**Division:**

Environmental Chemistry

**Session:**

Accurate Mass/High Resolution Mass Spectrometry for Environmental Monitoring & Remediation

**Title:**

Mass-suite: A novel Python package designed for mass spectrometry data analysis

**Authors:**

Ximin Hu, Derek Mar, Nozomi Suzuki, Bowei Zhang, David Beck, Edward Kolodziej

High-Resolution Mass Spectrometry (HRMS) has been widely used in chemical and biological analysis. Expanding from enhanced targeted analysis with precise results, non-targeted analysis utilized the data collection capacity of HRMS becomes more and more popular in recent years, especially in omics study and environmental analysis. However, comparing to the overwhelming data collection efficiency from the instrument, the analysis pipeline of such HRMS data for water quality assessment is still in its infancy, with many basic aspects of data reduction, analysis, and interpretation requires optimization.

Presented here is Mass-suite, a Python based open-source package that designed to better utilize HRMS data, especially for water quality assessment studies. The package provides flexible and various options to process the HRMS data: from basic functions, such as peak picking and data alignment; to advanced data analysis including statistical analysis, chemical fingerprint extraction or source quantification. Multiple algorithms were deployed within the package to enhance the data analysis performance including supervised machine learning and unsupervised clustering using Sci-kit learn Python package. The package also provides fully interactive access to the raw data, including visualization tools, formula prediction tools and MS2 online searching tool. Furthermore, the package developed in a modularized concept that different combinations of functions are available to accomplish desired tasks. The package has undergone tests on the feature detection coverage results in 99% detections out of a reference list of 400 known peaks from known environmental contaminants standard samples and proved to be work well with road water samples. Building up using the python core enables the utilization of cloud computational resource such as supercomputers or AWS, which frees up the space and cost of lab computers. By providing this package, we hope to open a new space for HRMS data analysis in environmental science field, resulting in more rapid and detailed studies in this area.