Supplementary Material:

Reviews and syntheses: Use and misuse of peak intensities from high resolution mass spectrometry in organic matter studies: opportunities for robust usage

William Kew[1], Allison Myers-Pigg[2], Christine H. Chang[2], Sean M. Colby[2], Josie Eder[1], Malak M. Tfaily[3], Jeffrey Hawkes[4], Rosalie K. Chu[1], James C. Stegen[2*]

[1]Environmental Molecular Sciences Laboratory, Richland, WA 99352, USA
[2]Pacific Northwest National Laboratory, Richland, WA 99352, USA
[3]Department of Environmental Science, University of Arizona, Tucson, AZ, 85719, USA
[4]Department of Chemistry, University of Uppsala, Uppsala, 75124, Sweden

*Correspondence to: James C. Stegen (James.Stegen@pnnl.gov)

1 Supplementary Figures
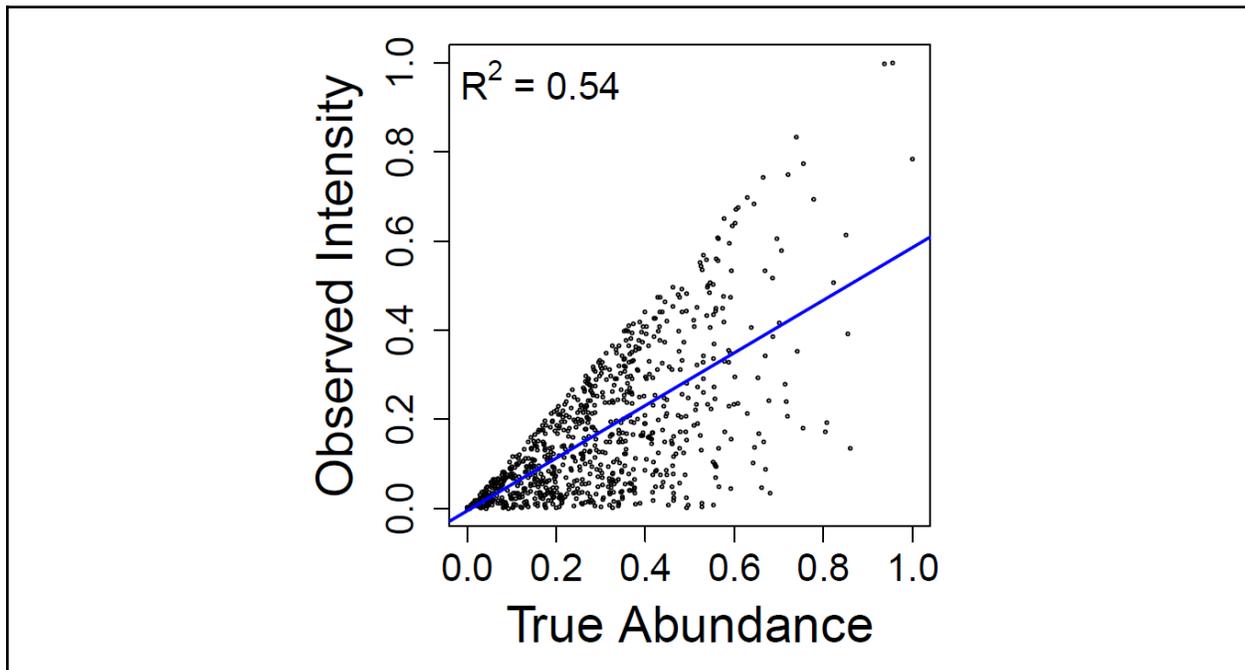
Supplementary Figure S1. Representative example of simulation model-generated estimates of observed peak intensity as a function of true abundance. Values are derived from one synthetic sample with 1000 peaks. The blue line is the linear regression model, and the associated R^2 value is near the median R^2 value for this relationship across simulations (see black line in Figure 10). The R^2 value near 0.5 indicates that error introduced in the simulation model significantly diminished the link between observed intensity and true abundance, though different simulation configurations will lead to different amounts of uncertainty.
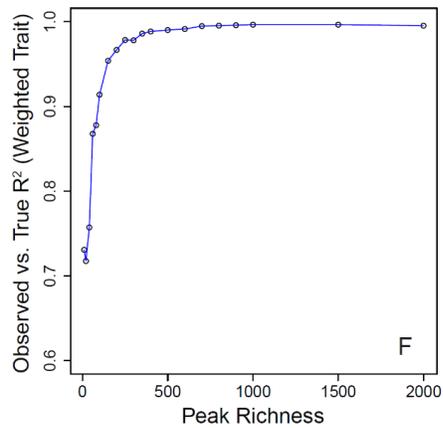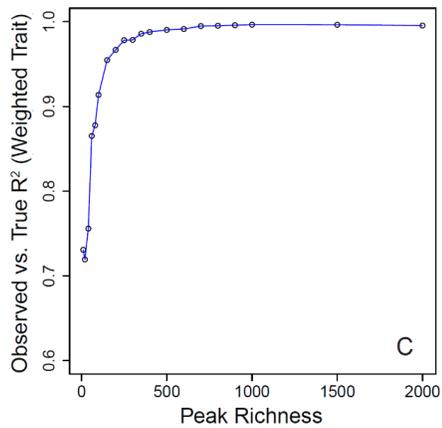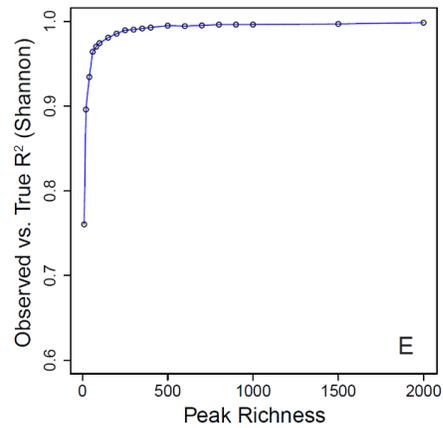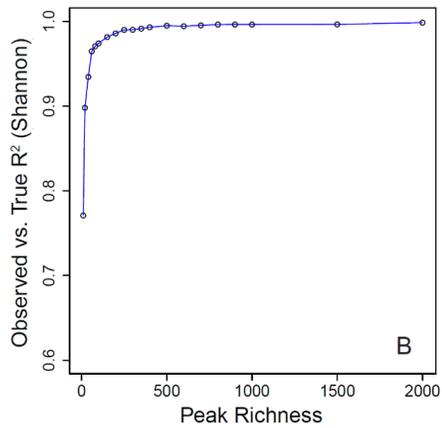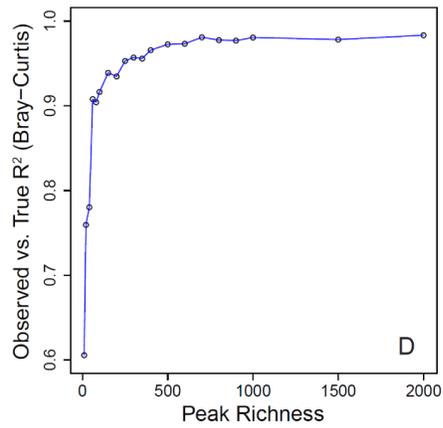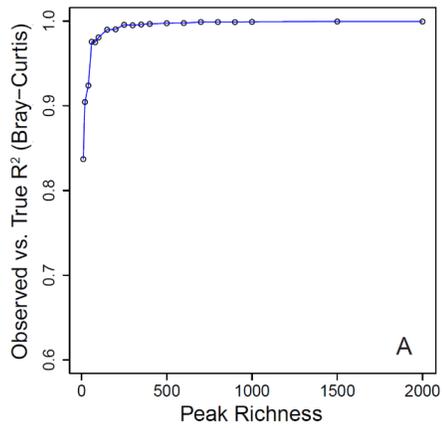
Supplementary Figure S2. The R² of the relationship between observed and true values of ecological metrics increases strongly with the number of peaks (i.e., peak richness). (A-C) The same error applied to a given peak between samples. (D-F) Different errors applied to a given peak between samples. Blue lines track the simulation model data points and are added to facilitate visual interpretation.

2 Supplementary Materials and Methods

2.1 Chemicals and Sample Preparation

Individual chemical standards (purity >95%) were acquired from IROA Technologies through the Phytochemical Metabolite Library of Standards (PHYTOMLS) and Mass Spectrometry Metabolite Library of Standards (MSMLS). Standards were initially dissolved per IROA protocol as follows: Chlorogenic acid, Cryptochlorogenic acid, Neochlorogenic acid and Ginkgolide C were all dissolved in ethanol (LCMS-grade, Optima, Thermo Scientific), Mangiferin and trehalose were dissolved in 95/5 water/methanol (v/v) (LCMS-grade, Optima, Thermo Scientific), and Aesculin, Enterodiol, Chicoric acid, Sinapic acid, and Phloridzin were all dissolved in methanol (LCMS-grade, Optima, Thermo Scientific), all at concentrations of 50 ppm (mg/l). These standards were further diluted by a stepped dilution ladder, from 5ppm to 100 ppt, into methanol (LCMS-grade, Optima, Thermo Scientific). Suwannee River Fulvic Acid (SRFA) (International Humic Substances Society, item 2S101F) was initially dissolved in LCMS grade water at a concentration of 1 mg/ml. Serial dilutions of this stock were made using LCMS grade methanol concentrations ranging from 40 ppm SRFA to 0 ppm SRFA (pure methanol).

2.2 Pure Compound Preparation (Figure 4 A-B)

Standards in methanol were analyzed on the instrument in a randomized sample and concentration order at the following concentrations 100 ppt, 200 ppt, 500 ppt, 1 ppb, 2 ppb, 5 ppb, 10 ppb, 20 ppb, 50 ppb, 100 ppb, 200 ppb, 500 ppb, 1 ppm and 2 ppm. Analyses were performed in triplicate. Two offline blanks were run after every injection. These data form main-text Figure 4 A-B.

2.3 Matrix Effect Preparations (Figure 4 C-E)

Matrix effects included 'inorganic' and 'organic' interferences. Inorganic interferences were simulated by preparation of two solid-phase extraction controls - one being a blank prepared from water (MilliQ), and the other prepared from Artificial River Water (ARW), a mineral water synthetic sample. ARW was prepared by dissolving the following into 20 liters of deionized water; 0.306 g silicic acid (Sigma Aldrich), 0.164 g potassium chloride (Fisher Chemical), 0.26 g magnesium carbonate (Fisher Chemical), 0.3 g sodium chloride (Sigma Aldrich), 1.34 g calcium sulfate (EM Science), and 3.00 g calcium carbonate (Fisher Chemical), no nitrate was added.

Both waters were solid-phase extracted as per the protocol of Dittmar et al. (2008), which includes the use of BondElut PPL (Agilent Technologies) sorbents. SPE cartridges were conditioned and equilibrated with methanol (1 ml, LCMS grade) and HCl(aq) (10mM, 1 ml), respectively. Samples were pre-acidified to pH 2 with hydrochloric acid (1M) prior to loading onto the SPE cartridge. Samples were washed with 3 x 15 ml of HCl(aq) (10mM), then dried under N2, before elution with methanol (1 ml, LCMS grade).

Organic interferences were simulated by addition of a complex organic matter standard, SRFA, at various concentrations (0 to 40 ppm). Individual standard molecules (aesculin, chlorogenic acid, enterodiol, ginkgolide C, mangiferin, phloridzin) were added to these matrices at a fixed concentration of 100 ppb from previous dilution ladder preparation.

Thus, the final samples comprise a matrix/solvent of either methanol (LCMS grade) or the product of SPE on water or ARW, in addition to a varying level of SRFA, and an individual standard compound. Samples were analyzed in triplicate in a randomized order. These samples form main-text Figure 4 C-E.

2.4 Mass Spectrometry Measurements

All data for this study was acquired on a 12 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer (FTICR MS) Bruker SolariX (Bruker, SolariX, Billerica, MA) located at PNNL in Richland, WA. The instrument

was equipped with an Infinity cell. Instrument settings were as follows: ESI source voltage +4.2kV, negative polarity, dry gas temperature 180°C, dry gas flow rate 4 l/min, ion accumulation time 50 ms, time of flight 0.65 ms. For each measurement, 144 transients of 1.67 s duration were co-added with a mass range of *m/z* 147 to 900 in a 4MW time domain, yielding a resolving power of ~400k at *m/z* 400. Samples were infused directly into the ESI source using a custom automated direct infusion cart that performed two offline blanks between each sample (Orton et al., 2018).

2.5 MS Data Analysis

Data were visually inspected using DataAnalysis (Bruker Daltonics, V5.0). Data processing was performed using CoreMS (v2.5b) (Corilo et al., 2021), a Python mass spectrometry framework, available online - https://github.com/EMSL-Computing/CoreMS, using Python v3.8. Briefly, the raw time domain data for each spectrum was loaded using the *ReadBrukerSolarix* function of CoreMS, followed by apodization and Fourier transformation. Frequency to mass conversion was performed as per a Ledford calibration using the instrument calibration constants. The mass spectra were then peak picked, and detected masses were cross referenced against the theoretical masses expected for the standard chemicals, allowing for a range of ion types including deprotonated ions, adduct ions (including Cl), and ion dimers and trimers, within an error tolerance of 2 mDa (where z=1). Identified signals were tabulated across spectra and samples for visualization and analysis using Pandas (Team, 2022) and Seaborn (Waskom, 2021) Python libraries. Peak height (apex) was used as the metric for signal intensity.

3 Supplementary References

Corilo, Y. E., Kew, W. R., and McCue, L. A.: EMSL-Computing/CoreMS: CoreMS 1.0.0, , https://doi.org/10.5281/zenodo.4641553, 2021.

Dittmar, T., Koch, B., Hertkorn, N., and Kattner, G.: A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater, Limnol. Oceanogr. Methods, 6, 230–235, https://doi.org/10.4319/lom.2008.6.230, 2008.

Orton, D. J., Tfaily, M. M., Moore, R. J., LaMarche, B. L., Zheng, X., Fillmore, T. L., Chu, R. K., Weitz, K. K., Monroe, M. E., Kelly, R. T., Smith, R. D., and Baker, E. S.: A Customizable Flow Injection System for Automated, High Throughput, and Time Sensitive Ion Mobility Spectrometry and Mass Spectrometry Measurements, Anal. Chem., 90, 737–744, https://doi.org/10.1021/acs.analchem.7b02986, 2018.

Team, T. P. D.: pandas-dev/pandas: Pandas, , https://doi.org/10.5281/ZENODO.3509134, 2022.

Waskom, M.: seaborn: statistical data visualization, J. Open Source Softw., 6, 3021, https://doi.org/10.21105/joss.03021, 2021.