Supplement of Biogeosciences, 21, 4665–4679, 2024
https://doi.org/10.5194/bg-21-4665-2024-supplement


Biogeosciences Open Access EGU

*Supplement of*

# Reviews and syntheses: Opportunities for robust use of peak intensities from high-resolution mass spectrometry in organic matter studies

**William Kew et al.**

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

## 1 Supplementary Figures and Table

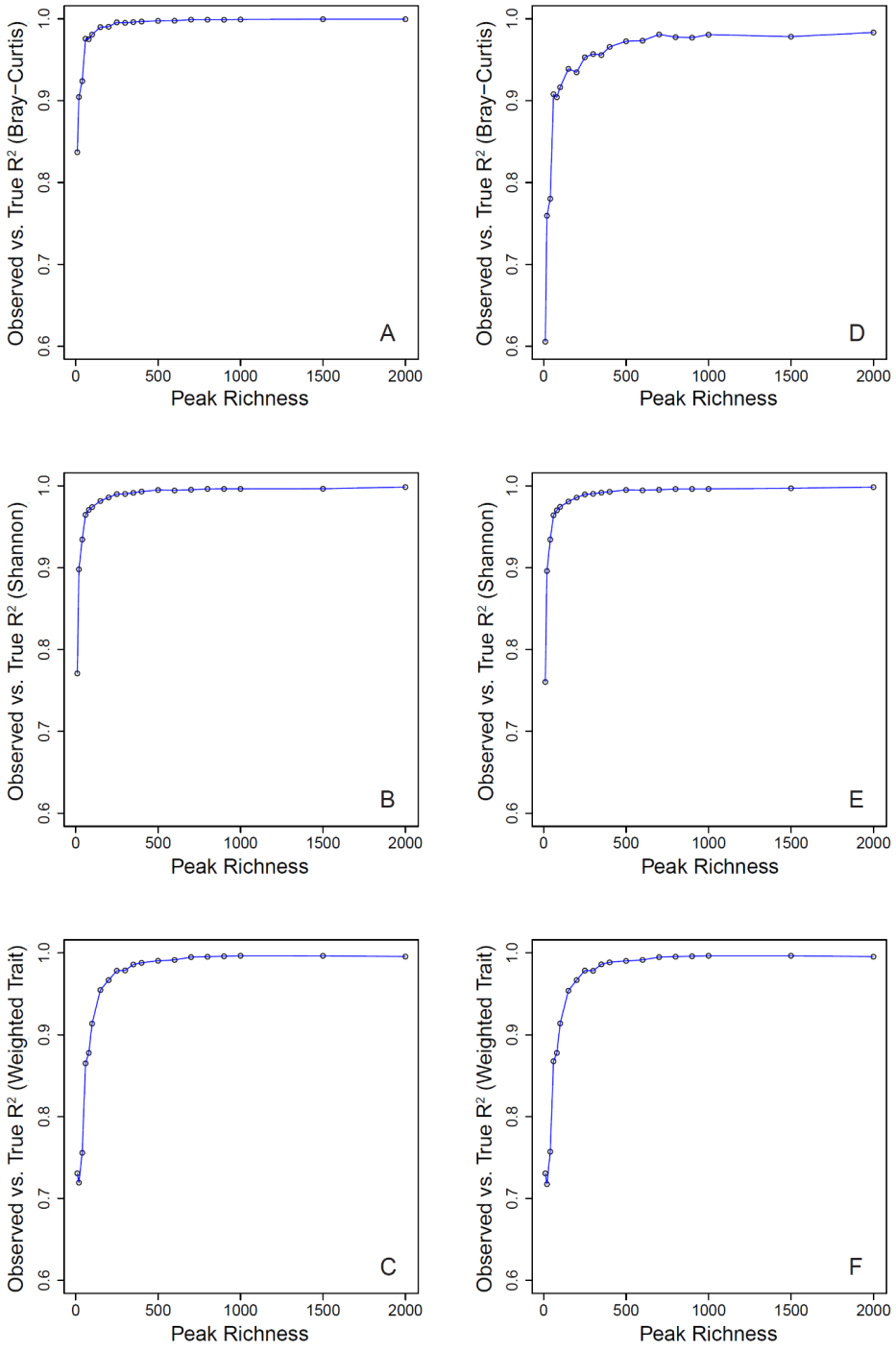


**Figure S1. The $R^2$ of the relationship between observed and true values of ecological metrics increases strongly with the number of peaks (i.e., peak richness).** (A-C) The same error applied to a given peak between samples, referred to as within-sample error in the main manuscript. (D-F) Different errors applied to a given peak between samples, referred to as between-sample error in the main manuscript. Blue lines track the simulation model data points and are added to facilitate visual interpretation.

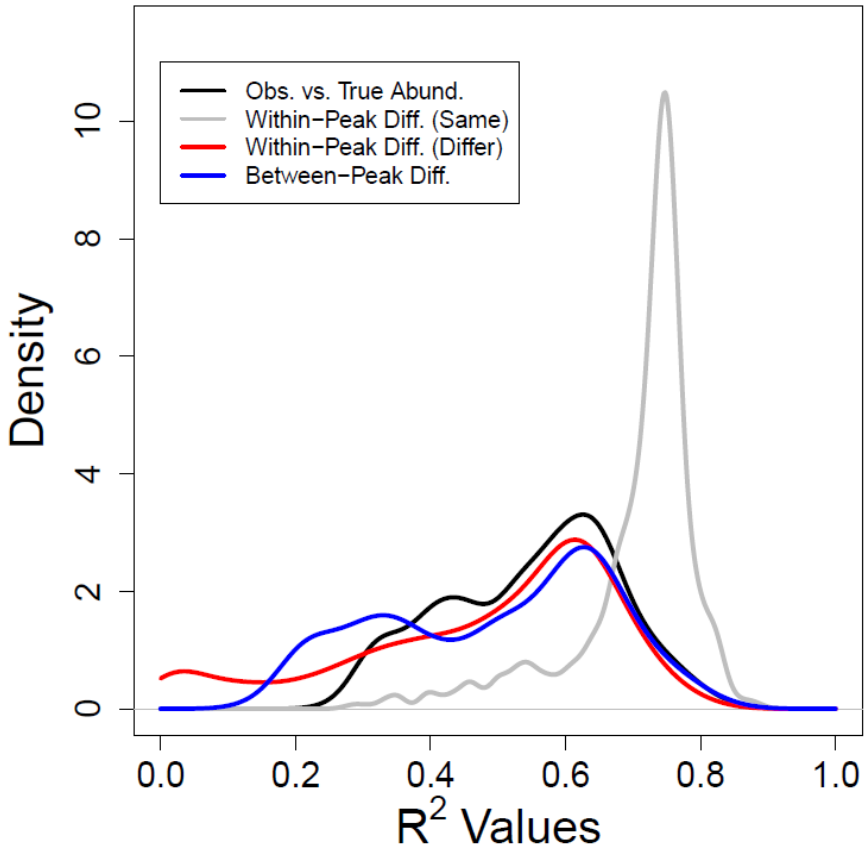


**Figure S2. Variation in observed intensity explained by true abundance.** Kernel density functions are shown for different relationships and types of error. Density functions were fit using $R^2$ values collated from across simulation iterations. Higher $R^2$ values indicate a stronger link (i.e., lower uncertainty) between observed intensities and true abundances. Black is for the relationship shown in Figure S4. Blue is for between-peak within-sample differences (example relationships shown in Figures S6A,B). Gray is for within-peak between-sample differences when the same peak-level error was used for both synthetic samples within a given simulation iteration (example relationship shown in Figure S6C). Red is for within-peak between-sample differences when different peak-level error was used across the synthetic samples within a given simulation iteration (example relationship shown in Figure S6D). While there are central tendencies in all four distributions, there is also significant variation in the degree to which observed intensities reflect true abundances.

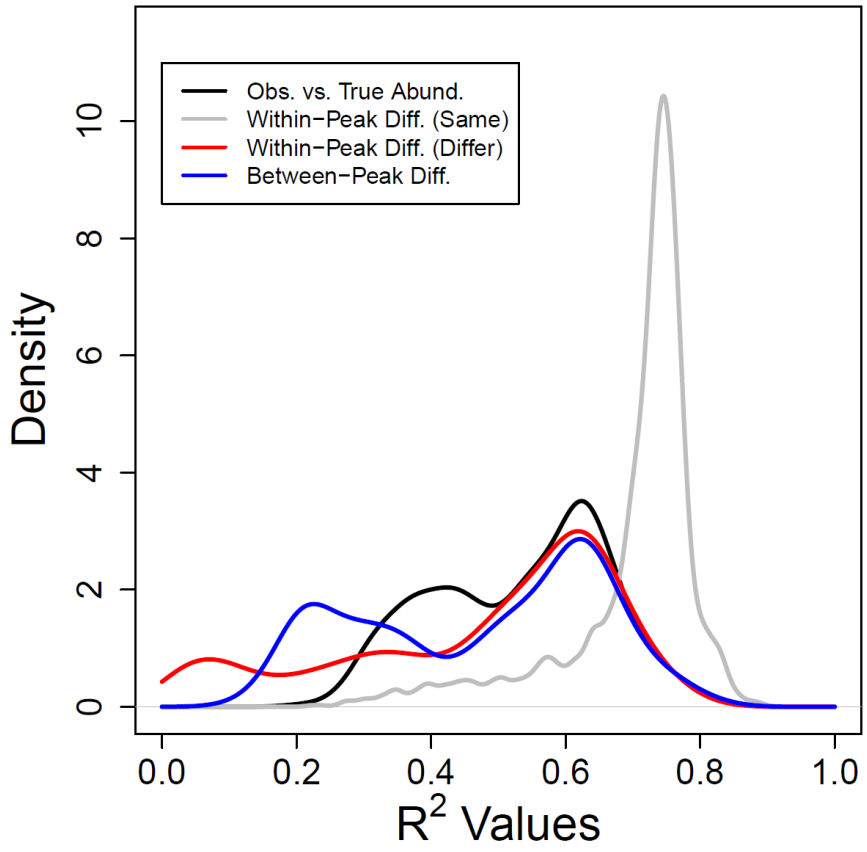


**Figure S3. Equivalent to Figure S2, but with a simulated error distribution with a range from 0 to 8. Variation in observed intensity explained by true abundance.** Kernel density functions are shown for different relationships and types of error. Density functions were fit using $R^2$ values collated from across simulation iterations. Higher $R^2$ values indicate a stronger link (i.e., lower uncertainty) between observed intensities and true abundances. Black is for the relationship shown in Figure S5. Blue is for between-peak within-sample differences (example relationships shown in Figures S7A,B). Gray is for within-peak between-sample differences when the same peak-level error was used for both synthetic samples within a given simulation iteration (example relationship shown in Figure S7C). Red is for within-peak between-sample differences when different peak-level error was used across the synthetic samples within a given simulation iteration (example relationship shown in Figure S7D). While there are central tendencies in all four distributions, there is also significant variation in the degree to which observed intensities reflect true abundances.

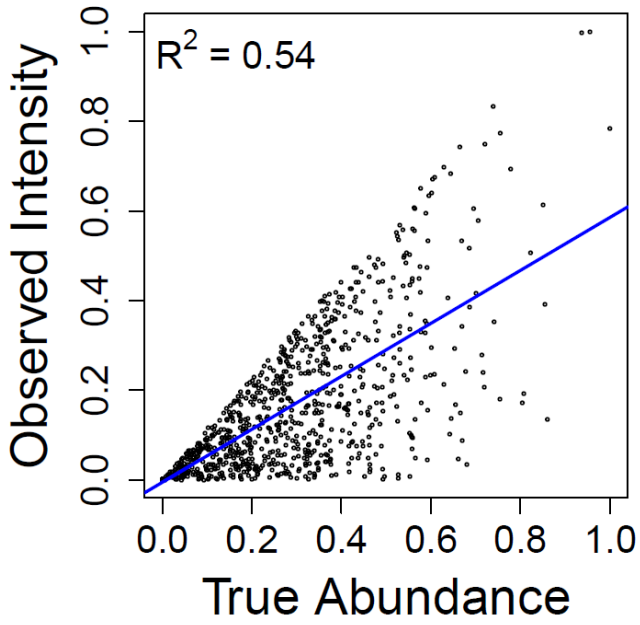


**Figure S4. Representative example of simulation model-generated estimates of observed peak intensity as a function of true abundance.** Values are derived from one synthetic sample with 1000 peaks. The blue line is the linear regression model, and the associated $R^2$ value is near the median $R^2$ value for this relationship across simulations (see black line in Figure S2). 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.

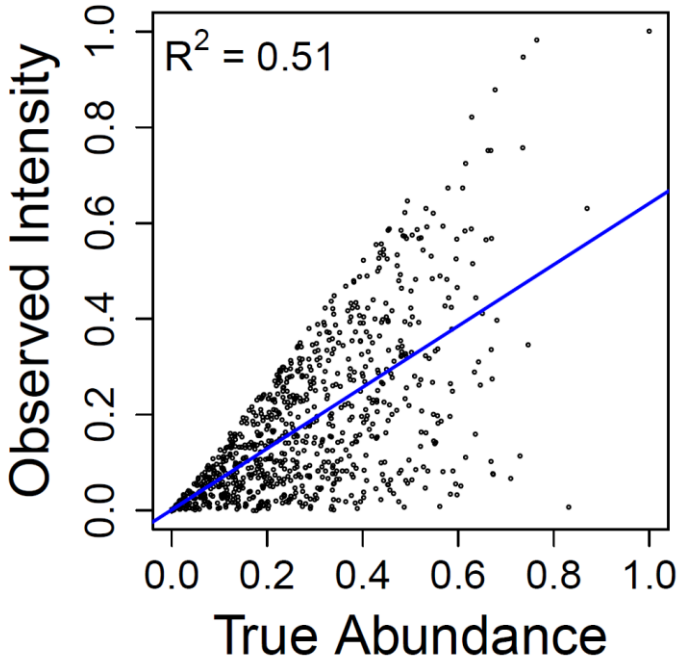

**Figure S5. Equivalent to Figure S4, but with a simulated error distribution with a range from 0 to 8. Representative example of simulation model-generated estimates of observed peak intensity as a function of true abundance.** Values are derived from one synthetic sample with 1000 peaks. The blue line 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 S3). 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.

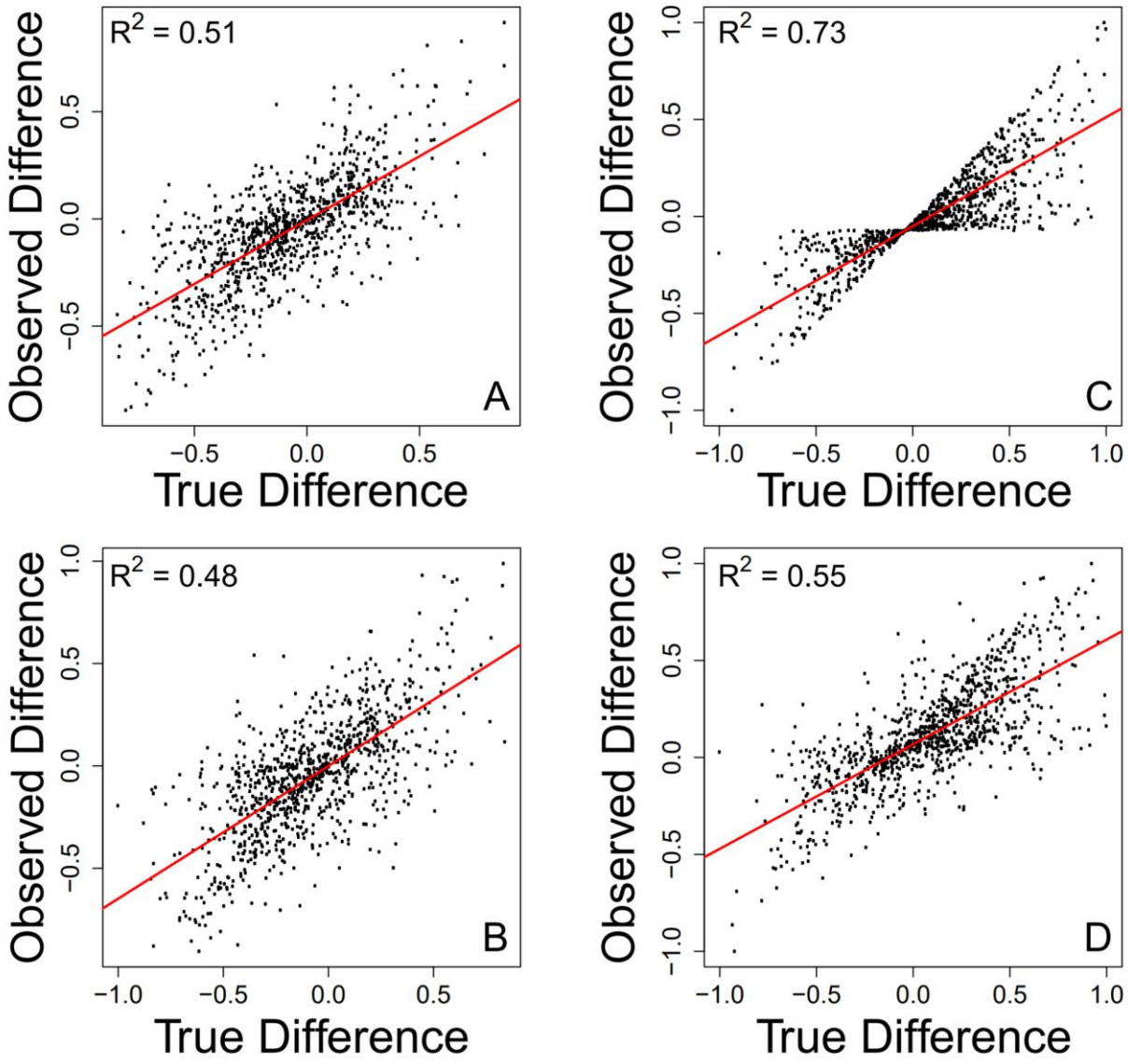


**Figure S6. Observed differences in peak intensity as a function of true differences in peak intensity across both within-peak and between-peak comparisons and across both kinds of error.** (A) Between-peak differences with the same error applied to a given peak between samples. (B) Between-peak differences with different errors applied to a given peak between samples. (C) Within-peak differences with the same error applied to a given peak between samples. (D) Within-peak differences with different errors applied to a given peak between samples. On all panels the red line represents the linear regression model, and the associated $R^2$ value is provided. On panel C the $R^2$ value should be interpreted with caution as the residuals are clearly heteroscedastic.

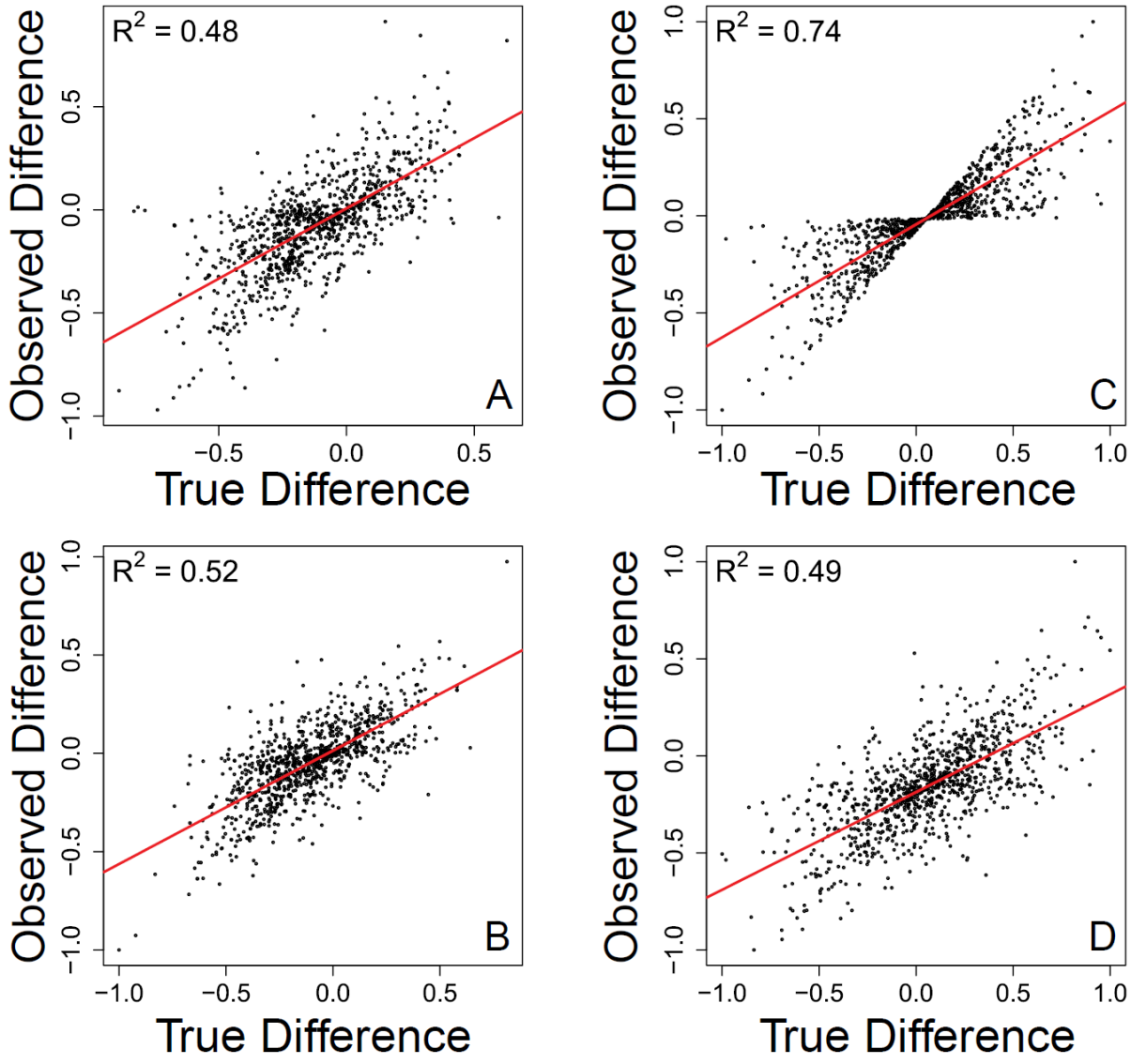


**Figure S7. Equivalent to Figure S6, but with a simulated error distribution with a range from 0 to 8. Observed differences in peak intensity as a function of true differences in peak intensity across both within-peak and between-peak comparisons and across both kinds of error.** (A) Between-peak differences with the same error applied to a given peak between samples. (B) Between-peak differences with different errors applied to a given peak between samples. (C) Within-peak differences with the same error applied to a given peak between samples. (D) Within-peak differences with different errors applied to a given peak between samples. On all panels the red line represents the linear regression model, and the associated $R^2$ value is provided. On panel C the $R^2$ value should be interpreted with caution as the residuals are clearly heteroscedastic.

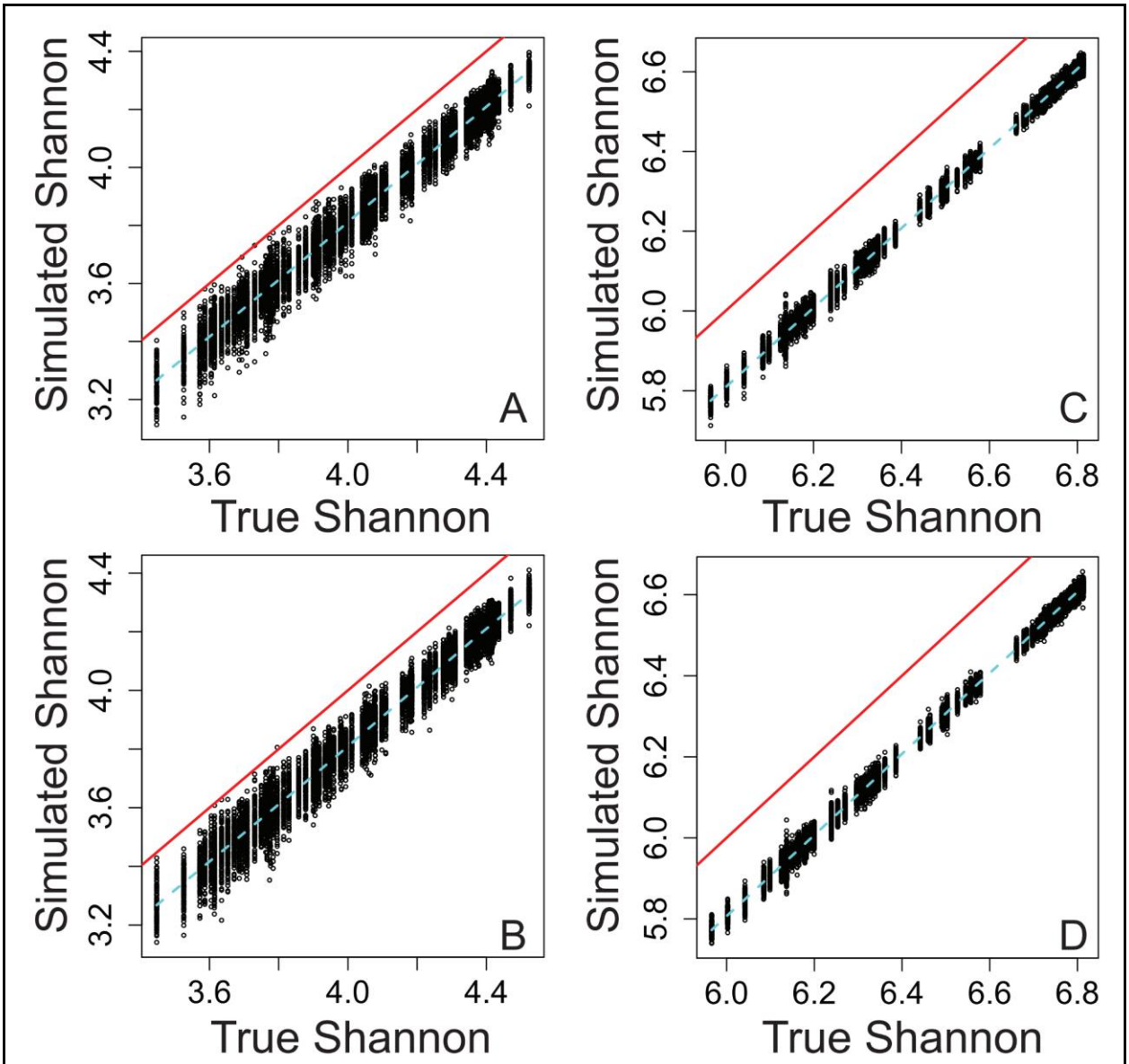


**Figure S8. Equivalent to Figure 7, but with a simulated error distribution with a range from 0 to 8. Shannon α-diversity that includes simulated error regressed against true Shannon, across different scenarios.** (A) The same error applied to a given peak between samples, and 100 peaks per sample. (B) Different errors applied to a given peak between samples, and 100 peaks per sample. (C) The same error applied to a given peak between samples, and 1000 peaks per sample. (D) Different errors applied to a given peak between samples, and 1000 peaks per sample. On all panels the red line represents the one-to-one line and the dashed line is a spline fit to the data. All data are from the simulation model.

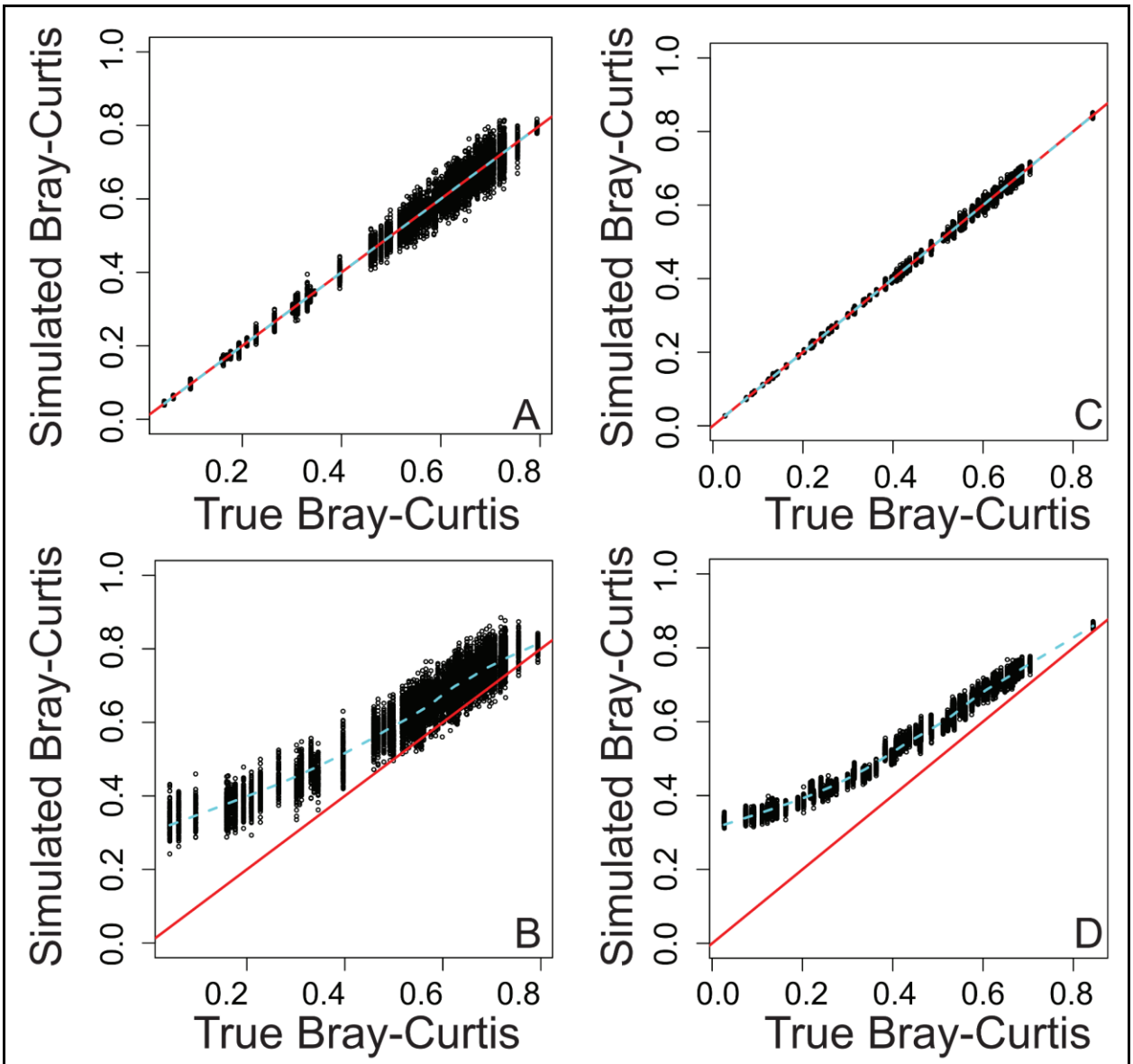


**Figure S9. Equivalent to Figure 8, but with a simulated error distribution with a range from 0 to 8. Bray-Curtis dissimilarity as a measure of β-diversity that includes simulated error regressed against true Bray-Curtis, across different scenarios.** (A) The same error applied to a given peak between samples, and 100 peaks per sample. (B) Different errors applied to a given peak between samples, and 100 peaks per sample. (C) The same error applied to a given peak between samples, and 1000 peaks per sample. (D) Different errors applied to a given peak between samples, and 1000 peaks per sample. On all panels the red line represents the one-to-one line and the dashed line is a spline fit to the data. All data are from the simulation model.

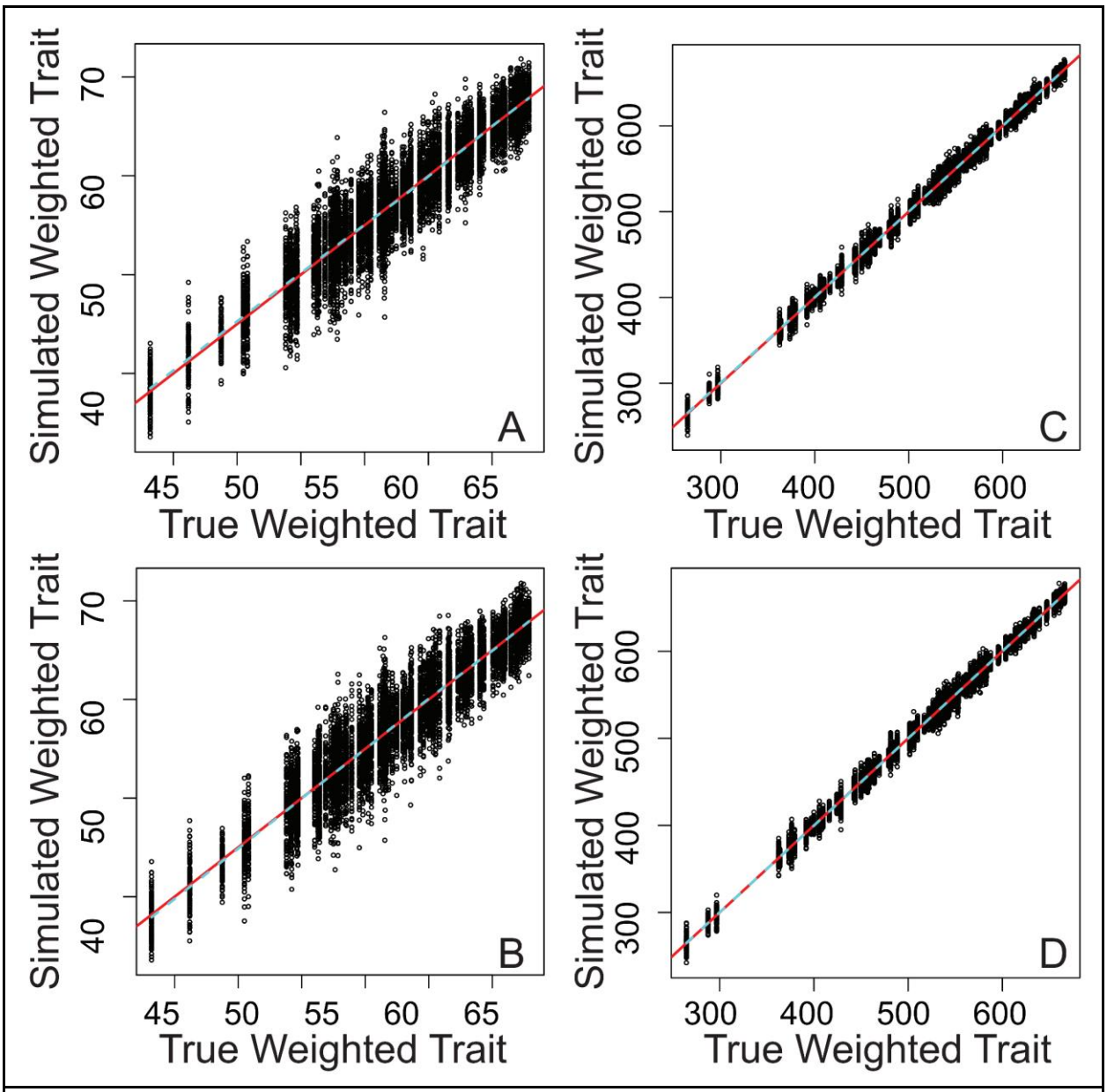


**Figure S10. Equivalent to Figure 9, but with a simulated error distribution with a range from 0 to 8. Mean peak-intensity-weighted trait values that include simulated error regressed against true mean peak-intensity-weighted trait values, across different scenarios.** (A) The same error applied to a given peak between samples, and 100 peaks per sample. (B) Different errors applied to a given peak between samples, and 100 peaks per sample. (C) The same error applied to a given peak between samples, and 1000 peaks per sample. (D) Different errors applied to a given peak between samples, and 1000 peaks per sample. On all panels the red line represents the one-to-one line and the dashed line is a spline fit to the data. All data are from the simulation model.

**Table S1. Statistics associated with analyses shown across panels in Figure 4.**

| r (pearson correlation coefficient) | p-value | chemical | panel | matrix |
|---|---|---|---|---|
| 0.970 | 0.00000000014 | Chicoric acid | A | MeOH |
| 0.936 | 0.00000000015 | Sinapic acid | A | MeOH |
| 0.945 | 0.00000000001 | Trehalose | A | MeOH |
| 0.978 | 0.00000004219 | Chlorogenic acid | B | MeOH |
| 0.992 | 0.00000012489 | Neochlorogenic acid | B | MeOH |
| 0.995 | 0.00000002368 | Cryptochlorogenic acid | B | MeOH |
| -0.749 | 0.00053452823 | Aesculin | C | MeOH |
| -0.840 | 0.00000682398 | Chlorogenic acid | C | MeOH |
| -0.833 | 0.00000966086 | Enterodiol | C | MeOH |
| -0.834 | 0.00001673756 | Ginkgolide C | C | MeOH |
| -0.540 | 0.02061742806 | Mangiferin | C | MeOH |
| -0.771 | 0.00011116595 | Phloridzin | C | MeOH |
| -0.775 | 0.00025727111 | Aesculin | D | BondElut |
| -0.922 | 0.00000005397 | Chlorogenic acid | D | BondElut |
| -0.739 | 0.00046310152 | Enterodiol | D | BondElut |
| 0.770 | 0.00018366257 | Ginkgolide C | D | BondElut |
| -0.655 | 0.00428291933 | Mangiferin | D | BondElut |
| 0.748 | 0.00035792179 | Phloridzin | D | BondElut |
| -0.868 | 0.00000300891 | Aesculin | E | BondElut-ARW |
| -0.679 | 0.00195476557 | Chlorogenic acid | E | BondElut-ARW |
| -0.940 | 0.00000000220 | Enterodiol | E | BondElut-ARW |
| 0.850 | 0.00001556556 | Ginkgolide C | E | BondElut-ARW |
| -0.404 | 0.10788543404 | Mangiferin | E | BondElut-ARW |
| 0.703 | 0.01074785364 | Phloridzin | E | BondElut-ARW |

## 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 - 2022) (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 peak intensity. Scatterplots including linear regressions were produced with Seaborn's 'regplot' function, using default bootstrapping and regression parameters. X-axis jitter was added to panels A and B to aid visualization of overlapping points. Pearson R correlation coefficients were calculated using the 'scipy' stats module and reported in Supplementary Information file "PaperFigure_AllEmpiricial_CorrelationCoefficients.xlsx".

### 2.6 Simulation Model

To generate synthetic data, we randomly assigned abundances to either 100 or 1000 peaks to study how the influences of errors change with the number of peaks; going above 1000 peaks did not change the outcomes (Fig. S1) and going below 100 peaks is unlikely to be relevant to many FTMS studies. Abundances were sampled with replacement from a Gaussian distribution that varied in mean and standard deviation across synthetic samples and across simulation iterations. Abundances were drawn twice to generate two independent samples per simulation, and the simulation was run 100 times for each number-of-peaks (100 or 1000 peaks per sample; referred to below as 'peak richness'). We varied the Gaussian distributions to generate synthetic samples varying in composition within and across simulations to ensure that the ecological metrics (see below) would vary across simulations. This step was necessary to evaluate metric performance across a broad range of metric values.

We simulated two types of error which can both be representative of variation in ionization efficiency. The goal was to generate synthetic data that mimicked our empirical and theoretical observations that indicate noise in the relationships between observed peak intensities and true abundances. For each type of error and within each iteration of the simulation, the error was introduced 100 times (i.e., 100 error iterations were nested within each sample-generation iteration). The first type of error (referred to as within-sample error) was designed to diminish the between-peak relationship between observed peak intensity and true abundance. To introduce this error, we multiplied the true abundance of each peak by a random number drawn from a uniform distribution ranging from 0 to 100. The inclusion of 0 indicates situations in which a given peak (i.e., ion) does not ionize well enough to be observed. The results should not be sensitive to the selected range, but as a sensitivity analysis, we also used a distribution of errors ranging from 0 to 8. Our empirical data suggest that this narrower range is appropriate (Fig. 4B), but simulation results were not affected by the selected error range; we present two versions of all figures based

on the simulation model, one version uses an error range of 0-100 and the other version uses an error range of 0-8. For each peak we multiplied the same random error by its abundance in each of the two synthetic samples within each iteration. This error-modified abundance of each peak in each synthetic sample was considered to be the observed peak intensity. We recognize that randomized errors do not perfectly reflect real-world variation in ionization efficiency. However, because the true impacts of matrix effects and individual molecular chemistries in complex mixtures are currently not known, the errors introduced in the model are simply used to diminish the relationship between observed peak intensities and true abundances.

Introducing error resulted in a relatively weak relationship between observed peak intensity and true abundance (median $R^2$ = ~0.5; see black lines in Figures S2, S3), with the amount of error increasing with true abundance (Figs. S4, S5). This relationship additionally supports our inclusion of error into the model as a means to simulate relatively weak relationships between observed peak intensity and true abundance. Between-peak differences in observed intensity were also weakly related to between-peak differences in true abundance (Figs. S6A, S7A), with a median $R^2$ of ~0.5 (see blue lines in Figures S2, S3). Because the same peak-level error-factor was used across both synthetic samples within a given simulation iteration, the within-peak between-sample differences in observed intensity were relatively strongly correlated to within-peak between-sample differences in true abundance (Figs. S6C, S7C), with a median $R^2$ of ~0.75 (see the gray line in Figures S2, S3). However, we suggest caution when interpreting the $R^2$ values associated with Figure S6C and S7C as the differences collapse when near zero, leading to heteroscedastic residuals that likely bias the $R^2$. This phenomenon can be explained by the fact that when two samples have essentially the same peak intensity for a given peak, introducing the same error to that peak in both samples has little influence on the between-sample difference in peak intensity.

The second type of error we introduced represents situations in which ionization efficiency varies across molecules – as in the first type of error – as well as across samples (referred to as between-sample error). Molecules may exhibit variations in ionization efficiency across samples due to changes in the composition of organic molecules and/or changes in inorganic solutes in the matrix (see above). To account for these effects, we multiplied the true abundance of each peak by a random number drawn from a uniform distribution ranging from 0 to 100; for sensitivity analysis, we also used an error distribution ranging from 0 to 8, which did not have meaningful influences on the results. For each iteration of the simulation, we introduced errors independently for the two synthetic samples. In this way, the simulated ionization efficiency for a given peak in a given synthetic sample was independent of its ionization efficiency in the other synthetic sample. The error-modified abundance of each peak in each synthetic sample was considered to be the observed peak intensity.

We observed a relatively large influence on observed peak intensities when allowing ionization efficiency to vary across samples. That is, the within-peak between-sample differences in observed intensity were weakly correlated to within-peak between-sample differences in true abundance (Figs. S6D, S7D), with a median $R^2$ of ~0.5 (see the red lines in Figures S2, S3). Compared to the same relationship that emerged under the first type of error, our results show a much weaker relationship between peak intensity and true abundance when ionization efficiency varies between samples (compare the gray and red lines in Figures S2 and S3). This result is expected, as variations in ionization efficiency add random noise to the within-peak between-sample differences in observed peak intensity. We note that the variation in ionization efficiency is independent between peaks for both the first and second types of error. The between-peak relationship summarized in Figures S2 and S3 (blue line) is, therefore, equivalent for both types of error, which is further supported by the strong similarity between Figures S6A and S6B (also true for Figures S7A and S7B).

To examine how both types of error influence ecological metrics, we used the initial true abundances and the error-modified abundances (i.e., observed peak intensity values) to calculate true and 'observed' values of within-sample Shannon diversity and between-sample Bray-Curtis. We also assigned an arbitrary trait value to each peak and calculated true and observed sample-level mean trait values; the mean values for each sample were weighted by true

abundance (true mean) or observed peak intensity (observed mean). We regressed observed values for Shannon diversity, Bray-Curtis, and mean traits against their true values, and performed this process independently for each level of peak richness. The resulting patterns for the case with error range from 0-100 are presented and discussed in the main text (Figs. 7-9). The patterns resulting from the case with error range from 0-8 showed the same patterns (Figs. S8-S10). This indicates no meaningful influence of the chosen error range within the simulation model.

## 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.