Goldschmidt 2023 Abstract
https://doi.org/10.7185/gold2023.15974

# Improved understanding of photochemical processing of dissolved organic matter by using machine learning approaches

**CHEN ZHAO**[1], XINYUE XU[2], HONGMEI CHEN[3], FENGWEN WANG[4], PENGHUI LI[5], CHEN HE[6], QUAN SHI[6], DR. YUANBI YI[1], XIAOMENG LI[2], SI-LIANG LI[7] AND DING HE[1]

[1]Department of Ocean Science and the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), The Hong Kong University of Science and Technology, Hong Kong, China

[2]Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, China

[3]State Key Laboratory for Marine Environmental Science, Institute of Marine Microbes and Ecospheres, College of Ocean and Earth Sciences, College of the Environment and Ecology, Xiamen University, Xiamen, China

[4]State Key Laboratory of Coal Mine Disaster Dynamics and Control, Department of Environmental Science, Chongqing University, Chongqing, China

[5]School of Marine Sciences, Sun Yat-sen University, Zhuhai, China

[6]State Key Laboratory of Heavy Oil Processing, China University of Petroleum, Beijing, China

[7]Institute of Surface-Earth System Science, School of Earth System Science, Tianjin University, Tianjin, China

Presenting Author: czhaobk@connect.ust.hk

Dissolved organic matter (DOM) sustains a substantial part of the organic matter transported seaward, where photochemical reactions significantly affect its transformation and fate. The state-of-the-art ultra-high resolution mass spectrometry, Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS), unprecedently allows for thousands of molecular formulas assignments within one DOM sample and provide an opportunity to directly link the molecular chemical composition and photochemical reactivity. The molecules linked to photochemical processing are operationally classified as photo-resistant, photo-labile, and photo-product types according to their occurrence before and after irradiation experiments. However, there will be "label conflicts" when considering more than one irradiation experiment, which means inconsistency of the assignments of molecules into specific photochemical types. For instance, a molecule assigned as a photo-resistant molecule in one irradiation experiment can be alternatively assigned as a photo-labile or photo-product molecule in another, leading to ambiguous understanding of the photochemical reactivity of specific molecules. The inconsistency of the fate of irradiated molecules among different experiments hampered our understanding of the roles the photochemical reactions have played, which cannot be properly addressed by traditional approaches. Here, we conducted irradiation experiments for samples from two large estuaries in China. Molecules occurred in irradiation experiments were characterized by the FT-ICR MS and assigned probabilistic labels to define their photochemical reactivity. These molecules with probabilistic labels were used to construct a learning database for establishing a suitable machine learning (ML) model. We further applied our well-trained ML model to "un-matched" (i.e., not detected in our irradiation experiments) molecules from five estuaries worldwide, to predict their photochemical reactivity. Results showed that numerous molecules with strong photo-lability can be captured solely by the ML model. Moreover, comparing DOM photochemical reactivity in five estuaries revealed that the riverine DOM chemistry largely determines their subsequent photochemical reactivity. We offer an expandable and renewable approach based on ML to compatibly integrate existing irradiation experiments and shed insight into DOM transformation and degradation processes. This approach can be also expanded to biological incubation or adsorption/desorption experiments which also suffer from the "label conflicts" problems with increasing experiments.