01-650 ug/L	Surface waters from estuary in India	Liao 2018[11]

Table 4.S3. Literature review of reported dissolved phase benzothiazoles in aquatic environments.

4.SI.16. Benzothiazole Oxidation Experiment Additional Information

As previously described in Kang et al. , the PAM-OFR generates OH radicals using UV lights at 185 and 254 nm.[12] The calibration for OH exposure was performed by introducing CO (1 ppm) into the PAM-OFR and measuring the decay in CO concentrations as a function of lamp voltage (results illustrated in figure S11). All described oxidation reactions are unseeded and therefore focus on particle nucleation rather than condensation onto seed aerosols within the OFR.[13,14,15,16] BT was introduced to the PAM-OFR by either flowing 2.5 LPM of zero air over a BT permeation tube (VICI) heated to 37 °C or introducing an instant plug injection of 62 ug of BT (96% Sigma Aldrich) dissolved in 5 uL of HPLC grade methanol (Fisher Scientific). In the plug experiment, the production of particles was characterized using a Scanning Mobility Particle Sizer (SMPS, TSI Incorporated) following a benzothiazole injection at 5 minutes. A 2.5 LPM zero air flow was added to obtain 5 LPM through the PAM-OFR. A separate injection of pure methanol showed no formation of aerosol mass. The relative humidity in the PAM-OFR was maintained between 50-70%. The aerosol flow produced after the PAM-OFR was not dried and thus contained water. Measuring produced SO2 from the BT oxidation was carried out by using a 43iQ Sulfur Dioxide Analyzer (ThermoFisher Scientific). Carulite (Carus Inc.) and a 0.2 µm Teflon filter (Hach Company, 47mm) was used to scrub ozone and aerosol from the stainless-steel sampling lines before entering the SO2 analyzer. During SO2 measurements, blanks were performed by either turning off the PAM UV lights (with BT) or by running the zero air through the PAM with lights kept on (without BT) and in both cases, SO2 levels were below ambient levels. In a continuous BT input oxidation experiment (using the permeation tube), the voltage on the PAM UV lights were changed to increase the equivalent days of aging by producing more OH radical. In this experiment, the voltage was increased to achieve an OH exposure of 3.8×10^{11} molecules cm-3s-1 (2.9 days equivalent ageing) then 4.6×10^{11} molecules cm-3 s-1 (3.5 days equivalent ageing), then 6.1×10^{11} molecules cm-3s-1 (4.8 days equivalent ageing). Particle production was not observed from introduction of BT into the PAM with lights off.

In addition to SO2, HSO4- was also measured in PAM oxidation products. HSO4- was observed in a sample collected from bubbling the downstream PAM aerosol line into HPLC grade acetonitrile (Fisher Scientific) and then analyzing the contents using a direct injection heated electrospray ionization (HESI) Orbitrap (ThermoFisher Scientific) mass spectrometer. The sample was injected at 5 µL/min and peaks were detected in negative mode with the capillary voltage set to 3.0 kV and capillary temperature set at 325 °C. The HESI gases were set to: sheath at 15, auxiliary at 5, and sweep at 0 (arbitrary units) and was calibrated before analysis using a negative mode calibration mix (Pierce ESI Ion Calibration Solutions, Thermo Fisher Scientific). The background subtracted mass spectra data can be found in Supporting Information Table S4.

Table 4.S4. Background subtracted BZT oxidation products from PAM sampling line, bubbled into ACN for 60 mins. Mass spectra collected using Orbitrap in negative mode, formula matched signals shown here are above 2% relative intensity.

Mass to charge ratio (m/z)	Intensity (counts)	Relative Intensity (to max signal)	Delta (ppm)	Composition $[M-H]^-$
83.02529	113643.3	54.36	2.41	C3 H3 O N2
109.04080	99471	47.59	0.67	C5 H5 O N2
96.96023	**74536.9**	**35.66**	**1.32**	**H O4 S**
122.03610	61766.7	29.55	1.21	C5 H4 O N3
124.05180	58808.5	28.13	1.08	C5 H6 O N3
120.0570	39885.8	19.08	1.93	C6 H6 N3
136.05180	34553.5	16.53	1.11	C6 H6 O N3
121.05220	31968.9	15.29	1.9	C5 H5 N4
132.0570	30341	14.51	1.71	C7 H6 N3
148.05180	28371	13.57	1.12	C7 H6 O N3
107.02520	27949.4	13.37	1.23	C5 H3 O N2
91.03033	24650.8	11.79	1.78	C5 H3 N2
92.02560	23525.3	11.25	1.99	C4 H2 N3
95.02524	22736.5	10.88	1.63	C4 H3 O N2
111.02010	21198.5	10.14	1.2	C4 H3 O2 N2
123.05650	21170.8	10.13	1.03	C6 H7 O N2
110.03620	20501	9.81	1.47	C4 H4 O N3
121.04090	19691.2	9.42	1.23	C6 H5 O N2
125.03580	19088.7	9.13	1.04	C5 H5 O2 N2
133.05230	18620.1	8.91	2.11	C6 H5 N4
163.06270	17925.2	8.58	1.11	C7 H7 O N4
93.00962	17713.3	8.47	1.97	C4 H O N2
108.02050	16848.6	8.06	1.08	C4 H2 O N3
130.04140	16252.8	7.78	2.59	C7 H4 N3
119.02530	15803.8	7.56	1.53	C6 H3 O N2
98.02489	15720.9	7.52	1.37	C4 H4 O2 N
134.03640	15519.6	7.42	3	C6 H4 O N3
105.04600	15297.7	7.32	2.05	C6 H5 N2
82.04126	14982.2	7.17	2.3	C3 H4 N3
97.04087	12930.6	6.19	1.38	C4 H5 O N2
82.03002	12848.3	6.15	2.25	C4 H4 O N
150.06750	12649.3	6.05	1.12	C7 H8 O N3
110.02490	12339.3	5.9	0.85	C5 H4 O2 N

138.06740	12336.3	5.9	1.07	C6 H8 O N3
133.04090	12327	5.9	1.36	C7 H5 O N2
116.02560	11853	5.67	1.85	C6 H2 N3
123.02020	11771.3	5.63	1.27	C5 H3 O2 N2
107.03670	11656.8	5.58	3.64	C4 H3 N4
137.03580	11259.8	5.39	1.11	C6 H5 O2 N2
114.01980	11237.8	5.38	1.08	C4 H4 O3 N
89.02457	11176	5.35	1.73	C3 H5 O3
118.04140	11134	5.33	2.5	C6 H4 N3
149.04710	11012.9	5.27	1.4	C6 H5 O N4
177.07830	10759.3	5.15	0.81	C8 H9 O N4
120.02060	10681	5.11	1.83	C5 H2 O N3
108.05680	9984.6	4.78	1.07	C5 H6 N3
99.02018	9696.8	4.64	1.77	C3 H3 O2 N2
113.03580	9548.8	4.57	0.99	C4 H5 O2 N2
96.00926	9355.4	4.48	1.66	C4 H2 O2 N
137.04710	8050.6	3.85	1.3	C5 H5 O N4
162.06750	7830.4	3.75	1.06	C8 H8 O N3
138.03110	7317.5	3.5	1.42	C5 H4 O2 N3
81.04601	7308.3	3.5	2.32	C4 H5 N2
152.04670	7268.5	3.48	1.24	C6 H6 O2 N3
159.06790	6833	3.27	2.04	C8 H7 N4
135.05650	6819	3.26	1.12	C7 H7 O N2
175.06280	6818.1	3.26	1.38	C8 H7 O N4
68.01439	6784.6	3.25	2.99	C3 H2 O N
109.00450	6711.4	3.21	1.21	C4 H O2 N2
160.05180	6486.3	3.1	1.3	C8 H6 O N3
96.05689	6339.8	3.03	1.73	C4 H6 N3
139.05150	6306.7	3.02	1.13	C6 H7 O2 N2
164.04670	6143.2	2.94	1.17	C7 H6 O2 N3
107.03820	6009.8	2.87	4.92	C6 H5 O N
150.03110	6006.5	2.87	1.4	C6 H4 O2 N3
86.02491	5973.8	2.86	1.82	C3 H4 O2 N
126.03110	5816.7	2.78	1.3	C4 H4 O2 N3
94.04125	5411.5	2.59	1.94	C4 H4 N3
140.04680	5383.6	2.58	1.48	C5 H6 O2 N3
112.04050	5143.5	2.46	0.85	C5 H6 O2 N
147.06800	5105.3	2.44	2.89	C7 H7 N4
134.07260	5034.4	2.41	1.35	C7 H8 N3

127.05140	4970	2.38	1.06	C5 H7 O2 N2
115.03030	4949.8	2.37	1.5	C7 H3 N2
144.05710	4836.7	2.31	2.34	C8 H6 N3
115.04020	4832.8	2.31	1.14	C5 H7 O3
146.07260	4761	2.28	1.67	C8 H8 N3
117.01950	4750.6	2.27	1.2	C4 H5 O4
104.02570	4742.3	2.27	2.41	C5 H2 N3
146.03630	4660.8	2.23	2.14	C7 H4 O N3
98.03613	4537.4	2.17	1.53	C3 H4 O N3
84.00926	4530.3	2.17	1.83	C3 H2 O2 N
94.02999	4472.8	2.14	1.61	C5 H4 O N

Figure 4.S11. SO_2 production from benzothiazole oxidation in PAM-OFR. Benzothiazole is introduced via custom permeation tube at a constant concentration of 12.8 ppb. Error bars indicate standard deviation of observed concentrations for each observation at the given equivalent ageing exposure setting (n = 20).

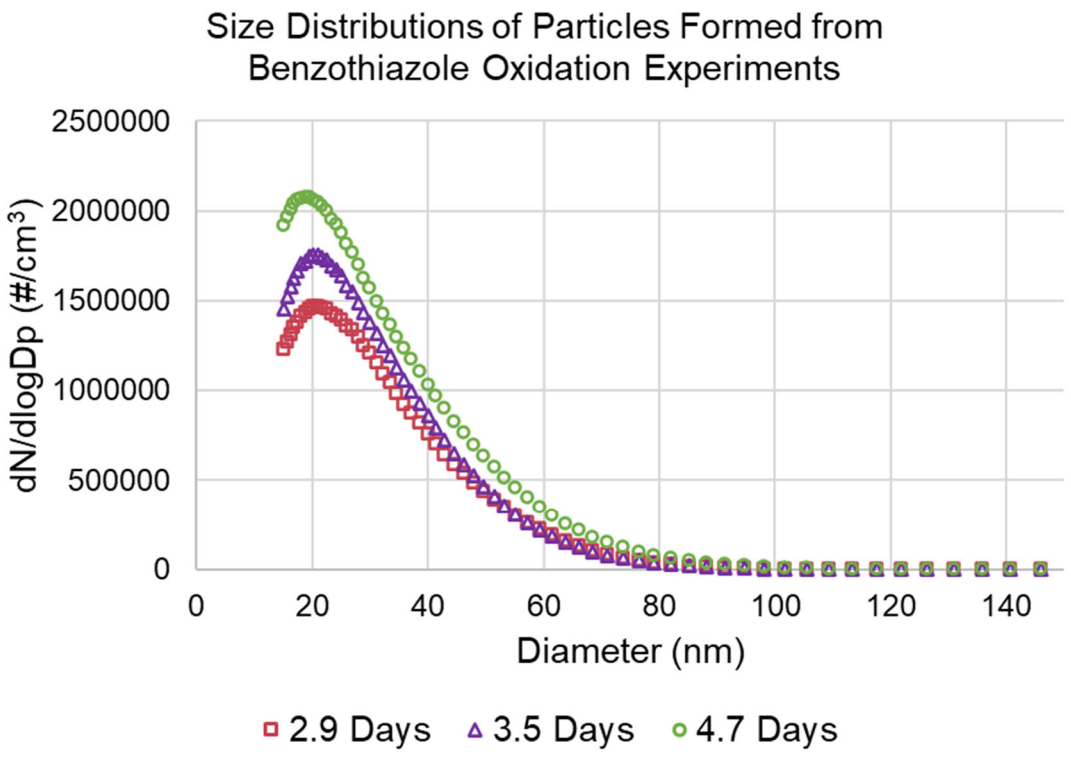


Figure 4.S12. Size distributions of aerosols formed from oxidation of benzothiazole in a PAM-OFR under OH exposure conditions simulating 2.9, 3.5, and 4.7 days of oxidative ageing.

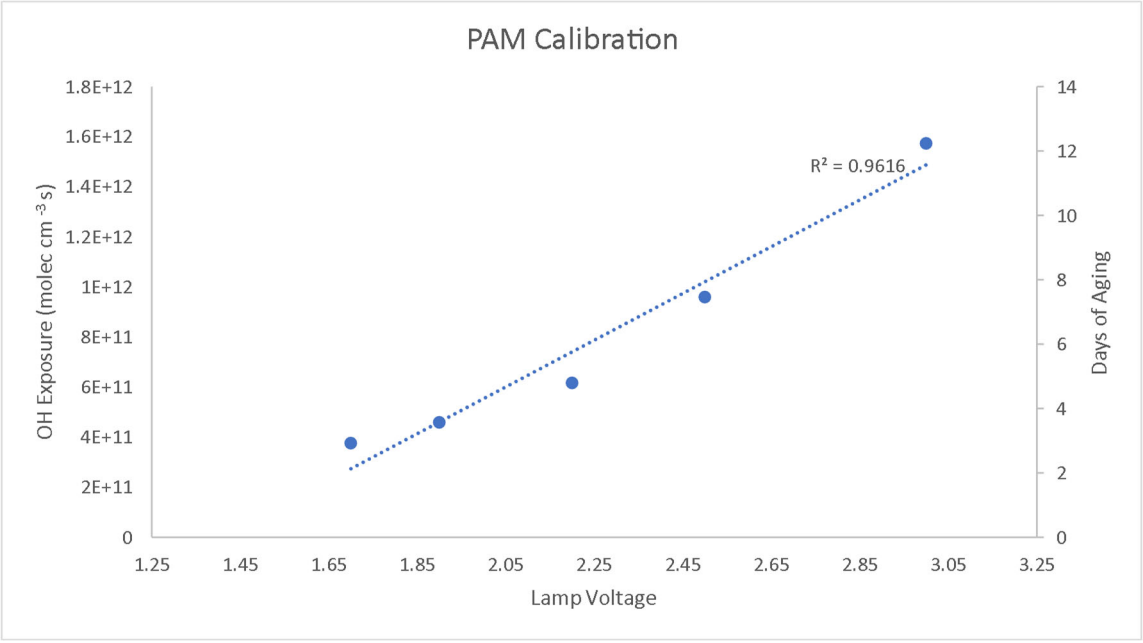


Figure 4.S13. PAM-OFR calibration curve determined by CO loss at increasing OH exposures.

Supporting Information References

(1) Yee, L. D.; Isaacman-Vanwertz, G.; Wernis, R. A.; Meng, M.; Rivera, V.; Kreisberg, N. M.; Hering, S. V; Bering, M. S.; Glasius, M.; Upshur, M. A.; Bé, A. G.; Thomson, R. J.; Geiger, F. M.; Offenberg, J. H.; Lewandowski, M.; Kourtchev, I.; Kalberer, M.; De Sá, S.; Martin, S. T.; Alexander, M. L.; Palm, B. B.; Hu, W.; Campuzano-Jost, P.; Day, D. A.; Jimenez, J. L.; Liu, Y.; Mckinney, K. A.; Artaxo, P.; Viegas, J.; Manzi, A.; Oliveira, M. B.; De Souza, R.; Machado, L. A. T.; Longo, K.; Goldstein, A. H. Observations of Sesquiterpenes and Their Oxidation Products in Central Amazonia during the Wet and Dry Seasons. *Atmos. Chem. Phys* **2018**, *18*, 10433–10457. https://doi.org/10.5194/acp-18-10433-2018.

(2) Sheu, R.; Marcotte, A.; Khare, P.; Charan, S.; Ditto, J. C.; Gentner, D. R. Advances in Offline Approaches for Chemically Speciated Measurements of Trace Gas-Phase Organic Compounds via Adsorbent Tubes in an Integrated Sampling-to-Analysis System. *J. Chromatogr. A* **2018**, *1575*, 80–90. https://doi.org/10.1016/j.chroma.2018.09.014.

(3) Dittmar, T.; Koch, B.; Hertkorn, N.; Kattner, G. A Simple and Efficient Method for the Solid-Phase Extraction of Dissolved Organic Matter (SPE-DOM) from Seawater. *Limnol. Oceanogr. Methods* **2008**, *6* (6), 230–235. https://doi.org/10.4319/lom.2008.6.230.

(4) Worton, D. R.; Decker, M.; Isaacman-VanWertz, G.; Chan, A. W. H.; Wilson, K. R.; Goldstein, A. H. Improved Molecular Level Identification of Organic Compounds Using Comprehensive Two-Dimensional Chromatography, Dual Ionization Energies and High Resolution Mass Spectrometry. *Analyst* **2017**, *142* (13), 2395–2403. https://doi.org/10.1039/c7an00625j.

(5) Lee, S.-H.; Mukherjee, S.; Brewer, B.; Ryan, R.; Yu, H.; Gangoda, M. A Laboratory Experiment To Measure Henry's Law Constants of Volatile Organic Compounds with a Bubble Column and a Gas Chromatography Flame Ionization Detector (GC-FID). *J. Chem. Educ.* **2012**, *90* (4), 495–499. https://doi.org/10.1021/ED200303X.

(6) Collins, T.; Hazard, L.; Hodgkiss, B.; Lawrence, S.; Parnell, E.; Terrill, E.; Varond, S.; Wilson, N.; Fabian, A.; Hampel, J.; O'Connell, K. *The La Jolla Shores Coastal Watershed Management Plan Final Report*; La Jolla, CA, 2008.

(7) Kloepfer, A.; Jekel, M.; Reemtsma, T. Occurrence, Sources, and Fate of Benzothiazoles in Municipal Wastewater Treatment Plants. *Environ. Sci. Technol.* **2005**, *39* (10), 3792–3798. https://doi.org/10.1021/es048141e.

(8) Reddy, C. M.; Quinn, J. G. Environmental Chemistry of Benzothiazeles Derived from Rubber. *Environ. Sci. Technol.* **1997**, *31* (10), 2847–2853. https://doi.org/10.1021/es970078o.

(9) Fiehn, O.; Reemtsma, T.; Jekel, M. Extraction and Analysis of Various Benzothiazoles from Industrial Wastewater. *Anal. Chim. Acta* **1994**, *295* (3), 297–305. https://doi.org/10.1016/0003-2670(94)80235-1.

(10) Zeng, E. Y.; Tran, K.; Young, D. Evaluation of Potential Molecular Markers for Urban Stormwater Runoff. *Environ. Monit. Assess.* **2004**, *90* (1–3), 23–43. https://doi.org/10.1023/B:EMAS.0000003564.24169.86.

(11) Liao, C.; Kim, U. J.; Kannan, K. A Review of Environmental Occurrence, Fate, Exposure, and Toxicity of Benzothiazoles. *Environ. Sci. Technol.* **2018**, *52* (9), 5007–5026. https://doi.org/10.1021/acs.est.7b05493.

(12) Kang, E.; Root, M. J.; Toohey, D. W.; Brune, W. H. Introducing the Concept of Potential Aerosol Mass (PAM). *Atmos. Chem. Phys.* **2007**, *7* (22), 5727–5744. https://doi.org/10.5194/acp-7-5727-2007.

(13) Zhao, R.; Zhang, Q.; Xu, X.; Zhao, W.; Yu, H.; Wang, W.; Zhang, Y.; Zhang, W. Effect of Experimental Conditions on Secondary Organic Aerosol Formation in an Oxidation Flow Reactor. *Atmos. Pollut. Res.* **2021**, *12* (3), 205–213. https://doi.org/10.1016/J.APR.2021.01.011.

(14) Ahlberg, E.; Falk, J.; Eriksson, A.; Holst, T.; Brune, W. H.; Kristensson, A.; Roldin, P.; Svenningsson, B. Secondary Organic Aerosol from VOC Mixtures in an Oxidation Flow Reactor. *Atmos. Environ.* **2017**, *161*, 210–220. https://doi.org/10.1016/J.ATMOSENV.2017.05.005.

(15) Lambe, A. T.; Chhabra, P. S.; Onasch, T. B.; Brune, W. H.; Hunter, J. F.; Kroll, J. H.; Cummings, M. J.; Brogan, J. F.; Parmar, Y.; Worsnop, D. R.; Kolb, C. E.; Davidovits, P. Effect of Oxidant Concentration, Exposure Time, and Seed Particles on Secondary Organic Aerosol Chemical Composition and Yield. *Atmos. Chem. Phys.* **2015**, *15* (6), 3063–3075. https://doi.org/10.5194/ACP-15-3063-2015.

(16) Palm, B. B.; Campuzano-Jost, P.; Ortega, A. M.; Day, D. A.; Kaser, L.; Jud, W.; Karl, T.; Hansel, A.; Hunter, J. F.; Cross, E. S.; Kroll, J. H.; Peng, Z.; Brune, W. H.; Jimenez, J. L. In Situ Secondary Organic Aerosol Formation from Ambient Pine Forest Air Using an Oxidation Flow Reactor. *Atmos. Chem. Phys.* **2016**, *16* (5), 2943–2970. https://doi.org/10.5194/ACP-16-2943-2016.

5 Chemical Signatures of Seasonally Unique Anthropogenic Influences on Organic Aerosol Composition in the Central Amazon

.

This work is adapted from:

Emily B. Franklin, Lindsay D. Yee, Rebecca Wernis, Gabriel Isaacman-VanWertz, Nathan Kreisberg, Robert Weber, Suzane de Sá, Brett Palm, Weiwei Hu, Pedro Campuzano-Jost, Douglas A. Day, Antonio Manzi, Paulo Artaxo, Rodrigo A. F. De Souza, Jose Jimenez, Scot Martin, Allen H. Goldstein, "Chemical Signatures of Seasonally Unique Anthropogenic Influences on Organic Aerosol Composition in the Central Amazon." This work is currently undergoing review with co-authors and is intended for submission to Atmospheric Chemistry and Physics in July 2022.

5.1 Abstract

Urbanization and human caused fires perturb the quantities and composition of fine organic aerosol material in the central Amazon, with significant ramifications for both radiative forcing and public health. These disturbances include not only direct emissions of primary aerosol material and precursors of secondary organic aerosol (SOA) from cities and fires, but also changes in the oxidation chemistry through which biogenic precursors form SOA. While SOA formation in the tropics has been observed and modelled to increase in the presence of anthropogenic pollutants, the mechanisms by which these enhancements occur and implications for particle chemistry remain incompletely characterized. During the Green Ocean Amazon (GoAmazon) field campaign of 2014/5, submicron aerosol samples were collected at the "T3" Manacapuru rural site, located 70km downwind of urban Manaus. These samples were analysed for speciated organic composition using TD-GCxGC-EI-HR-ToF-MS (thermal desorption two-dimensional gas chromatography with electron ionization and vacuum ultraviolet time of flight mass spectrometry), which yielded a library of ~1300 unique organic compounds which were traced across both seasons. Compounds were then grouped based on similar temporal variability, and anthropogenically influenced compound clusters were extracted from background measurements. These chemical signatures are compared between influence categories (urban plume vs. biomass burning) and between seasons (wet season vs. dry season) to establish season specific and source specific chemical profiles of influence. Between seasons, urban conditions produce consistently low carbon number high carbon oxidation state products compared to pristine and burning influenced conditions. Dry season products associated with both background and burning conditions are significantly more oxidized than their wet season counterparts, but include diverse high carbon number products not typically predicted or reported as biogenic or biomass burning products. There is a high degree of season-to-season aerosol composition variability, with only 52% of compounds traced present above detection limits in both seasons. Wet season background

conditions produced a range of high carbon number relatively low carbon oxidation state products that were never observed in the dry season. The dry season produced unique highly oxidized high carbon number products that were exclusively observed in periods of particularly intense biomass burning activity. Both of these populations may point to the importance of aqueous processing in Amazonian aerosol ageing processes, but further mechanistic insights are impeded by the limited degree of knowledge regarding product identities. Less than 12% of speciated and traced organic products are identifiable at an isomer specific level through matching to the NIST/EPA/NIH mass spectral database, and nearly all unique products observed were not identifiable. On a mass basis, unidentifiable organics are particularly important during strongly urban influenced periods of the wet season, when the fraction of GCxGC recovered organic mass attributed to unidentifiable organics rose to above 60%. These findings compositionally characterize anthropogenic influence on submicron organic aerosol in the tropics, identify key season to season differences in production signatures, and highlight high priority knowledge gaps in current speciated knowledge of ambient aerosol composition that are critical to improving understanding of how human emissions impact fine aerosol composition in the tropics.

5.2 Introduction

Human activities influence the composition and properties of ambient fine aerosol material under a broad array of conditions, with effects that are heavily dependent upon both the location and types of disturbances caused by humans and the composition and quantities of local natural emission sources. In the central Amazon, two broad classes of human activities have been shown to perturb the quantities and characteristics of fine aerosols: urbanization, and biomass burning.[1–4] These perturbations are important, because both have been demonstrated to significantly influence the radiative forcing,[2,5,6] and biomass burning during the Amazonian dry season causes $PM_{2.5}$ levels to frequently exceed WHO guidelines and has been associated with adverse public health outcomes for residents of the central Amazon.[7,8] In the central Amazon, fires are nearly entirely anthropogenic in nature. The high humidity frequent rainfall conditions make natural sources of fires extremely infrequent and unlikely, and fires are frequently set to clear land for soybean cultivation and cattle farming.[7,9,10] While fires occur throughout the year, burning activity is most intense during the dry season, which lasts from approximately August to October.[3] In addition to experiencing elevated biomass burning emissions, atmospheric residence times during the dry season are longer, which together contribute to dry season aerosol concentrations that are approximately an order of magnitude higher than the wet season, with a more oxidized bulk composition.[3,4,11]

The Amazonian wet season, which typically lasts from late January to mid April, is characterized by frequent rain, less significant biomass burning activity, and extremely low particle concentrations.[4,12] Because of the reduced influence from fires, perturbations from the more interseasonally consistent source of urban areas become more important during the wet season. In regions downwind of cities, urban emissions produce substantial changes in the composition and quantities of organic aerosol.[4] The enhancement of organic material in Amazonian regions downwind of cities, and specifically the city of Manaus, has been

the subject of intensive study through varying combinations of models and observations.[1,4,11,13–19] While some of the increased can be attributed to primary organic emissions from the city and formation of secondary aerosol material from anthropogenic organics,[17] the formation of SOA from biogenic volatile organic compounds (BVOCs) are also found to increase under influence from the urban plume and thereby contribute to elevated aerosol formation. The mechanisms driving these enhancements are incompletely characterized, and recent efforts to reproduce the enhancements using explicit molecular based models were not able to reproduce observed enhancements, although parameterized approaches have been more successful.[1,13] This indicates that our current understandings of how urban emissions impact biogenic SOA formation at the compound-specific level do not adequately capture real ambient dynamics.

The organic composition of the atmosphere is extremely complex, with the number of potential secondary and tertiary products that can theoretically be produced from well characterized reactive gas-phase precursors extending into the millions.[20,21] This complexity makes identifying key reaction pathways, such as those producing urban influenced biogenic SOA enhancement in the Amazon, highly challenging. Controlled laboratory experiments provide opportunities to better understand the complexity of ambient conditions by simplifying systems down to individual components which can enable the identification of key tracers of influence from a given precursor, but these controlled experiments are often conducted under conditions that are extremely dissimilar from reality. As reported in Porter et al., 2021, the vast majority of chamber oxidation experiments are carried out under conditions that are far drier, more concentrated, and more NOx impacted than the environments in which the vast majority of earth's SOA is being formed, including the central Amazon. Aqueous processing, which occurs when semivolatile organics partition into droplets and undergo aqueous-phase chemical reactions, are a potentially important mechanism controlling the composition of secondary organic material, particularly under humid conditions,[23] but the mechanisms by which aqueous processing alters the chemical fate of important precursors and the effects under ambient conditions in the Amazon remain incompletely characterized.

The present study investigates the organic composition of submicron aerosol at a speciated isomer-specific level to achieve the following goals: first, to identify the compositional fingerprints of anthropogenic influence on aerosol composition and compare these influences across seasons, and second, to identify critical knowledge gaps between compounds catalogued in mass spectral databases and real ambient aerosol organic material produced under unique conditions that will when fully characterized add important mechanistic insights to understandings of anthropogenic perturbations of biogenic secondary aerosols.

5.3 Methods

5.3.1 Green Ocean Amazon (GoAmazon 14/15) Field Campaign

5.3.1.1 Site Description

The Observations and Modelling of the Green Ocean Amazon (GoAmazon) experiment was an intensive multisite meteorology and atmospheric chemistry-focused field campaign carried out in the vicinity of Manaus, Brazil, located in the central Amazon. Measurements, including both flights and fixed site measurements, were conducted over the course of 2014 and 2015. The broader objectives of the campaign, descriptions of all fixed measurement sites, and meteorological context are discussed in detail in Martin et al., 2016. This work focuses on observations from the T3 field site, located 70 km from the city of Manaus and occasionally downwind of the city, depending on local meteorology and wind direction. This site was also periodically impacted by fires, which are primarily attributable to anthropogenic deforestation.[2,12] The measurements and analysis described below were restricted to two intensive operating periods (IOPs), during which intensive analysis by complementary specialized instrumentation suites was coordinated. The first of these (IOP1) was conducted from February 1 to March 31 2014 and coincided with the Amazon's wet season. The Amazon wet season is characterized by low particle backgrounds, frequent rain, and limited impact from fire activity. The second intensive operating period (IOP2) was conducted from August 15 to October 15 2014, coinciding with the dry season. In the dry season, rain is less frequent, particle backgrounds are higher, and anthropogenic burning activity is significantly enhanced.[2,9,10]

5.3.1.2 Filter Sample Collection and Selection

During both IOP's, submicron aerosol material was collected on filters (Pallflex Tissuequartz) using a custom designed sequential sampler (Aerosol Devices, Inc.) which excluded supermicron particles using a greaseless cyclone (BGI). Due to the high humidity conditions, reducing the relative humidity of sampled air was necessary to prevent water condensation onto the filter material. This was achieved by sampling air through 2.6 m of 2 cm ID copper tubing which was maintained below the dew point temperature of the trailer housing the sequential sampler, allowing excess water to be removed. The water condenser also reduced potential artifacts from I/SVOC condensation onto filter material, as semivolatiles preferentially condensed into the water trap and were removed. This sampler was previously described in Yee et al., 2018, which provides additional design details. Field blanks were collected in each filter holder of the sequential sampler weekly. Samples were immediately frozen following collection and were maintained frozen prior to analysis. During the wet season, filters were collected with approximately 12 hour time resolution, with "day" filters collected between 6:00 and 18:00 local time and "night" filters collected between 18:00 and 6:00 local time, with non-collection periods for filter replacement lasting not more than 15 minutes between samples. During the dry season, filters were collected with 4 hour time resolution. "Pre-dawn" filters were collected between 2:00 and 6:00, "morning" filters were collected between 6:00 and 10:00, "mid-day" filters were collected between 10:00 and 14:00, "afternoon" filters were collected between 14:00 and 18:00, "evening" filters were collected between 18:00 and 22:00, and "mid-night" filters

were collected between 22:00 and 2:00. Over 200 total filters were collected during both IOP's. A selection of samples from each season were selected to coincide with data collection periods from key complementary instruments (described in "Supporting Measurements"), known periods of interesting atmospheric conditions including deposition of biomass burning material from Africa, and days that have been the subject of previous observational and observation-model comparison.[3,4,11,13,14,16,18] In total, 54 samples were selected from the wet season, spanning 27 days of measurements. Selected samples spanned from February 8th to March 20th, and included a 15-day continuous measurement period from March 6th to March 20th. 129 samples were selected from the dry season, equating to 21.5 days of dry season coverage. Selected samples fell between September 8th and October 11th and included a 13-day continuous analysis period from September 18th to October 1st.

5.3.2 Sample Analysis

5.3.2.1 Offline sample analysis by thermal desorption two-dimensional gas chromatography coupled with high resolution time of flight mass spectrometry (TD-GCxGC-EI-HR-ToF-MS)

All selected samples from both IOP's of the GoAmazon campaign were analyzed for speciated organic composition using thermal desorption two-dimensional gas chromatography coupled with high resolution time of flight mass spectrometry (TD-GCxGC-EI-HR-ToF-MS, hereafter abbreviated GCxGC). The instrumental configuration used in this work, including subcomponent manufacturers, column materials, and thermal methods is described in Franklin et al., 2021, and a complete methodological analysis of the instrument discussed in Worton et al., 2017.[26] Briefly, this instrument functions as follows. First, aliquots of filter material of standardized areas were doped with an internal standard solution (described in Franklin et al., 2022) and introduced into the thermal desorption system. Using information about total organic mass loading during filter collection times from supplementary measurements by the AMS (see "Supporting Measurements"), the number of aliquots run was adjusted to maintain a relatively consistent total organic mass loading into the instrument. The approach of varying filter area by loading was chosen because the filter collection time averaged organic aerosol concentrations at the study site varied over orders of magnitude, and a fixed sample size approach would have rendered the organic composition of biogenic background or pristine samples incompletely characterizable, as the majority of analytes would be below detection limits. In the thermal desorption unit, the filter material is heated to evaporate the organic constituents into a helium carrier gas flow. The helium carrier gas is enriched with MSTFA (n-methyl-n-trimethylsilyl-trifluoro-acetamide) by headspace sweep over a pure liquid reservoir. MSTFA is a derivitization agent which replaces the active hydrogen in analyte OH groups with Si-$(CH_3)_3$ groups to significantly improve recovery of polar organics. The derivatized and volatilized sample analytes are then focused on a cooled inlet system before being rapidly heated to simultaneously introduce all analytes to the head of the first GC column. Analytes are then separated by two GC columns in sequence, the first using a semistandard nonpolar stationary phase to separate analytes by volatility while the second

uses a polar stationary phase to separate analytes by polarity. The transition between first and second column is cryomodulated, with the effluent from the first column collected in guard column material cooled by a -80 ºC cold air jet before being simultaneously released to the second column by a hot jet in intervals of 2.1 seconds. The separated analytes are then detected by 70 eV EI HR-ToF-MS (Tofwerk, m/Δm 4000). This technique is sensitive to organic compounds between C13 and C26 n-alkane volatility equivalents.

5.3.2.2 Compilation of a custom mass spectral library and timeline creation

The raw data outputs from the GCxGC instrument were processed by GCimage, which performed automated baseline correction, "blob detection" (the 2-D equivalent of 1-D peak detection), and retention index configuration based on the elution times of a deuterated alkane series included in the internal standard solution. This produced processed sample summaries in which each detected organic is characterized by its first-dimension retention index, second dimension retention time, total ion chromatograph volume, and 70 eV EI mass spectrum. This output is visualized in Figure 5.1, in which a raw chromatograph from a burn impacted time in the wet season is overlaid with circles proportional in size to the volume of each detected compound and color coded, with red indicating sample compounds, yellow indicating internal standard compounds, and blue indicating compounds excluded from analysis due to their presence in field blanks. A custom mass spectral library including the mass spectra and retention indices of unique Amazon organic aerosol analytes was compiled from 11 representative samples, summarized in Table 1. As the goal of this work is to trace as many unique analytes as possible over both seasons, library curation focused on maximizing template sample to template sample differences. Template samples were selected based on analysis provided in De Sá et al., 2018 & 2019 combined with a visual assessment of the total ion chromatographs to identify the appearance of unique populations of compounds and cover a range of atmospheric conditions, including differing levels of urban and biomass burning influence and different times of day. In addition to the compounds identified in the 11 template samples, 15 compounds not present in the template samples but observed in significant quantities at other times were added to the custom mass spectral library. The use of template images for custom mass spectral library curation and trace organic analysis in similar GCxGC applications has been previously described in Zhang et al., 2018.

Initial library curation including the unique compounds observed across the 11 template samples and the 15 manually added compounds yielded an initial mass spectral library of >1500 unique organic species. Using the NIST14 software, all samples were searched against this custom library, and timelines of each compound were created by tracing positive matches across all samples. After timeline creation, a number of low abundance compounds were removed from analysis due to low quality mass spectra and/or infrequent appearance above detection limits, yielding a final dataset of 1,325 unique organic compounds traced across both IOPs.

5.3.2.3 Analyte categorization: directly quantifiable, identifiable, and not identifiable

Quantification and characterization of known and unidentifiable compounds observed in this dataset was the subject of new methodological developments, and the process of categorizing GoAmazon organic analytes is discussed in full in Franklin et al., 2022. Briefly, compounds were categorized as directly quantifiable, identifiable, and not identifiable as follows. A custom 130 component external calibration standard including a range of compounds previously observed under similar conditions and spanning a range of functional group types, volatilities, and polarities was run at various points throughout sample analysis to create calibration curves (discussed in full in "Quantification" below). The chromatographs of the external standard were searched against the custom Amazon sample library, revealing that of the 1325 traced compounds, 63 directly matched the external standard compounds. These species were defined as "directly quantifiable." Next, the template libraries were searched against the NIST/NIH/EPA Main mass spectral database. Following methodology described in Worton et al., 2018, analyte compounds were defined as "identifiable" if they presented a high (>800) mass spectral match factor with the database entry and general agreement between compound retention indices listed in either the NIST/NIH/EPA database or reported in the literature. In restricting the "identifiable" compounds list to those which can be explicitly matched to fully isomerically characterized entries in mass spectral databases, this work adheres to the definition outlined in Nozière et al., 2015, which states "An organic compound is fully identified only if its molecular structure is entirely known, including its isomeric and spatial (stereo) configuration."

5.3.2.4 Quantification

Quantification of GoAmazon compounds, including a full list of compounds comprising the authentic calibration external standard, is described in depth in Franklin et al., 2022. At 5 points throughout the analysis of GoAmazon samples, the external calibration standard was analysed on the GCxGC instruments at 5 sequential mass loadings above a zero point to create 6 point calibration curves. The external standard (stored at -20ºC in a solution of 1:1 methanol:chloroform) was introduced into the instrument by applying it to aliquots of baked tissuequartz filter material identical to that used for GoAmazon aerosol collection. The external standard doped aliquots were subsequently doped with internal standard solution at mass loadings identical to those used on aerosol filter samples before being introduced to the GCxGC instrument to maximize continuity between calibration curves and samples. Each calibration standard and each sample analyte was assigned to its nearest 3 internal standard compounds within a set search radius, and its volume was then normalized by the relative volumes (meaning that the volume of each internal standard is normalized by its own average volume) of the three nearest internal standard compounds. This step corrects for two phenomena that can complicate quantification in GCxGC. First, it corrects for matrix effects, which is when highly loaded samples produce artificially enhanced signal, and second, it corrects for changes in instrument condition. The slope of each calibration curve, in units of (ng external standard

run)/(internal standard normalized volume recorded) are then computed to yield a calibration factor for each external standard compound. The 63 directly matched GoAmazon analyte compounds along with a selection of additional compounds that were identifiable, highly chemically similar to external standard compounds, and close to those compounds in GCxGC space were then quantified using either the quantification factor from the calibration run which occurred closest to the sample analysis run or the average of the two closest calibration runs in cases where two calibration curves bookended an analysis period. This approach has been used in a range of GCxGC applications,[25,28,30,31] and uncertainties are estimated at ± 10% for compounds directly quantified by external standard and ± 50% for proxy quantified compounds.

The not identifiable compounds were quantified using Ch3MS-RF.[27] Briefly, a random forest-based model was trained to predict compound quantification factors based on their mass spectra, position in GCxGC space (eg retention index and second dimension retention time), and when during the analysis period they were processed from the authentic calibration standard and using these patterns to predict what the quantification factor of a given unknown analyte should be at any point during the analysis window. Median random forest-based quantification factor prediction error (when tested on a reserved portion of the external standard not used in model training) was -2%, while the median absolute error was 37%. For both quantification methods, once a mass had been calculated for each sample run, that mass was scaled by the number of aliquots of filter material, the size of the collection area of the filter, the duration of sampling, and the sampling flow rate to yield a time averaged concentration of each organic in air.

5.3.2.5 Chemical Properties Characterization

Developed and originally described in Kroll et al., 2011,[32] carbon number-average carbon oxidation state space diagrams, also known as "Kroll diagrams," map compound properties based on how many carbon atoms the compound has in its ambient (not derivatized) state (n_c) and the estimated average carbon oxidation states of those compounds ($\overline{OS_c}$). This visualization is useful for multiple reasons. First, different areas of Kroll diagram space are associated with different chemical properties. For example, high carbon number (n_c >20) high carbon oxidation state ($\overline{OS_c}$ > -1) compounds are typically water-soluble organic compounds, while other regions are associated with classes of secondary aerosol material produced by different degrees of atmospheric oxidation. Second, changes in the distribution of products throughout this space are indicative of important chemical processes during the atmospheric oxidation of organic material. For example, movement from low to high carbon numbers is indicative of oligomerization, movement towards higher carbon oxidation states while maintaining a consistent carbon number backbone is indicative of functionalization, while decreased carbon numbers coupled with increased carbon oxidation states are the product of fragmentation towards a final product of CO_2.

Both carbon number and average carbon oxidation state are readily extractable or calculable from chemical formulae. For all identifiable compounds therefore, the formulae

of their identified standard or NIST/EPA/NIH database matches were converted to their underivatized forms and n_c and $\overline{OS_c}$ were directly calculated. For the not identifiable compounds Ch3MS-RF was used to predict n_c and $\overline{OS_c}$ based on compound retention indices and mass spectra. A full discussion of Ch3MS-RF, along with optimization and performance evaluation in predicting properties for sample compounds analyzed in this work, is presented in Franklin et al., 2022.[27] When evaluated on the identifiable compounds that did not overlap with the compounds in the external calibration standard, mean absolute error for carbon number predictions was 1.8, while mean absolute error in average carbon oxidation state prediction was .25.

5.3.2.6 Event source attribution by dynamic time warping hierarchical clustering

The compounds identified and traced during the wet season were clustered by dynamic time warping hierarchical clustering, following a methodology previously described in Chapter 3. First, the 100 most abundant compounds were extracted from the rest in order to reduce the potential influence of signal to noise ratio in assigning cluster variability. These compounds were then grouped by dynamic time warping hierarchical clustering.[33]. Clustering of the wet season compounds was optimized by a 9 cluster solution based on a combined analysis of the silhouette index, the Davies-Bouldin index, and the modified Davies-Bouldin index.[34–36] However, upon analysis of the components of each cluster, it was revealed that commonly reported tracers of biomass burning such as levoglucosan had been assigned to the same cluster as compounds such as phthalic acid. This indicates that relative to the high degree of variability among the background and biogenic compounds observed at the field site, the urban and biomass burning influences occurred during similar time frames, likely due to meteorology. To differentiate between the compounds associated with burning activity and those associated with urban emissions, all compounds assigned to the combined anthropogenic influence cluster were extracted and re-clustered. This yielded an optimized 4-cluster solution based on the Davies Bouldin and Modified Davies Bouldin indices. In this clustering scheme, all previously established burning and urban influence tracer compounds were assigned to distinct clusters. Combined with the 8 undisputed clusters from the original analysis, this yielded a final solution of 12 optimized clusters of similar temporal variability in the wet season. For each cluster, an average time profile was calculated by averaging the z-scored abundance of each compound in the cluster at each point in time. All of the additional traced compounds from the wet season were then evaluated for their Pearson correlation with each of the cluster profiles and assigned to the cluster with which they demonstrated the strongest (most positive) correlation.

Clustering of the compounds observed in the dry season was performed using the same methodology previously described for the wet season. Clustering was optimized at a 10-cluster solution based on minimization of the Modified Davies-Bouldin index. In the wet season, tracer compounds typically associated with urban and burning influence were not assigned to the same clusters, and as such additional sub-clustering was deemed unnecessary. Given the higher time resolution of dry season measurements and the

consistent diurnal profiles of urban and burning influence (urban influence typically peaking at night, biomass burning influence peaking during the day), steps were taken to ensure that background biogenic compounds were not assigned to either of these groups based only on sharing a similar diurnal profile. This was achieved by normalizing the time series of each compound by its average abundance during each filter collection period and re clustering the 100 most abundant compounds. When normalized by time of day, the compounds assigned to biomass burning and urban associated clusters (defined by literature review of identifiable components as described below) remained consistent, indicating that these compounds are significantly impacted by burning and urban influence events and do not only share typical diurnal variability.

Clusters were assigned to groups of background/biogenic, urban influenced, or biomass burning influenced based on a combination of literature review of identifiable constituents, analysis of the correlation between mean cluster profiles and selected primary tracers of urban or biomass burning influence, and analysis of correlation between mean cluster profiles and AMS-PMF factors (described in "Supporting Measurements"). Levoglucosan was selected as the primary biomass burning tracer, as it was observed in significant quantities and its utility as a tracer of burn activity has been widely characterized and utilized.[37] Phthalic acid was selected as the primary urban influence tracer, as it has been identified as an urban tracer in a variety of contexts and its abundance peaked during periods previously identified as urban influenced in De Sá et al., 2018, 2019. A summary of selected identifiable organic constituents previously established as tracers of biogenic, urban, and biomass burning organic aerosol formation or emissions is presented in Table 2. Importantly, this work groups compounds by temporal variability and event source rather than mass origin, meaning that biogenic secondary organic aerosol products whose production is significantly enhanced by the presence of anthropogenic oxidants or that are formed exclusively in the presence of those oxidants are defined as urban or burn influenced. In the wet season, three clusters were identified as urban influenced, two clusters were identified and biomass burning influenced, and seven were defined as background/biogenic. In the dry season, one cluster was identified as urban influenced, two clusters were identified and biomass burning influenced, and seven clusters were defined as background/biogenic.

5.3.3 Supporting Measurements

Several additional measurements were made at the T3 site that provide important context for the analysis presented in this work. Real-time particle phase bulk measurements were made by high resolution time-of-flight aerosol mass spectrometer (AMS, Aerodyne). These measurements are described in detail in De Sá et al., 2018, 2019. The aerosol mass spectrometer characterized the fractional contribution of organics, sulfate, nitrate, ammonium, and chloride to total submicron aerosol mass. In addition, the organic fraction was analysed using positive matrix factorization resolve factors and attribute organic mass to different compositional source groups. Complementary speciated measurements of individual organics were conducted made by semi-volatile thermal desorption aerosol gas chromatograph (SV-TAG), as described in Yee et al., 2018. SV-TAG measures total (gas

+ particle) and particle concentrations of organics to identify the partitioning of semivolatile species between gas and aerosol phases.

5.4 Results and Discussion

5.4.1 Anthropogenic Perturbations of Aerosol Chemical Property Distributions

Burning influenced, urban influenced, and background/biogenic compounds inhabit different though overlapping areas of chemical properties space. Each event source group contains a diverse group of species and the chemical property spaces of each population overlap, as is indicated by the overlapping standard deviation error bars in Figure 5.2. However, differences between property distributions of each population of compounds (both interseasonally and between event source groups in each season) are statistically significant (t-test p-values <.05), with the following exceptions. The wet and dry season urban influenced clusters are not significantly different in either dimension, and the average carbon numbers of wet season burning influenced and background/biogenic compounds are not significantly different.

As illustrated in Figure 5.2, urban influenced organics inhabit consistently lower carbon number and more oxidized regions of the Kroll diagram compared to the other influence categories, with an average carbon number of 9.3 and average carbon oxidation state of -.49 in the wet season and an average carbon number of 9.2 and average carbon oxidation state of -.46 in the dry season. The urban influenced clusters are also most tightly grouped with respect to carbon number, with carbon number standard deviations of 2.7 and 2.8 (wet and dry season respectively), compared to standard deviations consistently > for all other groupings. This phenomenon is illustrated in the example sample distribution mappings provided in Panel B of figures 5.3 and 5.4; in both the wet season and the dry season, urban influence appears as a relatively tightly clustered group of species in the smallest and most oxidized corner of $\overline{OS_c} - n_c$ space. Given the parameterization of precursor oligomerization, fragmentation, and functionalization outlined in Kroll et al., 2011, this is likely indicative of higher degrees of fragmentation in the secondary aerosol formation pathways forming these products, which could at least partially contribute to the phenomenon reported in Shilling et al., 2018, which found less organic aerosol mass enhancement from secondary aerosol formation than expected in the GoAmazon urban plume. The position of the average chemical properties of the both sets of urban attributed measurements also point to likely contributions from enhanced secondary aerosol production, specifically from monoterpenes. At an average carbon number of slightly greater than 9, more than half of the products attributed to urban influence have 10 or fewer carbons, placing them within the chemical properties spaces of previously established oxidation pathways for monoterpenes in chamber conditions.[32,38] Together, these observations indicate that the oxidant conditions in the urban plume shift biogenic precursor oxidation towards more oxygenated and lower carbon number products, and that these perturbation processes may play an important role in the fate of monoterpenes.

In contrast to the urban influenced compounds, biomass burning influenced products span a wide range of properties and demonstrate a significant interseasonal drift, with dry season burning products on average higher carbon number and more oxidized. This is primarily attributable to a population of high carbon number oxidized compounds that are present in the dry season but not the wet season, as will be discussed in greater detail in *Observations of Seasonally Unique Organic Aerosol Products*. Some of these products are observable in the example property distribution of a biomass burning influenced sample from the dry season, illustrated in Figure 5.4 panel A. The burning influenced compound grouping contains a compositionally diverse suite of chemicals covering a particularly large range of carbon numbers, with carbon number standard deviations of 5.3 and 6.0 for the wet and dry seasons respectively. This grouping includes products previously established as primary biomass burning products, two of which are listed in Table 5.2. These two well characterized primary products, a 6 carbon sugar, levoglucosan, and a 24 carbon alkanoic acid, tetracosanoic acid, are illustrative of two disparate but consistently co-observed populations of biomass burning tracers, specifically low carbon number and highly oxidized sugars and high carbon number less oxidized alkanoic acids and hydrocarbons. Both groups can be seen in panel A of figures 5.3 and 5.4, in which levoglucosan is illustrated by the largest orange circle. In addition to these previously characterized primary products, there are a large number of burning attributed compounds falling inside the region of the Kroll diagram typically representative of secondary organic aerosol products,[32] and despite the interseasonal shift both wet and dry seasons' mean measures of biomass burning attributed aerosol properties fall within this region. As is discussed in *Diversity, Properties, and Importance of Unidentifiable Organics* these compounds are largely not identifiable by mass spectral database match and represent opportunities for additional investigation into speciated secondary aerosol production from biomass burning precursors.

Background/biogenic compounds are by number the most diverse compound grouping, making of 40% of compounds traced in the dry season and 49% of compounds traced during the wet season and covering ranges of both carbon number and carbon oxidation state properties similar to those assigned to biomass burning source groups. Similarly to the biomass burning compounds, the seasonal compositional spaces are significantly different largely due to a population of seasonally unique products; as discussed in *Observations of Seasonally Unique Organic Aerosol Products* and illustrated in panel A of figure 5.5, unique pristine products are observed in the dry season that are typically higher carbon number and relatively less oxidized, contributing to wet season background/biogenic products being higher carbon number and lower average carbon oxidation state compared to similarly categorized dry season products.

Previous work has established that the bulk organic composition of Amazonian submicron organic aerosol ass assessed by AMS is more oxidized in the dry season than in the wet season. While the composition of the urban influenced grouping remains statistically unchanged, the chemical distribution spaces of both biomass burning influenced and background/biogenic organic aerosol is significantly shifted towards more

oxygenated species. While GCxGC analysis is not sensitive to the entire organic carbon pool and the compositional analysis presented in Figure 5.2 is not mass weighted, rendering the findings not directly comparable between speciated and bulk measurements, the general agreement between bulk and speciated observations highlights the utility of investigating speciated compositional shifts to better understand changes in bulk properties. Shifts in the non-mass weighted compositional distributions indicate that not only the relative abundance of products formed but the characteristics of products themselves play an influential role in interseasonal differences in atmospheric chemistry.

5.4.2 Observations of Seasonally Unique Organic Aerosol Products

Given that all samples were collected from the same location, the interseasonal consistency in the individual identities of submicron organic aerosol products detected was surprisingly low. Of the 1,325 compounds traced, only 52% were observed in both seasons. 31% of traced compounds were observed exclusively in the wet season, while 17% were observed exclusively in the dry season. This breakdown, along with the event source breakdown and chemical properties distributions of seasonally unique and commonly observed compounds, is illustrated in Figure 5.5. While some of the interseasonal differences may be attributable to limits of detection (meaning that some "seasonally unique" products may have been present in the other season but always below detection limits), this does not diminish the relevance of the unique products observed. Importantly, species were only removed from analysis in early stages if they were not traceable in either season, and compounds were only classified as seasonally unique if they were never observed above detection limits in the other season. A consistent limit of detection based on contribution to the total ion chromatograph was applied to all samples, making the assignments of seasonally unique products the most conservative assessment possible given instrumental limitations.

Of the compounds exclusively observed in the dry season, most were attributed to biomass burning (75%), while 15% were attributed to urban influence and 10% attributed to background/biogenic sourced. By numbers, the wet season compounds were mostly attributed to background/biogenic sources (61%), while 25% were attributed to urban influence and 14% were attributed to biomass burning. A quantitative analysis of organic aerosol mass attributed to seasonally unique species during the ~2 week continuous analysis periods of each season (Figure 5.6) reveals the following findings. Unique biomass burning attributed compounds (Figure 5.6 panel B) dominate the seasonally unique mass concentrations but are highly episodic, particularly at the beginning of the analysis period. This is largely due to the significance and short duration of the appearance of the unique population of high carbon number oxidized compounds illustrated in the top row of Figure 5.5 panel C, which appeared only under particularly intense burning influenced conditions. Based on their location in the Kroll diagram, these compounds are likely high carbon number water soluble organic carbon, which has been previously identified in biomass burning emissions and has been attributed to both primary emissions[39] and atmospheric oligomerization processes, specifically aqueous-phase photooxidation.[40] Potential explanations for the unique observations of these compounds

in the dry season include preferential loss of these products to wet deposition during the wet season, likely to be particularly important due to their solubilities, production timescales requiring the longer atmospheric residence times and greater light availability typical of the dry season compared to the wet season, or differences in primary emissions from differing combustion conditions.

During the wet season, uniquely produced biomass burning products contributed relatively little to the observed organic mass compared to the urban influenced and background/biogenic groupings, as illustrated in Figure 5.6. While the wet season unique urban influenced products exhibited relatively similar (though slightly higher carbon number) properties compared to the commonly observed urban influenced organics, the wet season background/biogenic grouping contains a population of highly compositionally distinct products. These products, illustrated in the bottom row of Figure 5.5 Panel A, include a diverse group of high carbon number (< 15) relatively low carbon oxidation state species. While previous work has identified the formation high carbon oligomers from aqueous processing of biogenic precursors in cloud droplets, the production of these products is predicted to increase with in the presence of elevated urban oxidant conditions, which is explicitly not the case for the compounds reported in this work.[23] In fact, 5 of the highest carbon number, most oxidized members of this unique product group are anticorrelated with the urban plume tracer, producing Pearson correlation coefficients of < .4. As discussed below in *Diversity, Properties, and Importance of Unidentifiable Organics* and mirroring the findings for the unique dry season biomass burning influenced products, the majority of these uniquely produces wet season biogenic species are not identifiable by database match, rendering father mechanistic insights into how these species are being produced beyond the scope of this work. That said, the presence and properties of interseasonally unique organic aerosol components point to differences in secondary aerosol formation processes that could guide the needed conditions (humidity, temperature, concentrations) for future laboratory oxidation experiments to replicate previously unestablished oxidation processes.

5.4.3 Diversity, Properties, and Importance of Unidentifiable Organics

Of the organic aerosol constituents catalogued in this work, very few were identifiable by match to an authentic standard or entry in the NIST/NIH/EPA mass spectral databases. Despite the compositional differences previously described, this observation was highly consistent between seasons: 91% of compounds traced during the wet season were not identifiable, while 90% of compounds traced in the dry season were not identifiable. Of the three influence categories described in this work, urban influenced organics were by numbers the least well known, with 95% of dry season and 96% of wet season compounds not identifiable by database match. Burning associated compounds were consistently better known, with 85% of dry season and 82% of wet season biomass burning species available as standards or present in mass spectral databases, while the background/biogenic compounds fell in a middle ground, with 90% and 89% of wet season compounds identifiable, respectively. The properties of identifiable and not identifiable compounds in from each season segregated by event influence source are illustrated in Figure 5.7. Both

of the seasonally and compositionally unique populations described in *Observations of Seasonally Unique Organic Aerosol Products*, specifically the unique high carbon number dry season organics associated with biomass burning and the unique high carbon number wet season organics observed under background conditions were entirely not identifiable. The relevance of these species to the design of future oxidation experiments is discussed in *Implications for Future Laboratory Studies* below.

The contributions of unidentifiable compounds to organic aerosol mass vary substantially, but are consistently dominant under wet season highly urban influenced conditions, as illustrated in Figure 5.8. During significantly urban influenced sampling periods, defined as times when the urban influence plume tracer was measured at > 1 standard deviation above its mean, consistently > 60% of recovered aerosol mass was attributed to unidentifiable compounds. This observation agrees with and partially explains the findings reported in Mouchel-Vallon et al., 2020, which found that while model parameterizations using the volatility basis set were able to reproduce observed aerosol enhancements within the Manaus urban plume, explicit product modelling was not able to match the observed enhancement. While the chemical oxidation modelling is not limited to products catalogued in mass spectral databases, the significance of unidentifiable organics to total wet season submicron organic aerosol concentrations under urban influenced conditions highlights the following challenge in improving explicit modelling of these conditions. Because the majority of individual urban influenced products cannot be identified and a dominant mass fraction (at least within the composition ranges to which GCxGC is sensitive) of organic material under urban influenced conditions is attributable to these unidentifiable compounds, the production mechanisms relevant to plume influenced secondary aerosol production cannot currently be structurally verified. Authentic synthesis and characterization of these products will significantly enhance mechanistic understandings of how urban perturbations of secondary aerosol formation occur, as will, as discussed below, future laboratory experiments.

5.4.4 Implications for Future Laboratory Studies

Replication of the unidentifiable products observed in Amazonian organic aerosol that are introduced here will play a critical role in expanding the scope of findings presented in this work. While still highly complex, emissions of gas-phase reactive organic compounds are less complex and better characterized than their aerosol phase oxidation products.[20,21,41] As discussed in *Diversity, Properties, and Importance of Unidentifiable Organics*, the number and diversity of not identifiable products observed in this work is quite high; the time and effort required to synthesize new compounds is prohibitively high to provide a primary solution to improving characterization of these complex mixtures. Laboratory oxidation experiments provide extremely valuable opportunities to advance understanding of these important compound populations; when aerosol produced from laboratory oxidation experiments is analyzed using the same instruments and protocols as ambient samples, unidentifiable compounds that are matched between experimental and ambient samples can be attributed to likely precursor sources and/or oxidation conditions. However, in order for these comparisons to be useful, experiments must be conducted under

conditions that are sufficiently similar to those of the real atmosphere, and as discussed in Porter et al., 2021 this is usually not the case. For this application humidity is likely to be particularly important, as most laboratory oxidation experiments are conducted under extremely low humidity conditions compared to the high humidity of the Amazon rainforest. Heterogeneous chemistry and aqueous-phase processing also deserve consideration, as aqueous-phase oligomerization has been previously reported and could potentially explain the not identifiable and seasonally unique high carbon number populations illustrated in Figure 5.5. The event source influence attributions and chemical properties distributions of chemically interesting compound populations such as those highlighted throughout this work can be utilized to design targeted laboratory oxidation experiments to improve mechanistic insights into the complexity of ambient aerosol formation.

5.5 Conclusion

In this work, we present a speciated isomer-specific analysis of the organic composition of submicron organic aerosol collected under a range of urban and biomass burning influenced ambient conditions over two seasons in a forested region of the Amazon rainforest. Based on this analysis, we conclude that oxidants from the Manaus plume influence the fate of biogenic precursors, producing more highly oxygenated products and shifting oxidation pathways towards more fragmentation compared to burning influenced and pristine conditions. We additionally report observations of interseasonally unique products, the most notable of which include high carbon number highly oxygenated biomass burning products in the wet season and high carbon number less oxidized products observed under pristine conditions in the wet season. These findings indicate that different atmospheric conditions produce not only differing mixtures of secondary products but also entirely unique products, likely attributable to differences in atmospheric lifetimes and oxidant conditions. The absence of the uniquely observed organic species from mass spectral databases inhibits more precise mechanistic insights into how uniquely produced species are formed and highlights the importance of replicating real ambient oxidation conditions under laboratory settings and fully isomerically characterizing products of known precursors to better understand the processes underpinning interseasonal organic aerosol formation sensitivity to different anthropogenic perturbations. Interseasonally common compounds are likewise incompletely characterized in mass spectral databases and all unidentifiable products substantially contribute to total submicron aerosol mass particularly under urban influenced wet season conditions. As urbanization intensifies and atmospheric conditions evolve, mechanistic understandings of these and other processes critical to the composition and quantity of organic aerosol produced from the central Amazon will be critical for understanding and predicting potential impacts on public health and climate.

5.6 Acknowledgements

We gratefully acknowledge the support of the National Science Foundation Graduate Research Fellowship Program (DGE- 1752814), as well as the Department of Energy Atmospheric System Research program (DE-SC0020051). The described work was

conducted under DOE-ASR support and benefits from samples collected at the GoAmazon field campaign, a DOE-ASR supported observational campaign. We acknowledge the support from the Central Office of the Large Scale Biosphere Atmosphere Experiment in Amazonia (LBA), the Instituto Nacional de Pesquisas da Amazonia (INPA), and the Instituto Nacional de Pesquisas Espaciais (INPE). The work was conducted under 2009/15235-8 of the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

5.7 References

(1) Shrivastava, M.; Andreae, M. O.; Artaxo, P.; Barbosa, H. M. J.; Berg, L. K.; Brito, J.; Ching, J.; Easter, R. C.; Fan, J.; Fast, J. D.; Feng, Z.; Fuentes, J. D.; Glasius, M.; Goldstein, A. H.; Gomes Alves, E.; Gomes, H.; Gu, D.; Guenther, A.; Jathar, S. H.; Kim, S.; Liu, Y.; Lou, S.; Martin, S. T.; Mcneill, V. F.; Medeiros, A.; De Sá, S. S.; Shilling, J. E.; Springston, S. R.; Souza, R. A. F.; Thornton, J. A.; Isaacman-Vanwertz, G.; Yee, L. D.; Ynoue, R.; Zaveri, R. A.; Zelenyuk, A.; Zhao, C. Urban Pollution Greatly Enhances Formation of Natural Aerosols over the Amazon Rainforest. *Nat. Commun.* **2019**. https://doi.org/10.1038/s41467-019-08909-4.

(2) Artaxo, P.; Rizzo, L. V.; Brito, J. F.; Barbosa, H. M. J.; Arana, A.; Sena, E. T.; Cirino, G. G.; Bastos, W.; Martin, S. T.; Andreae, M. O. Atmospheric Aerosols in Amazonia and Land Use Change: From Natural Biogenic to Biomass Burning Conditions. *Faraday Discuss.* **2013**, *165* (0), 203–235. https://doi.org/10.1039/C3FD00052D.

(3) De Sá, S. S.; Rizzo, L. V; Palm, B. B.; Campuzano-Jost, P.; Day, D. A.; Yee, L. D.; Wernis, R.; Isaacman-Vanwertz, G.; Brito, J.; Carbone, S.; Liu, Y. J.; Sedlacek, A.; Springston, S.; Goldstein, A. H.; Barbosa, H. M. J.; Alexander, M. L.; Artaxo, P.; Jimenez, J. L.; Martin, S. T.; Paulson, J. A. Contributions of Biomass-Burning, Urban, and Biogenic Emissions to the Concentrations and Light-Absorbing Properties of Particulate Matter in Central Amazonia during the Dry Season. *Atmos. Chem. Phys* **2019**, *19*, 7973–8001. https://doi.org/10.5194/acp-19-7973-2019.

(4) De Sá, S. S.; Palm, B. B.; Campuzano-Jost, P.; Day, D. A.; Hu, W.; Isaacman-Vanwertz, G.; Yee, L. D.; Brito, J.; Carbone, S.; Ribeiro, I. O.; Cirino, G. G.; Liu, Y.; Thalman, R.; Sedlacek, A.; Funk, A.; Schumacher, C.; Shilling, J. E.; Schneider, J.; Artaxo, P.; Goldstein, A. H.; Souza, R. A. F.; Wang, J.; Mckinney, K. A.; Barbosa, H.; Alexander, M. L.; Jimenez, J. L.; Martin, S. T. Urban Influence on the Concentration and Composition of Submicron Particulate Matter in Central Amazonia. *Atmos. Chem. Phys* **2018**, *18*, 12185–12206. https://doi.org/10.5194/acp-18-12185-2018.

(5) Liu, L.; Cheng, Y.; Wang, S.; Wei, C.; Pöhlker, M. L.; Pöhlker, C.; Artaxo, P.;

Shrivastava, M.; Andreae, M. O.; Pöschl, U.; Su, H. Impact of Biomass Burning Aerosols on Radiation, Clouds, and Precipitation over the Amazon: Relative Importance of Aerosol-Cloud and Aerosol-Radiation Interactions. *Atmos. Chem. Phys.* **2020**, *20* (21), 13283–13301. https://doi.org/10.5194/ACP-20-13283-2020.

(6) Zhao, B.; Fast, J. D.; Donahue, N. M.; Shrivastava, M.; Schervish, M.; Shilling, J. E.; Gordon, H.; Wang, J.; Gao, Y.; Zaveri, R. A.; Liu, Y.; Gaudet, B. Impact of Urban Pollution on Organic-Mediated New-Particle Formation and Particle Number Concentration in the Amazon Rainforest. *Environ. Sci. Technol.* **2021**, *55* (8), 4357–4367. https://doi.org/10.1021/ACS.EST.0C07465/SUPPL_FILE/ES0C07465_SI_001.PDF.

(7) Urrutia-Pereira, M.; Rizzo, L. V.; Chong-Neto, H. J.; Solé, D. Impact of Exposure to Smoke from Biomass Burning in the Amazon Rain Forest on Human Health. *J. Bras. Pneumol.* **2021**, *47* (5), 11. https://doi.org/10.36416/1806-3756/E20210219.

(8) Nawaz, M. O.; Henze, D. K. Premature Deaths in Brazil Associated With Long-Term Exposure to PM2.5 From Amazon Fires Between 2016 and 2019. *GeoHealth* **2020**, *4* (8), e2020GH000268. https://doi.org/10.1029/2020GH000268.

(9) Davidson, E. A.; De Araüjo, A. C.; Artaxo, P.; Balch, J. K.; Brown, I. F.; Mercedes, M. M.; Coe, M. T.; Defries, R. S.; Keller, M.; Longo, M.; Munger, J. W.; Schroeder, W.; Soares-Filho, B. S.; Souza, C. M.; Wofsy, S. C. The Amazon Basin in Transition. *Nat. 2012 4817381* **2012**, *481* (7381), 321–328. https://doi.org/10.1038/nature10717.

(10) Cardil, A.; de-Miguel, S.; Silva, C. A.; Reich, P. B.; Calkin, D.; S Brancalion, P. H.; Vibrans, A. C.; P Gamarra, J. G.; Zhou, M.; Pijanowski, B. C.; Hui, C.; Crowther, T. W.; Hérault, B.; Piotto, D.; Salas-Eljatib, C.; North Broadbent, E.; Almeyda Zambrano, A. M.; Picard, N.; O C Aragão, L. E.; Bastin, J.-F.; Routh, D.; van den Hoogen, J.; Peri, P. L.; Liang, J. Recent Deforestation Drove the Spike in Amazonian Fires Environmental Research Letters Recent Deforestation Drove the Spike in Amazonian Fires. *Environ. Res. Lett.* **2020**. https://doi.org/10.1088/1748-9326/abcac7.

(11) Shilling, J. E.; Pekour, M. S.; Fortner, E. C.; Artaxo, P.; De Sá, S.; Hubbe, J. M.; Longo, K. M.; Machado, L. A. T.; Martin, S. T.; Springston, S. R.; Tomlinson, J.; Wang, J. Aircraft Observations of the Chemical Composition and Aging of Aerosol in the Manaus Urban Plume during GoAmazon 2014/5. *Atmos. Chem. Phys.* **2018**, *18* (14), 10773–10797. https://doi.org/10.5194/ACP-18-10773-2018.

(12) Martin, S. T.; Artaxo, P.; Machado, L. A. T.; Manzi, A. O.; Souza, R. A. F.; Schumacher, C.; Wang, J.; Andreae, M. O.; Barbosa, H. M. J.; Fan, J.; Fisch, G.; Goldstein, A. H.; Guenther, A.; Jimenez, J. L.; Pöschl, U.; Dias, M. A. S.; Smith, J. N.; Wendisch, M. Introduction: Observations and Modeling of the Green Ocean Amazon (GoAmazon2014/5). *Atmos. Chem. Phys* **2016**, *16*, 4785–4797. https://doi.org/10.5194/acp-16-4785-2016.

(13) Mouchel-Vallon, C.; Lee-Taylor, J.; Hodzic, A.; Artaxo, P.; Aumont, B.; Camredon, M.; Gurarie, D.; Jimenez, J. L.; Lenschow, D. H.; Martin, S. T.; Nascimento, J.; Orlando, J. J.; B. Palm, B.; Shilling, J. E.; Shrivastava, M.; Madronich, S. Exploration of Oxidative Chemistry and Secondary Organic Aerosol Formation in the Amazon during the Wet Season: Explicit Modeling of the Manaus Urban Plume with GECKO-A. *Atmos. Chem. Phys.* **2020**, *20* (10), 5995–6014. https://doi.org/10.5194/ACP-20-5995-2020.

(14) Rafee, S. A. A.; Martins, L. D.; Kawashima, A. B.; Almeida, D. S.; Morais, M. V. B.; Souza, R. V. A.; Oliveira, M. B. L.; Souza, R. A. F.; Medeiros, A. S. S.; Urbina, V.; Freitas, E. D.; Martin, S. T.; Martins, J. A. Contributions of Mobile, Stationary and Biogenic Sources to Air Pollution in the Amazon Rainforest: A Numerical Study with the WRF-Chem Model. *Atmos. Chem. Phys* **2017**, *17*, 7977–7995. https://doi.org/10.5194/acp-17-7977-2017.

(15) Cirino, G.; Brito, J.; Barbosa, H. M. J.; Rizzo, L. V.; Tunved, P.; de Sá, S. S.; Jimenez, J. L.; Palm, B. B.; Carbone, S.; Lavric, J. V.; Souza, R. A. F.; Wolff, S.; Walter, D.; Tota, J.; Oliveira, M. B. L.; Martin, S. T.; Artaxo, P. Observations of Manaus Urban Plume Evolution and Interaction with Biogenic Emissions in GoAmazon 2014/5. *Atmos. Environ.* **2018**, *191*, 513–524. https://doi.org/10.1016/J.ATMOSENV.2018.08.031.

(16) Nascimento, J. P.; Bela, M. M.; Meller, B. B.; Banducci, A. L.; Rizzo, L. V.; Liduvino Vara-Vela, A.; Barbosa, H. M. J.; Gomes, H.; Rafee, S. A. A.; Franco, M. A.; Carbone, S.; Cirino, G. G.; Souza, R. A. F.; Mckeen, S. A.; Artaxo, P. Aerosols from Anthropogenic and Biogenic Sources and Their Interactions-Modeling Aerosol Formation, Optical Properties, and Impacts over the Central Amazon Basin. *Atmos. Chem. Phys.* **2021**, *21* (9), 6755–6779. https://doi.org/10.5194/ACP-21-6755-2021.

(17) Palm, B. B.; De Sá, S. S.; Day, D. A.; Campuzano-Jost, P.; Hu, W.; Seco, R.; Sjostedt, S. J.; Park, J. H.; Guenther, A. B.; Kim, S.; Brito, J.; Wurm, F.; Artaxo, P.; Thalman, R.; Wang, J.; Yee, L. D.; Wernis, R.; Isaacman-VanWertz, G.; Goldstein, A. H.; Liu, Y.; Springston, S. R.; Souza, R.; Newburn, M. K.; Lizabeth Alexander, M.; Martin, S. T.; Jimenez, J. L. Secondary Organic Aerosol Formation

from Ambient Air in an Oxidation Flow Reactor in Central Amazonia. *Atmos. Chem. Phys.* **2018**, *18* (1), 467–493. https://doi.org/10.5194/ACP-18-467-2018.

(18) Glasius, M.; Bering, M. S.; Yee, L. D.; De Sá, S. S.; Isaacman-VanWertz, G.; Wernis, R. A.; Barbosa, H. M. J.; Alexander, M. L.; Palm, B. B.; Hu, W.; Campuzano-Jost, P.; Day, D. A.; Jimenez, J. L.; Shrivastava, M.; Martin, S. T.; Goldstein, A. H. Organosulfates in Aerosols Downwind of an Urban Region in Central Amazon. *Environ. Sci. Process. Impacts* **2018**, *20* (11), 1546–1558. https://doi.org/10.1039/C8EM00413G.

(19) Liu, Y.; Seco, R.; Kim, S.; Guenther, A. B.; Goldstein, A. H.; Keutsch, F. N.; Springston, S. R.; Watson, T. B.; Artaxo, P.; Souza, R. A. F.; Mckinney, K. A.; Scot, † ‡; Martin, T. Isoprene Photo-Oxidation Products Quantify the Effect of Pollution on Hydroxyl Radicals over Amazonia. **2018**.

(20) Goldstein, A. H.; Galbally, I. E. Known and Unexplored Organic Constituents in the Earth's Atmosphere. *Environ. Sci. Technol.* **2007**, *41* (5), 1514–1521. https://doi.org/10.1021/ES072476P.

(21) Heald, C. L.; Kroll, J. H. The Fuel of Atmospheric Chemistry: Toward a Complete Description of Reactive Organic Carbon. *Sci. Adv.* **2020**, *6* (6). https://doi.org/10.1126/SCIADV.AAY8967/SUPPL_FILE/AAY8967_SM.PDF.

(22) Porter, W. C.; Jimenez, J. L.; Barsanti, K. C. Quantifying Atmospheric Parameter Ranges for Ambient Secondary Organic Aerosol Formation. *ACS Earth Sp. Chem.* **2021**, *5* (9), 2380–2397. https://doi.org/10.1021/ACSEARTHSPACECHEM.1C00090/SUPPL_FILE/SP1C00090_SI_003.XLSX.

(23) Ervens, B.; Turpin, B. J.; Weber, R. J. Secondary Organic Aerosol Formation in Cloud Droplets and Aqueous Particles (AqSOA): A Review of Laboratory, Field and Model Studies. *Atmos. Chem. Phys* **2011**, *11*, 11069–11102. https://doi.org/10.5194/acp-11-11069-2011.

(24) Yee, L. D.; Isaacman-Vanwertz, G.; Wernis, R. A.; Meng, M.; Rivera, V.; Kreisberg, N. M.; Hering, S. V; Bering, M. S.; Glasius, M.; Upshur, M. A.; Bé, A. G.; Thomson, R. J.; Geiger, F. M.; Offenberg, J. H.; Lewandowski, M.; Kourtchev, I.; Kalberer, M.; De Sá, S.; Martin, S. T.; Alexander, M. L.; Palm, B. B.; Hu, W.; Campuzano-Jost, P.; Day, D. A.; Jimenez, J. L.; Liu, Y.; Mckinney, K. A.; Artaxo, P.; Viegas, J.; Manzi, A.; Oliveira, M. B.; De Souza, R.; Machado, L. A. T.; Longo, K.; Goldstein, A. H. Observations of Sesquiterpenes and Their Oxidation Products in Central Amazonia during the Wet and Dry Seasons. *Atmos. Chem.*

Phys **2018**, *18*, 10433–10457. https://doi.org/10.5194/acp-18-10433-2018.

(25) Franklin, E. B.; Alves, M. R.; Moore, A. N.; Kilgour, D. B.; Novak, G. A.; Mayer, K.; Sauer, J. S.; Weber, R. J.; Dang, D.; Winter, M.; Lee, C.; Cappa, C. D.; Bertram, T. H.; Prather, K. A.; Grassian, V. H.; Goldstein, A. H. Atmospheric Benzothiazoles in a Coastal Marine Environment. *Environ. Sci. Technol.* **2021**, *55* (23), 15705–15714. https://doi.org/10.1021/ACS.EST.1C04422/SUPPL_FILE/ES1C04422_SI_001.PDF.

(26) Worton, D. R.; Decker, M.; Isaacman-VanWertz, G.; Chan, A. W. H.; Wilson, K. R.; Goldstein, A. H. Improved Molecular Level Identification of Organic Compounds Using Comprehensive Two-Dimensional Chromatography, Dual Ionization Energies and High Resolution Mass Spectrometry. *Analyst* **2017**, *142* (13), 2395–2403. https://doi.org/10.1039/C7AN00625J.

(27) Franklin, E. B.; Yee, L. D.; Aumont, B.; Weber, R. J.; Grigas, P.; Goldstein, A. Ch3MS-RF: A Random Forest Model for Chemical Characterization and Improved Quantification of Unidentified Atmospheric Organics Detected by Chromatography-Mass Spectrometry Techniques. *Atmos. Meas. Tech- Discuss.* **2022**. https://doi.org/10.5194/amt-2022-99.

(28) Zhang, H.; Yee, L. D.; Lee, B. H.; Curtis, M. P.; Worton, D. R.; Isaacman-Vanwertz, G.; Offenberg, J. H.; Lewandowski, M.; Kleindienst, T. E.; Beaver, M. R.; Holder, A. L.; Lonneman, W. A.; Docherty, K. S.; Jaoui, M.; Pye, H. O. T.; Hu, W.; Day, D. A.; Campuzano-Jost, P.; Jimenez, J. L.; Guo, H.; Weber, R. J.; De Gouw, J.; Koss, A. R.; Edgerton, E. S.; Brune, W.; Mohr, C.; Lopez-Hilfiker, F. D.; Lutz, A.; Kreisberg, N. M.; Spielman, S. R.; Hering, S. V; Wilson, K. R.; Thornton, J. A.; Goldstein, A. H. Monoterpenes Are the Largest Source of Summertime Organic Aerosol in the Southeastern United States. *PNAS* **2018**. https://doi.org/10.1073/pnas.1717513115.

(29) Nozière, B.; Kalberer, M.; Claeys, M.; Allan, J.; D'Anna, B.; Decesari, S.; Finessi, E.; Glasius, M.; Grgić, I.; Hamilton, J. F.; Hoffmann, T.; Iinuma, Y.; Jaoui, M.; Kahnt, A.; Kampf, C. J.; Kourtchev, I.; Maenhaut, W.; Marsden, N.; Saarikoski, S.; Schnelle-Kreis, J.; Surratt, J. D.; Szidat, S.; Szmigielski, R.; Wisthaler, A. The Molecular Identification of Organic Compounds in the Atmosphere: State of the Art and Challenges. *Chem. Rev.* **2015**, *115* (10), 3919–3983. https://doi.org/10.1021/CR5003485/ASSET/IMAGES/CR5003485.SOCIAL.JPEG_V03.

(30) Jen, C. N.; Hatch, L. E.; Selimovic, V.; Yokelson, R. J.; Weber, R.; Fernandez, A.

E.; Kreisberg, N. M.; Barsanti, K. C.; Goldstein, A. H. Speciated and Total Emission Factors of Particulate Organics from Burning Western US Wildland Fuels and Their Dependence on Combustion Efficiency. *Atmos. Chem. Phys.* **2019**, *19* (2), 1013–1026. https://doi.org/10.5194/ACP-19-1013-2019.

(31) Liang, Y.; Jen, C. N.; Weber, R. J.; Misztal, P. K.; Goldstein, A. H. Chemical Composition of PM2.5 in October 2017 Northern California Wildfire Plumes. *Atmos. Chem. Phys.* **2021**, *21* (7), 5719–5737. https://doi.org/10.5194/ACP-21-5719-2021.

(32) Kroll, J. H.; Donahue, N. M.; Jimenez, J. L.; Kessler, S. H.; Canagaratna, M. R.; Wilson, K. R.; Altieri, K. E.; Mazzoleni, L. R.; Wozniak, A. S.; Bluhm, H.; Mysak, E. R.; Smith, J. D.; Kolb, C. E.; Worsnop, D. R. Carbon Oxidation State as a Metric for Describing the Chemistry of Atmospheric Organic Aerosol. *Nat. Chem. 2010 32* **2011**, *3* (2), 133–139. https://doi.org/10.1038/nchem.948.

(33) Łuczak, M. Hierarchical Clustering of Time Series Data with Parametric Derivative Dynamic Time Warping. *Expert Syst. Appl.* **2016**, *62*, 116–130. https://doi.org/10.1016/J.ESWA.2016.06.012.

(34) Rousseeuw, P. J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Math.* **1987**, *20* (C), 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

(35) Davies, D. L.; Bouldin, D. W. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *PAMI-1* (2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909.

(36) Kim, M.; Ramakrishna, R. S. New Indices for Cluster Validity Assessment. *Pattern Recognit. Lett.* **2005**, *26* (15), 2353–2363. https://doi.org/10.1016/J.PATREC.2005.04.007.

(37) Bhattarai, H.; Saikawa, E.; Wan, X.; Zhu, H.; Ram, K.; Gao, S.; Kang, S.; Zhang, Q.; Zhang, Y.; Wu, G.; Wang, X.; Kawamura, K.; Fu, P.; Cong, Z. Levoglucosan as a Tracer of Biomass Burning: Recent Progress and Perspectives. *Atmos. Res.* **2019**, *220*, 20–33. https://doi.org/10.1016/J.ATMOSRES.2019.01.004.

(38) Faiola, C. L.; Pullinen, I.; Buchholz, A.; Khalaj, F.; Ylisirniö, A.; Kari, E.; Miettinen, P.; Holopainen, J. K.; Kivimäenpää, M.; Schobesberger, S.; Yli-Juuti, T.; Virtanen, A. Secondary Organic Aerosol Formation from Healthy and Aphid-Stressed Scots Pine Emissions. *ACS Earth Sp. Chem.* **2019**, *3* (9), 1756–1772. https://doi.org/10.1021/ACSEARTHSPACECHEM.9B00118/ASSET/IMAGES/L

ARGE/SP9B00118_0011.JPEG.

(39) Norwood, M. J.; Louchouarn, P.; Kuo, L. J.; Harvey, O. R. Characterization and Biodegradation of Water-Soluble Biomarkers and Organic Carbon Extracted from Low Temperature Chars. *Org. Geochem.* **2013**, *56*, 111–119. https://doi.org/10.1016/J.ORGGEOCHEM.2012.12.008.

(40) Cai, J.; Zeng, X.; Zhi, G.; Gligorovski, S.; Sheng, G.; Yu, Z.; Wang, X.; Peng, P. Molecular Composition and Photochemical Evolution of Water-Soluble Organic Carbon (WSOC) Extracted from Field Biomass Burning Aerosols Using High-Resolution Mass Spectrometry. *Atmos. Chem. Phys.* **2020**, *20* (10), 6115–6128. https://doi.org/10.5194/ACP-20-6115-2020.

(41) Jardine, A. B.; Jardine, K. J.; Fuentes, J. D.; Martin, S. T.; Martins, G.; Durgante, F.; Carneiro, V.; Higuchi, N.; Manzi, A. O.; Chambers, J. Q. Highly Reactive Light-Dependent Monoterpenes in the Amazon. *Geophys. Res. Lett.* **2015**, *42* (5), 1576–1583. https://doi.org/10.1002/2014GL062573.

(42) Simoneit, B. R. T.; Schauer, J. J.; Nolte, C. G.; Oros, D. R.; Elias, V. O.; Fraser, M. P.; Rogge, W. F.; Cass, G. R. Levoglucosan, a Tracer for Cellulose in Biomass Burning and Atmospheric Particles. *Atmos. Environ.* **1999**, *33* (2), 173–182. https://doi.org/10.1016/S1352-2310(98)00145-9.

(43) Oros, D. R.; Simoneit, B. R. T. Identification and Emission Factors of Molecular Tracers in Organic Aerosols from Biomass Burning Part 2. Deciduous Trees. *Appl. Geochemistry* **2001**, *16* (13), 1545–1565. https://doi.org/10.1016/S0883-2927(01)00022-1.

(44) Al-Naiema, I. M.; Stone, E. A. Evaluation of Anthropogenic Secondary Organic Aerosol Tracersfrom Aromatic Hydrocarbons. *Atmos. Chem. Phys.* **2017**, *17* (3), 2053–2065. https://doi.org/10.5194/ACP-17-2053-2017.

(45) Al-Naiema, I. M.; Offenberg, J. H.; Madler, C. J.; Lewandowski, M.; Kettler, J.; Fang, T.; Stone, E. A. Secondary Organic Aerosols from Aromatic Hydrocarbons and Their Contribution to Fine Particulate Matter in Atlanta, Georgia. *Atmos. Environ.* **2020**, *223*, 117227. https://doi.org/10.1016/J.ATMOSENV.2019.117227.

(46) Wang, Q.; He, X.; Zhou, M.; Huang, D. D.; Qiao, L.; Zhu, S.; Ma, Y. G.; Wang, H. L.; Li, L.; Huang, C.; Huang, X. H. H.; Xu, W.; Worsnop, D.; Goldstein, A. H.; Guo, H.; Yu, J. Z.; Huang, C.; Yu, J. Z. Hourly Measurements of Organic Molecular Markers in Urban Shanghai, China: Primary Organic Aerosol Source Identification and Observation of Cooking Aerosol Aging. *ACS Earth Sp. Chem.*

2020, *4* (9), 1670–1685.
https://doi.org/10.1021/ACSEARTHSPACECHEM.0C00205/ASSET/IMAGES/LARGE/SP0C00205_0013.JPEG.

5.8 Tables and Figures

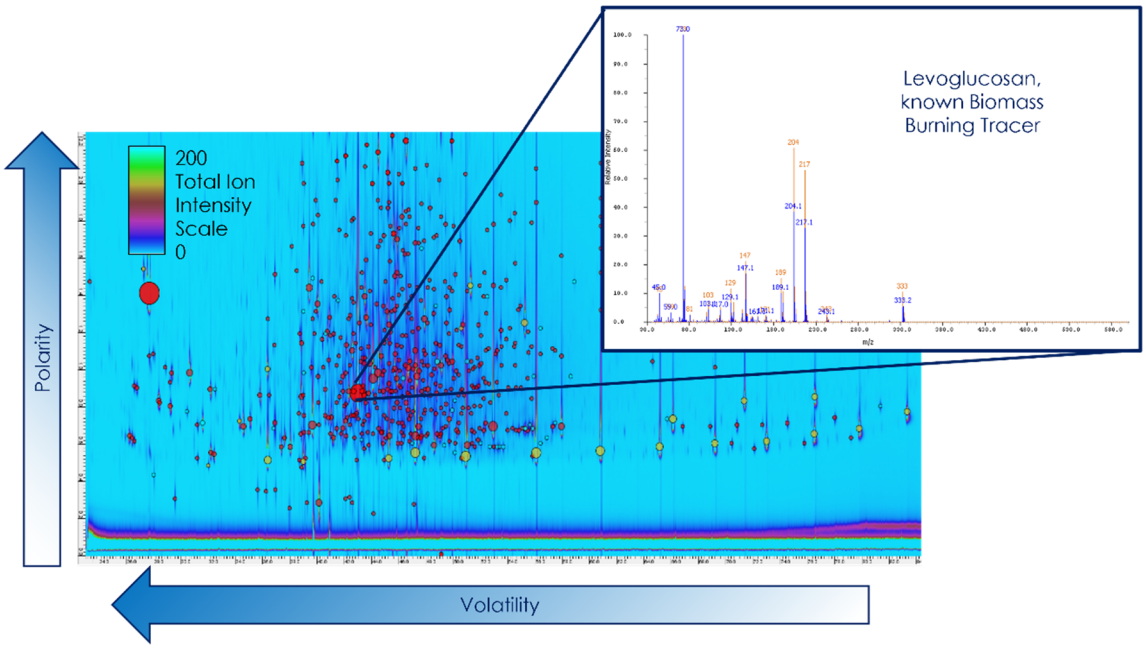


Figure 5.1. GCxGC chromatogram, of submicron organic aerosol collected at the T3 GoAmazon field site from 6:00-18:00 March 17, 2014. Analytes are separated by both volatility and polarity with two GC columns in sequence, with decreasing volatility in the x dimension and increasing polarity in the y dimension. Circles indicate detected and traced compounds, with size proportional to compound volume. Red indicates detected and traced analytes, yellow indicates internal standards, and blue indicates field blank contaminations that were removed from analysis. Each detected compound is characterized by a structure-specific 70 eV EI mass spectrum, as illustrated for levoglucosan.

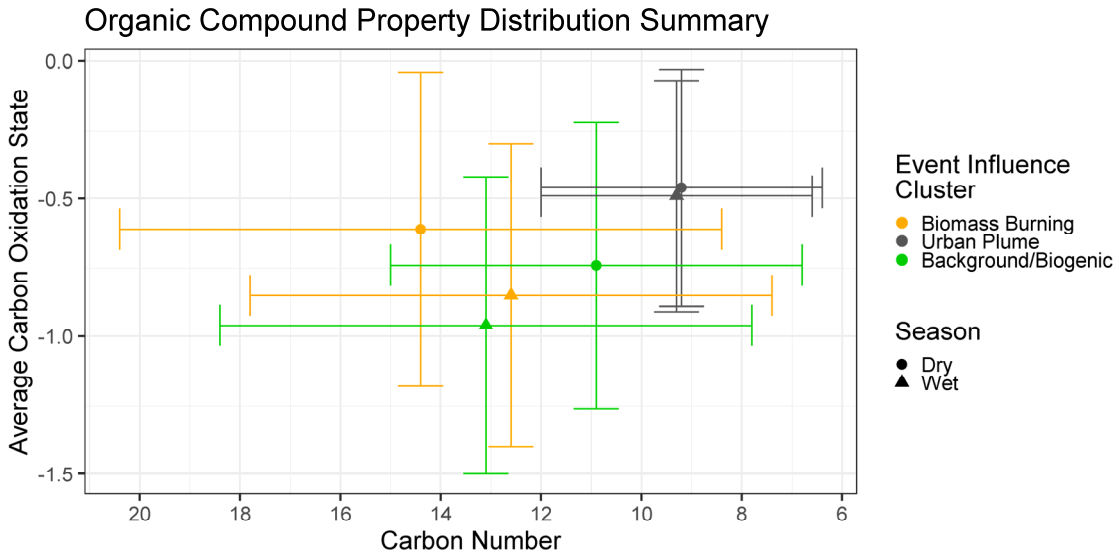


Figure 5.2. Summary of chemical properties of organic compounds speciated and traced over the wet and dry seasons of the GoAmazon field campaign. Marker positions indicate the unweighted average properties of each population of compounds in Kroll diagram space, and error bars indicate the standard deviation of each property for each grouping. Compounds are assigned to event influence clusters of biomass burning (orange), urban plume influenced (grey) and background/biogenic (green).

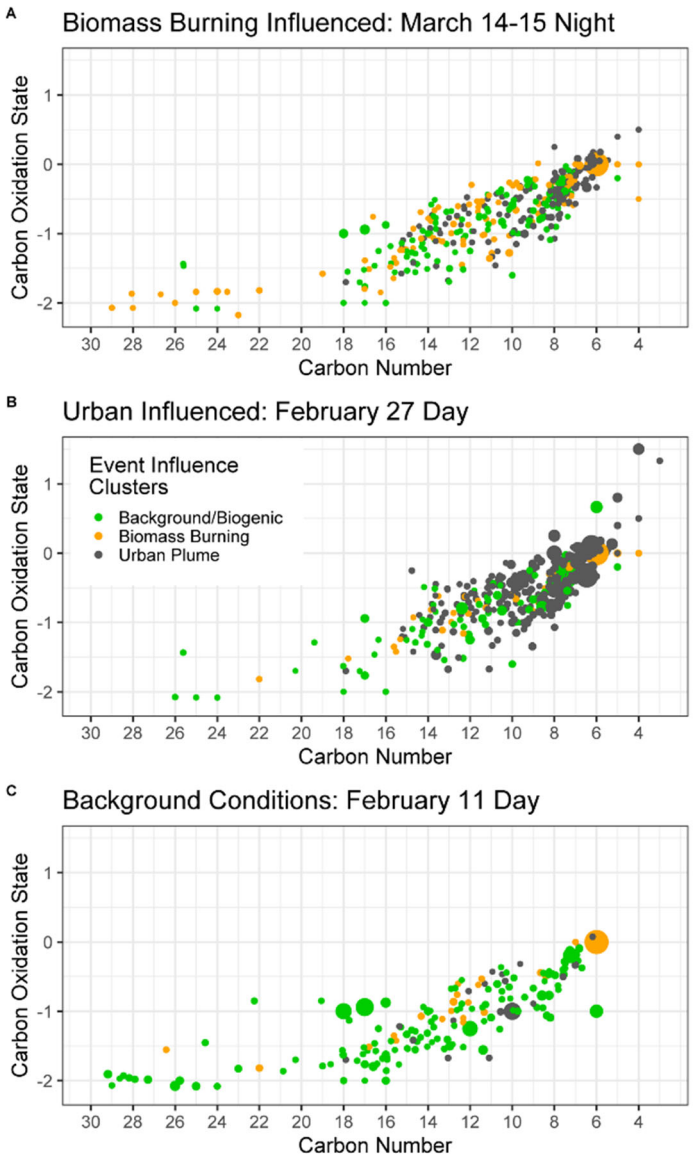


Figure 5.3. Example compound properties distributions in Kroll diagram space for fire influenced (panel A), urban influenced (panel B), and pristine conditions (panel C) of submicron organic aerosol collected during the wet season of the 2014 GoAmazon field campaign. Each individual compound is represented by a circle whose size is proportional to its concentration.

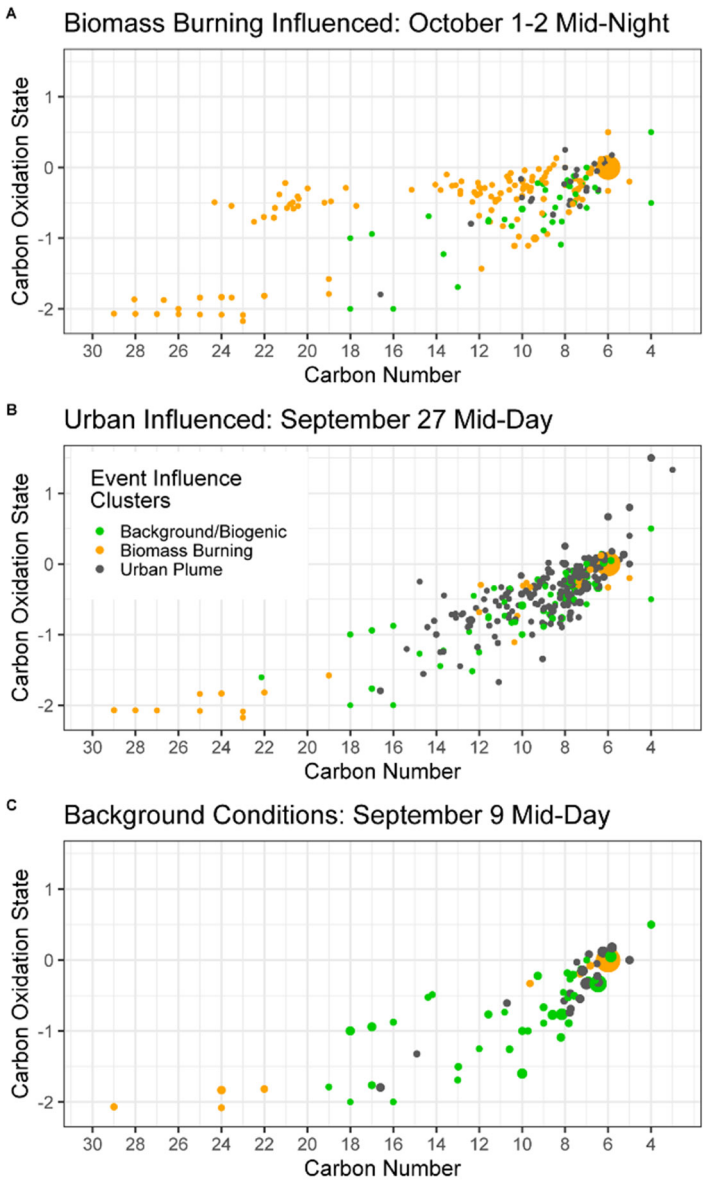


Figure 5.4. Example compound properties distributions in Kroll diagram space for fire influenced (panel A), urban influenced (panel B), and pristine conditions (panel C) of submicron organic aerosol collected during the dry season of the 2014 GoAmazon field campaign. Each individual compound is represented by a circle whose size is proportional to its concentration.

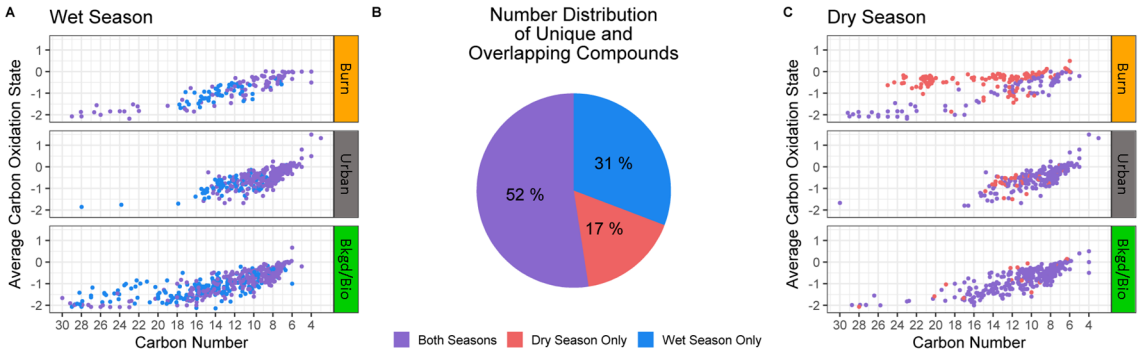


Figure 5.5. Compositional analysis of seasonally unique (exclusively observed in either the wet or dry seasons) and commonly observed organic compounds from GCxGC analysis of submicron aerosol collected during the GoAmazon field campaign. Panel A illustrates the properties distributions of wet season compounds in $\overline{OS_c} - n_c$ space segregated by the event source clusters to which those compounds were assigned, with unique compounds indicated in blue and commonly observed products illustrated in purple. Panel B illustrates the fractional contribution of wet season unique, dry season unique, and consistently observed species to the total number of compounds traced. Panel C illustrates the properties distributions of dry season compounds in $\overline{OS_c} - n_c$ space segregated by the event source clusters to which those compounds were assigned.

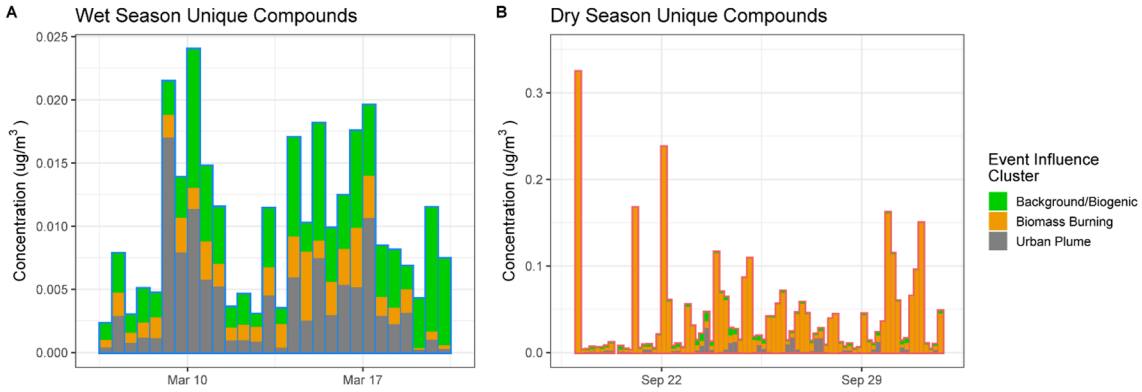


Figure 5.6. Concentration time series of summed seasonally unique organic aerosol compounds observed in the GoAmazon field campaign during the continuous analysis periods of the wet season (Panel A, outlined in blue) and dry season (Panel B, outlined in red). Mass attributed to biomass burning event influence is indicated in orange, mass attributed to urban influence is indicated in grey, and mass attributed to background/biogenic species is indicated in green.

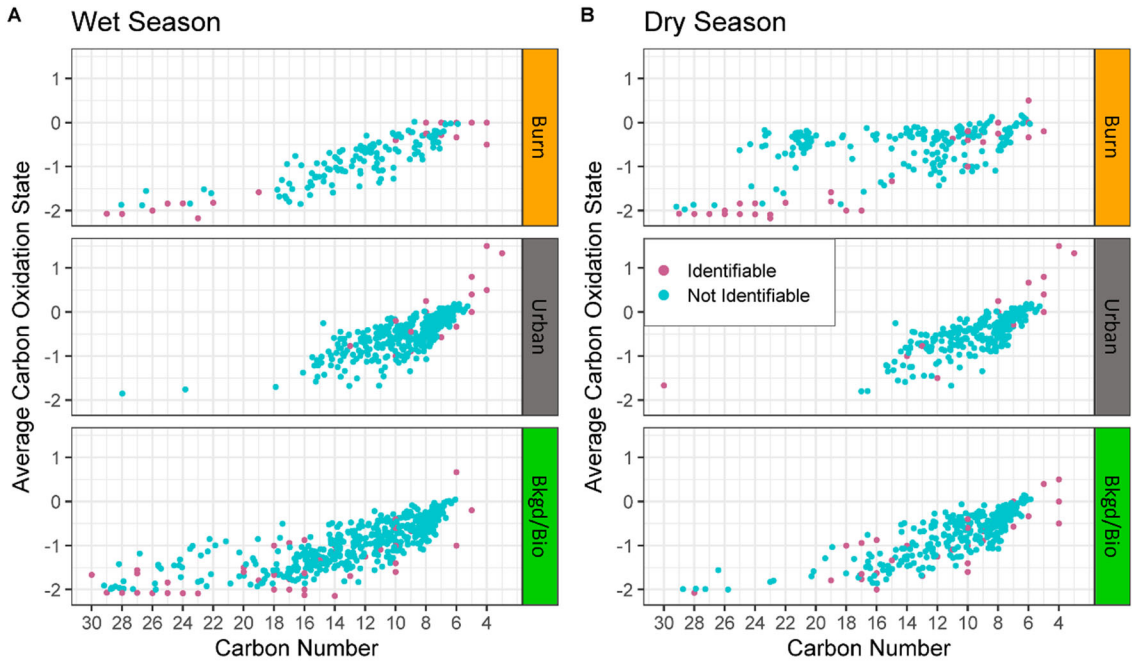


Figure 5.7. Chemical properties distributions in Kroll diagram space of identifiable (pink) and not identifiable (teal) organic compounds identified in submicron aerosol collected at the GoAmazon field campaign. Panel A illustrates compounds observed in the wet season, vertically segregated by attribution to burning influenced, urban influenced, and background/biogenic event source groups. Panel B illustrates compounds observed in the dry season, similarly vertically segregated.

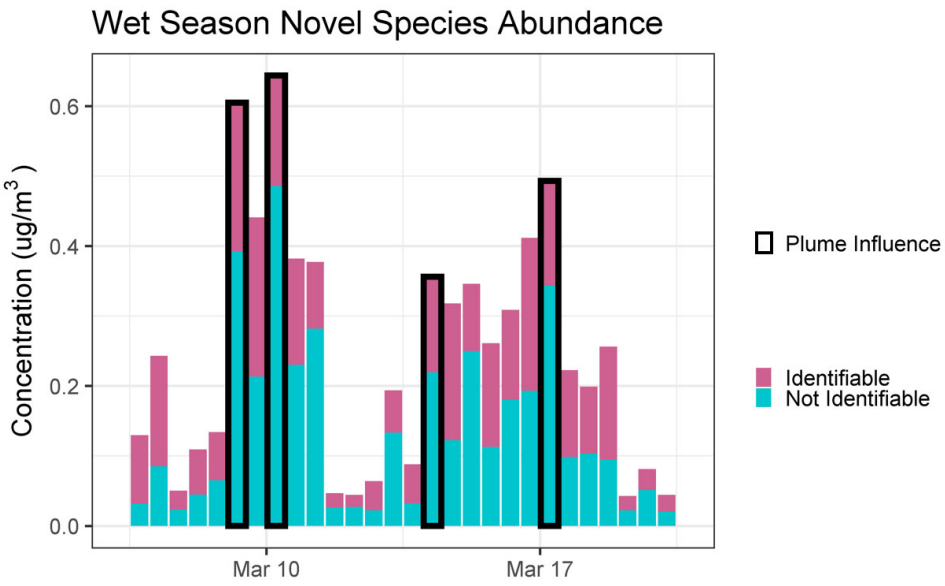


Figure 5.8. Contributions of identifiable (pink) and not identifiable (teal) compounds to submicron organic aerosol concentration during the wet season during the GoAmazon field campaign in 2014. Days significantly impacted by emissions from the urban plume are outlined in black, with "plume influence" defined as periods when the urban influence tracer was measured at >1 standard deviation above its mean in the wet season.

Table 5.1. Summary of template samples used in compilation of custom mass spectral library for GCxGC analysis of submicron organic aerosol samples collected in the GoAmazon field campaign.

Season	Sample Collection Period	Description
Wet	2/9/2014 6:00 – 18:00	Urban Impacted
Wet	2/13/2014 6:00- 18:00	Low Pollution
Wet	2/16/2014 6:00- 18:00	Low Pollution
Wet	2/16/2014 18:00- 2/17/2014 6:00	Low Pollution
Wet	3/9/2014 6:00-18:00	Urban and Burn Impacted
Wet	3/14/2014 6:00-18:00	Urban and Burn Impacted
Wet	3/17/2014 6:00-18:00	Urban Impacted
Dry	9/8/2014 6:00-18:00	Low Pollution
Dry	9/14/2014 2:00- 6:00	Urban Impacted
Dry	9/23/2014 14:00 – 18:00	Burn Impacted
Dry	9/27/2014 14:00 – 18:00	Burn Impacted

Table 5.2. Example identifiable tracer compounds assigned to source groups created from hierarchical clustering of GCxGC speciated analysis of submicron aerosol collected during the GoAmazon field campaign.

Assigned Event Source Group	Compound	Notes	Reference
Biomass Burning	Levoglucosan	Cellulose combustion	Bhattarai et al., 2019 Simoneit et al., 1999
	Tetracosanoic acid		Oros and Simoneit, 2001
Urban Emissions Plume	Phthalic Acid	Oxidation product of polycyclic aromatic hydrocarbons	Al-Naiema and Stone, 2017 Al-Naiema et al., 2020
	Azelaic acid	Cooking aerosol tracer	Wang et al., 2020
Background/Biogenic	Pinic Acid	Monoterpene oxidation product	Christoffersen et al., 1998

6 Conclusions and Future Work

In this work, anthropogenic influences on two classes of natural aerosol, coastal marine aerosol and tropical organic aerosol, are compositionally characterized with an aim of contributing to the body of knowledge required to advance analysis and prediction of aerosol impacts on public health and climate. Given the high degree of chemical complexity of organic aerosol composition, the development of new methodologies that take advantage of the full scope of information produced by advances in instrumentation was embraced, resulting in the development of the Ch3MS-RF model, which has broad applications throughout environmental chemistry. In Chapter 1, the motivation of this work, namely the importance of organic aerosols for climate and public health, knowledge gaps in ambient organic aerosol source composition and environmental transformations, are described, and the current state of knowledge on marine and tropical biogenic aerosols are introduced. In Chapter 2, the Ch3MS-RF model is introduced, described, and evaluated, specifically for the application of predicting properties of tropical organic aerosol species. This model predicts atmospherically relevant properties of organic compounds based only on their mass spectra and retention indices, making it applicable data produced by widely utilized GC-MS instrumentation, as well as GCxGC-MS. This model is utilized in Chapters 3 and 5 to investigate human influences on costal sea spray aerosol and tropical organic aerosol, respectively. The analysis of coastal sea spray aerosol presented in Chapter 3 finds that anthropogenic pollutants, in particular personal care products, oils, and PAHs, contribute substantially to the organic content of sea spray aerosol in coastal regions, but that microbiological blooms have the capacity to transform this organic content, producing biogenic and biologically transformed products that are largely missing from current mass spectral databases. Chapter 4 focuses on one specific class of anthropogenic marine pollutants, the benzothiazoles, and finds that these species are emitted from coastal waters in both gas and aerosol phases, emissions that cannot be fully characterized by or predicted from concentrations observed dissolved in ocean water. This chapter additionally explores the atmospheric oxidation chemistry of benzothiazole, finding that it has the capacity to contribute to secondary aerosol formation and the coastal marine sulfur budget. In Chapter 5, anthropogenic influences on organic aerosol in the Amazon Rainforest are investigated during both the wet and dry seasons. Findings indicate that fires and urban pollution impact organic composition differently, but that urban influences on aerosol composition remain compositionally consistent season to season while fires produce seasonally unique and compositionally distinct compounds. In this final chapter of the dissertation, key opportunities for future advances in each of this work's main research thrusts are outlined. The primary priorities suggested here include future developments of Ch3MS-RF and speciated atmospheric chemistry data science more generally, increased characterization of organic emissions from ambient and simulated polluted marine environments, and development of oxidation experiments to more closely simulate the conditions of ambient tropical oxidation chemistry to elucidate critical aerosol formation and ageing processes.

6.1 Expanded Applications of Machine Learning in Speciated Atmospheric Chemistry Data Analysis

As is true for most complex environmental mixtures,[1,2] the majority of compounds separated and detected from the marine and tropical organic aerosol material reported in Chapters 3 and 5 are not identifiable via comparisons to available mass spectral libraries. To address this challenge, in Chapter 2 this work presents Ch3MS-RF, a machine learning model for predicting the chemical properties of unidentifiable organic aerosol constituents. Four properties, specifically O:C ratio, volatility, average carbon oxidation state, and carbon number were modeled with success. Accurate prediction of these properties represents significant opportunities in atmospheric chemistry, as they allow compound populations to be represented and visualized in Kroll diagram[3] and volatility basis set[4] feature spaces, which are broadly utilized in the atmospheric chemistry community and particularly useful for chemical modelling applications.[5] Given that Ch3MS-RF predicts vapor pressures with greater accuracy than formula-based parameterizations, volatility basis set parameterizations using GC-MS and Ch3MS-RF are likely to be more accurate than those using methods that characterize organic composition by formulae, for example by soft ionization techniques.

The number of properties that can be predicted by Ch3MS-RF is not limited to those included in Chapter 2 however, and future developments to explore predictive capabilities across additional property spaces represent significant opportunities for growth. For example, prediction of molecular weight when coupled with the established volatility prediction capabilities would enable compounds to be characterized by the molecular corridors framework developed and described in Li et al., 2016 and Shiraiwa et al., 2014.[6,7] This framework has been identified as particularly useful in tracing the evolution of secondary aerosol constituents from volatile precursors. Random forest modeling is not limited to continuous variable prediction; the methodology is adaptable to categorization, and with very limited adjustments could be used to identify groupings, for example by chemical functional groups (acid, aldehyde, aromatic, etc). While initial efforts at functional group categorization stalled at early stages due to challenges related to how compounds with multiple functional group types should be categorized, binary predictions of whether or not a given compound contains a functional group of interest have shown significant promise. Successful implementation of functional group prediction would be useful for a variety of applications, including identifying oxidant-specific product classes and improving comparisons between chamber and ambient samples to yield additional information about the product classes that are currently not reproduced under controlled laboratory conditions, such as the seasonally unique Amazonian organic aerosol compounds discussed in Chapter 5.

In addition to expanding the capabilities and applications of Ch3MS-RF, the atmospheric chemistry community can significantly benefit from adapting other chemistry focused machine learning and data science-focused platforms pioneered in other subfields. One highly applicable example of this is the Global Natural Products Social Molecular Networking (GNPS) platform, which was initially developed for metabolomics

applications and is based on a neural networking methodology.[8] This tool goes beyond mass spectral matching for identification purposes to create associations or families of molecules based on spectral similarities. In atmospheric applications, this methodology could yield critical chemical categorization of unidentifiable atmospheric organics, including identifying groupings of products from a single precursor in the case of secondary aerosol formation or clusters of similarly sourced contributions to sea spray aerosol composition. This tool is likely to be of particular use in organizing and identifying key functional relationships within UCB-GLOBES, the database of ambient and experimentally produced organic compounds observed across space and time described in Chapters 3 and 5. When combined with the time series based groupings utilized within this work, which group compounds by similar event sources, compositional groupings may enable future analysis to determine not only what events cause emissions of any given compound of interest but also the precursor or precursor class that contributed the mass. This capability would mark a significant step forward in speciated atmospheric organic analysis and increase the mechanistic insights that can be gained from ambient measurements.

6.2 Marine Pollution Aerosolization in Coastal Environments

Chapters 3 and 4 describe the emissions of organic pollutants from the coastal ocean, with an emphasis on chemical composition of primary sea spray aerosol. Chapter 3, which focuses on a single pollutant class, the benzothiazoles, finds that the concentrations and speciation of organic pollutants observed in nascent SSA are not reflective of the dissolved organic phase. This supports previous literature, which has found that many organic pollutants become highly concentrated in the sea surface microlayer, which leads to significant enrichments in sea spray aerosol compared to bulk water.[9–12] Walsh et al., 2017 has proposed sea spray aerosols as an important route of exposure for coastal communities to become exposed to hazardous materials including pesticides, heavy metals and phthalates.[13] However, as noted in that work, epidemiological evidence for aerosol exposure is difficult to establish due to the fact that coastal communities chronically exposed to coastal sea spray aerosol also often consume local fish, another well established and more easily monitored route of exposure. Chronic human exposure to anthropogenic pollutants via sea spray aerosol inhalation is beyond the scope of this work and the chamber sea spray aerosol production methodology presented here. However, the findings presented in this work, particularly those discussed in Chapter 3, indicate that additional investigations of the mechanics of pollutant transfer to primary sea spray aerosol warrant future inquiry, as do ambient observations of sea spray aerosol composition in more polluted coastal contexts.

A summary of hazardous marine pollutants that have been reported in coastal ocean water is presented in Table 1, along with the health impacts associated with chronic exposures to those pollutants. Some have been previously measured in sea spray aerosol, but measurements of the speciated organic composition of sea spray aerosol remain sparse, and investigations of the health effects of exposures to pollutants via sea spray aerosols even more so. For example, the disposal of large quantities of DDT off the coast of Los

Angeles has been identified as a significant source of pesticides to the ocean floor, leading to sediment contamination levels ~40 times those observed at highly impacted surface superfund sites.[14] A speciated analysis of sea spray aerosol composition at the coast of Los Angeles or on the Channel Islands with pesticide and pesticide metabolite detection capabilities could yield critical context to the marine aerosol DDT exposure mechanism proposed in Walsh et al., 2017 and provide coastal communities with important health-relevant information. Priorities for future sea spray aerosol organic analysis include pesticides, pharmaceuticals, and personal care products (including UV filters). Enrichment of the sea surface microlayer with anthropogenic compounds also has implications for marine microbiology and the emissions of biogenic compounds from the ocean, as the sea surface microlayer has its own biologically rich ecosystem.[9] The interactions between anthropogenic pollutants and microbiology in the sea surface microlayer and resulting impacts on sea spray aerosol composition are an important area of future environmental engineering research, with implications for public health and atmospheric chemistry in coastal regions.

Table 6.1. Classes of organic pollutants previously observed in coastal ocean water and/or sea spray aerosols, along with current public health concerns related to chronic exposure to those pollutants.

Compound Class	Observations in Coastal Oceans Water	Observations in SSA	Health Impacts
PAHs	Frias 2010[15], Manodori 2006[16]	Chapter 3, Cincinelli 2001[10]	Carcinogenic[17]
Benzothiazoles	Chapter 4, Liao 2018[18]	Chapter 4, Chapter 3	Carcinogenic[18]
Phthalates	Walsh 2017[13]	Chapter 3, Cincinelli 2001[10]	Endocrine disruptors[19]
Pesticides	Cox 1972, Frias 2010[15]		Carcinogenic, neurotoxic[20]
UV Filters	Tovar-Sanchez 2013[21]	Chapter 3	Potential endocrine distruptors[22]
Siloxanes	Hong, 2014[23]	Chapter 3	
Pharmaceuticals	Gaw 2014[24]		Antibiotic resistance (indirect), incompletely characterized[25]
Heavy Metals	Walsh 2017[13]	Li 2018[26]	Cognitive delays, neurocognitive disorders, behavioral issues, respiratory problems, cancer[27]
PCB's	Frias 2010[15], Manodori 2006[16]		Endocrine disruptor[28]

6.3 Reproducing Tropical Secondary Aerosol Formation and Perturbations in a Laboratory

In Chapter 5, the influences of fires and urban emissions on organic aerosol in the central Amazon are traced and compositionally characterized, revealing shifts in product formation and seasonally unique compound populations formed under pristine and pollution impacted conditions. One key finding limiting the ability to gain additional mechanistic insights from the seasonally unique and source-specific species is the novelty of the uniquely produced products, which are largely not currently listed in mass spectral databases, making them nearly impossible to definitively identify. Additional exploratory research has been conducted to attempt to expand characterization and source attribution of the unidentifiable compounds, specifically through analyzing aerosols produced by single precursor oxidation studies of monoterpenes and sesquiterpenes. The results of this comparative work are limited but informative; while many unidentifiable particle-phase compounds were formed in the chamber oxidation experiments, few overlapped with those produced in the Amazon samples. Specifically, only 48 out of 1378 unidentifiable spectra from the Amazon matched compounds produced by oxidation experiments utilizing alpha pinene, limonene, beta caryophyllene, and myrcene (each individually, not in conjunction), all of which are significant terpene emissions in the Amazon.[29,30] None of the unidentifiable Amazon compounds reproduced in chamber oxidation experiments belong to the dry season unique high carbon number biomass burning group or the seasonally unique wet season biogenic populations highlighted in Figure 5.5.

Replicating unique compounds observed in ambient conditions through laboratory oxidation experiments is critical to gaining additional mechanistic insights into how secondary aerosol is formed in the atmosphere. However, doing so requires replicating real ambient oxidation conditions in ways that can be difficult to achieve and are not commonly utilized. As discussed in Porter et al., 2021,[31] the conditions typically used in controlled laboratory oxidation experiments are concentrated in a property space rarely seen in the atmosphere, namely dry, 'room temperature,' and highly concentrated with both precursors and oxidants. For the specific example of better replicating Amazonian conditions and reproducing unique anthropogenic perturbation-associated species, conducting experiments under high humidity conditions and expanding analysis into the role of aqueous processing are likely to prove particularly important. In coordination with the expanded data science methods described in section 6.1, designing chamber oxidation experiments to replicate ambient conditions of interest and analyzing products formed from those experiments identically to ambient samples will enable significant advances in our understanding of the processes by which human emissions alter biogenic secondary aerosol formation in the tropics.

6.4 References

(1) Worton, D. R.; Decker, M.; Isaacman-VanWertz, G.; Chan, A. W. H.; Wilson, K. R.; Goldstein, A. H. Improved Molecular Level Identification of Organic Compounds Using Comprehensive Two-Dimensional Chromatography, Dual Ionization Energies and High Resolution Mass Spectrometry. *Analyst* **2017**, *142* (13), 2395–2403. https://doi.org/10.1039/c7an00625j.

(2) Hamilton, J. F.; Webb, P. J.; Lewis, A. C.; Hopkins, J. R.; Smith, S.; Davy, P. Atmospheric Chemistry and Physics Partially Oxidised Organic Components in Urban Aerosol Using GCXGC-TOF/MS. *Atmos. Chem. Phys* **2004**, *4*, 1279–1290.

(3) Kroll, J. H.; Donahue, N. M.; Jimenez, J. L.; Kessler, S. H.; Canagaratna, M. R.; Wilson, K. R.; Altieri, K. E.; Mazzoleni, L. R.; Wozniak, A. S.; Bluhm, H.; Mysak, E. R.; Smith, J. D.; Kolb, C. E.; Worsnop, D. R. Carbon Oxidation State as a Metric for Describing the Chemistry of Atmospheric Organic Aerosol. *Nat. Chem. 2010 32* **2011**, *3* (2), 133–139. https://doi.org/10.1038/nchem.948.

(4) Donahue, N. M.; Robinson, A.; Stanier, C. O.; Pandis, S. N. Coupled Partitioning, Dilution, and Chemical Aging of Semivolatile Organics. *Environ. Sci. Technol.* **2006**, *40* (8), 2635–2643. https://doi.org/10.1021/ES052297C.

(5) Shrivastava, M.; Fast, J.; Easter, R.; Gustafson, W. I.; Zaveri, R. A.; Jimenez, J. L.; Saide, P.; Hodzic, A. Modeling Organic Aerosols in a Megacity: Comparison of Simple and Complex Representations of the Volatility Basis Set Approach. *Atmos. Chem. Phys.* **2011**, *11* (13), 6639–6662. https://doi.org/10.5194/ACP-11-6639-2011.

(6) Shiraiwa, M.; Berkemeier, T.; Schilling-Fahnestock, K. A.; Seinfeld, J. H.; Pöschl, U. Molecular Corridors and Kinetic Regimes in the Multiphase Chemical Evolution of Secondary Organic Aerosol. *Atmos. Chem. Phys.* **2014**, *14* (16), 8323–8341. https://doi.org/10.5194/ACP-14-8323-2014.

(7) Li, Y.; Pöschl, U.; Shiraiwa, M. Molecular Corridors and Parameterizations of Volatility in the Chemical Evolution of Organic Aerosols. *Atmos. Chem. Phys.* **2016**, *16* (5), 3327–3344. https://doi.org/10.5194/ACP-16-3327-2016.

(8) Nothias, L. F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M.; Aicheler, F.; Aksenov, A. A.; Alka, O.; Allard, P. M.; Barsch, A.; Cachet, X.; Caraballo-Rodriguez, A. M.; Da Silva, R. R.; Dang, T.; Garg, N.; Gauglitz, J. M.; Gurevich, A.; Isaac, G.; Jarmusch, A. K.; Kameník, Z.; Kang, K. Bin; Kessler, N.; Koester, I.; Korf, A.; Le Gouellec, A.; Ludwig, M.; Martin H, C.; McCall, L. I.; McSayles, J.; Meyer, S. W.; Mohimani, H.; Morsy, M.; Moyne, O.; Neumann, S.; Neuweger, H.; Nguyen, N. H.; Nothias-Esposito, M.; Paolini, J.; Phelan, V. V.; Pluskal, T.; Quinn, R. A.; Rogers, S.; Shrestha, B.; Tripathi, A.; van der Hooft, J. J. J.; Vargas, F.; Weldon, K. C.; Witting, M.; Yang, H.; Zhang, Z.; Zubeil, F.; Kohlbacher, O.; Böcker, S.; Alexandrov, T.; Bandeira, N.; Wang, M.; Dorrestein, P. C. Feature-Based

Molecular Networking in the GNPS Analysis Environment. *Nat. Methods 2020 179* **2020**, *17* (9), 905–908. https://doi.org/10.1038/s41592-020-0933-6.

(9) Wurl, O.; Obbard, J. P. A Review of Pollutants in the Sea-Surface Microlayer (SML): A Unique Habitat for Marine Organisms. *Mar. Pollut. Bull.* **2004**, *48* (11–12), 1016–1030. https://doi.org/10.1016/J.MARPOLBUL.2004.03.016.

(10) Cincinelli, A.; Stortini, A. M.; Perugini, M.; Checchini, L.; Lepri, L. Organic Pollutants in Sea-Surface Microlayer and Aerosol in the Coastal Environment of Leghorn—(Tyrrhenian Sea). *Mar. Chem.* **2001**, *76* (1–2), 77–98. https://doi.org/10.1016/S0304-4203(01)00049-4.

(11) Engel, A.; Bange, H. W.; Cunliffe, M.; Burrows, S. M.; Friedrichs, G.; Galgani, L.; Herrmann, H.; Hertkorn, N.; Johnson, M.; Liss, P. S.; Quinn, P. K.; Schartau, M.; Soloviev, A.; Stolle, C.; Upstill-Goddard, R. C.; van Pinxteren, M.; Zäncker, B. The Ocean's Vital Skin: Toward an Integrated Understanding of the Sea Surface Microlayer. *Front. Mar. Sci.* **2017**, *4* (MAY), 165. https://doi.org/10.3389/FMARS.2017.00165/BIBTEX.

(12) Cross, J. N.; Hardy, J. T.; Hose, J. E.; Hershelman, G. P.; Antrim, L. D.; Gossett, R. W.; Crecelius, E. A. Contaminant Concentrations and Toxicity of Sea-Surface Microlayer near Los Angeles, California. *Mar. Environ. Res.* **1987**, *23* (4), 307–323. https://doi.org/10.1016/0141-1136(87)90024-9.

(13) Walsh, J. J.; Lenes, J. M.; Weisberg, R. H.; Zheng, L.; Hu, C.; Fanning, K. A.; Snyder, R.; Smith, J. More Surprises in the Global Greenhouse: Human Health Impacts from Recent Toxic Marine Aerosol Formations, Due to Centennial Alterations of World-Wide Coastal Food Webs. *Marine Pollution Bulletin*. Elsevier Ltd March 15, 2017, pp 9–40. https://doi.org/10.1016/j.marpolbul.2016.12.053.

(14) Kivenson, V.; Lemkau, K. L.; Pizarro, O.; Yoerger, D. R.; Kaiser, C.; Nelson, R. K.; Carmichael, C.; Paul, B. G.; Reddy, C. M.; Valentine, D. L. Ocean Dumping of Containerized DDT Waste Was a Sloppy Process. *Environ. Sci. Technol.* **2019**, *53* (6), 2971–2980. https://doi.org/10.1021/ACS.EST.8B05859/ASSET/IMAGES/LARGE/ES-2018-058598_0004.JPEG.

(15) Frias, J. P. G. L.; Sobral, P.; Ferreira, A. M. Organic Pollutants in Microplastics from Two Beaches of the Portuguese Coast. *Mar. Pollut. Bull.* **2010**, *60* (11), 1988–1992. https://doi.org/10.1016/J.MARPOLBUL.2010.07.030.

(16) Manodori, L.; Gambaro, A.; Piazza, R.; Ferrari, S.; Stortini, A. M.; Moret, I.; Capodaglio, G. PCBs and PAHs in Sea-Surface Microlayer and Sub-Surface Water Samples of the Venice Lagoon (Italy). *Mar. Pollut. Bull.* **2006**, *52* (2), 184–192. https://doi.org/10.1016/J.MARPOLBUL.2005.08.017.

(17) Kim, K. H.; Jahan, S. A.; Kabir, E.; Brown, R. J. C. A Review of Airborne Polycyclic Aromatic Hydrocarbons (PAHs) and Their Human Health Effects. *Environ. Int.* **2013**, *60*, 71–80. https://doi.org/10.1016/J.ENVINT.2013.07.019.

(18) Liao, C.; Kim, U. J.; Kannan, K. A Review of Environmental Occurrence, Fate, Exposure, and Toxicity of Benzothiazoles. *Environ. Sci. Technol.* **2018**, *52* (9), 5007–5026. https://doi.org/10.1021/acs.est.7b05493.

(19) Kay, V. R.; Chambers, C.; Foster, W. G. Reproductive and Developmental Effects of Phthalate Diesters in Females. *http://dx.doi.org/10.3109/10408444.2013.766149* **2013**, *43* (3), 200–219. https://doi.org/10.3109/10408444.2013.766149.

(20) Alavanja, M. C. R.; Hoppin, J. A.; Kamel, F. Health Effects of Chronic Pesticide Exposure: Cancer and Neurotoxicity. *Annu. Rev. Public Health* **2004**, *25*, 155–197. https://doi.org/10.1146/ANNUREV.PUBLHEALTH.25.101802.123020.

(21) Tovar-Sánchez, A.; Sánchez-Quiles, D.; Basterretxea, G.; Benedé, J. L.; Chisvert, A.; Salvador, A.; Moreno-Garrido, I.; Blasco, J. Sunscreen Products as Emerging Pollutants to Coastal Waters. *PLoS One* **2013**, *8* (6), e65451. https://doi.org/10.1371/JOURNAL.PONE.0065451.

(22) Huang, Y.; Law, J. C. F.; Lam, T. K.; Leung, K. S. Y. Risks of Organic UV Filters: A Review of Environmental and Human Health Concern Studies. *Sci. Total Environ.* **2021**, *755* (Pt 1). https://doi.org/10.1016/J.SCITOTENV.2020.142486.

(23) Hong, W. J.; Jia, H.; Liu, C.; Zhang, Z.; Sun, Y.; Li, Y. F. Distribution, Source, Fate and Bioaccumulation of Methyl Siloxanes in Marine Environment. *Environ. Pollut.* **2014**, *191*, 175–181. https://doi.org/10.1016/j.envpol.2014.04.033.

(24) Gaw, S.; Thomas, K. V.; Hutchinson, T. H. Sources, Impacts and Trends of Pharmaceuticals in the Marine and Coastal Environment. *Philos. Trans. R. Soc. B Biol. Sci.* **2014**, *369* (1656). https://doi.org/10.1098/RSTB.2013.0572.

(25) Nassiri Koopaei, N.; Abdollahi, M. Health Risks Associated with the Pharmaceuticals in Wastewater. *DARU J. Pharm. Sci. 2017 251* **2017**, *25* (1), 1–7. https://doi.org/10.1186/S40199-017-0176-Y.

(26) Li, S.; Du, L.; Tsona, N. T.; Wang, W. The Interaction of Trace Heavy Metal with Lipid Monolayer in the Sea Surface Microlayer. *Chemosphere* **2018**, *196*, 323–330. https://doi.org/10.1016/J.CHEMOSPHERE.2017.12.157.

(27) Al osman, M.; Yang, F.; Massey, I. Y. Exposure Routes and Health Effects of Heavy Metals on Children. *BioMetals 2019 324* **2019**, *32* (4), 563–573. https://doi.org/10.1007/S10534-019-00193-5.

(28) Brouwer, A.; Longnecker, M. P.; Birnbaum, L. S.; Cogliano, J.; Kostyniak, P.; Moore, J.; Schantz, S.; Winneke, G. Characterization of Potential Endocrine-Related Health Effects at Low-Dose Levels of Exposure to PCBs. *Environ. Health Perspect.* **1999**, *107* (SUPPL. 4), 639–649. https://doi.org/10.1289/EHP.99107S4639.

(29) Yee, L. D.; Isaacman-Vanwertz, G.; Wernis, R. A.; Meng, M.; Rivera, V.; Kreisberg, N. M.; Hering, S. V; Bering, M. S.; Glasius, M.; Upshur, M. A.; Bé, A. G.; Thomson, R. J.; Geiger, F. M.; Offenberg, J. H.; Lewandowski, M.; Kourtchev,

I.; Kalberer, M.; De Sá, S.; Martin, S. T.; Alexander, M. L.; Palm, B. B.; Hu, W.; Campuzano-Jost, P.; Day, D. A.; Jimenez, J. L.; Liu, Y.; Mckinney, K. A.; Artaxo, P.; Viegas, J.; Manzi, A.; Oliveira, M. B.; De Souza, R.; Machado, L. A. T.; Longo, K.; Goldstein, A. H. Observations of Sesquiterpenes and Their Oxidation Products in Central Amazonia during the Wet and Dry Seasons. *Atmos. Chem. Phys* **2018**, *18*, 10433–10457. https://doi.org/10.5194/acp-18-10433-2018.

(30) Jardine, A. B.; Jardine, K. J.; Fuentes, J. D.; Martin, S. T.; Martins, G.; Durgante, F.; Carneiro, V.; Higuchi, N.; Manzi, A. O.; Chambers, J. Q. Highly Reactive Light-Dependent Monoterpenes in the Amazon. *Geophys. Res. Lett.* **2015**, *42* (5), 1576–1583. https://doi.org/10.1002/2014GL062573.

(31) Porter, W. C.; Jimenez, J. L.; Barsanti, K. C. Quantifying Atmospheric Parameter Ranges for Ambient Secondary Organic Aerosol Formation. *ACS Earth Sp. Chem.* **2021**, *5* (9), 2380–2397. https://doi.org/10.1021/ACSEARTHSPACECHEM.1C00090/SUPPL_FILE/SP1C00090_SI_003.XLSX.