Experimental assessment of EAIRMS normalization methodologies for environmental stable isotopes

Morgan Schwartz[1], Sawyer Balint[1], Drew Fowler[2], Stella Linnekogel[3], Sáde Cromratie Clemons[4], and Laura K. Burkemper[5]

[1]US Environmental Protection Agency Office of Research and Development Atlantic Coastal Environmental Sciences Division
[2]US Geological Survey Lower Mississippi-Gulf Water Science Center Baton Rouge LA Office
[3]Centre for Ecology & Hydrology Lancaster Site
[4]University of Colorado Boulder Department of Geography
[5]The University of New Mexico

November 19, 2023

Abstract

RATIONALE: In stable isotope mass spectrometry, isotope values are normalized to internationally recognized reference scales using certified reference materials and working standards. Numerous techniques exist for performing this normalization, but these methodologies need to be experimentally assessed to compare their impact on reproducibility of isotope results. **METHODS**: We tested normalization methods by the number of standards used, their matrix, their isotope range, and whether normalization required extrapolating beyond the isotope range. Using 8 certified reference materials and 5 working standards on a ThermoFinnigan Delta-V IRMS and Elementar VisION IRMS for nitrogen and carbon isotope composition via solid combustion with an elemental analyzer, we computed every possible isotope normalization (n=6272). Additionally, we assessed how sample matrix impacted linearity effects on both instruments. **RESULTS**: Normalizations composed of three or four reference materials had better performance than one-point and two-point methods, especially when the normalization was matrix-mixed or extrapolated, and normalizations with an isotope range greater than 15were more accurate under these conditions. Normalizations that were matrix-matched and were not extrapolated exhibited the highest accuracy. Linearity effects were found to exceed instrument precision by two orders of magnitude irrespective of sample matrix and were not predicted by reference gas diagnostics. **CONCLUSIONS**: To maximize interlaboratory comparability of isotope results, operators of EAIRMS systems should use at least 3 calibration standards to construct their normalizations, select standards with a large isotope range to avoid extrapolation, and match the matrix of their standards to their samples to the best extent possible.

Experimental assessment of EAIRMS normalization methodologies for environmental stable isotopes

Sawyer Balint[1]*, Morgan Schwartz[2]*, Drew Fowler[3+], Stella Linnekogel[4+], Sáde Cromratie Clemons[5], Laura K. Burkemper[6]

[1]ORISE Participant, U.S. EPA Atlantic Coastal Environmental Sciences Division, Narragansett, RI. 02882

[2]U.S. EPA Atlantic Coastal Environmental Sciences Division, Narragansett, RI. 02882

[3]U.S. Geological Survey, Louisiana Cooperative Fish and Wildlife Research Unit, Baton Rouge, LA. 70810

[4]UK Centre for Ecology & Hydrology Lancaster, Library Avenue, Bailrigg, Lancaster, LA1 4AP, United Kingdom

[5]Department of Geography, University of Boulder, Boulder, CO

[6]Center for Stable Isotopes, University of New Mexico, Albuquerque, NM

* Authors share first authorship

[+]Authors contributed equally to this manuscript

Corresponding Author :

Sawyer Balint

sjbalint@bu.edu

(518) 258-9510

Morgan Schwartz schwartz.morgan@epa.gov

(401) 782-9624

CRediT author statement :

Sawyer Balint : Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing – Original Draft, Visualization, Supervision, Project administration.

Morgan Schwartz : Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Supervision, Project administration.

Drew Fowler : Conceptualization, Methodology, Investigation, Writing – Review & Editing.

Stella Linnekogel : Conceptualization, Methodology, Investigation, Writing – Review & Editing.

Sade Cromratie Clemons: Conceptualization, Methodology, Investigation, Writing – Review & Editing.

Laura K. Burkemper: Conceptualization, Methodology, Validation, Investigation, Resources, Data Curation, Writing – Review & Editing, Supervision, Project administration.

Acknowledgements:

The authors would like to thank Viorel Atudorei, Seth Newsome, and Zachary Sharp for their assistance in the planning and execution of IsoCamp 2022, where the experimental design for this project was conceived. Additionally, the authors thank Renee Brooks, Michaela Cashman, and Autumn Oczkowski for review and feedback on the manuscript. Stella Linnekogel (UKCEH) acknowledges support from the NERC project Diurnal Variation in Soil Nitrous oxide Emissions (DIVINE): drivers and mechanisms (NE/V000837/1). This article has been subjected to Agency review and has been approved for publication. This work was supported in part by an appointment to the Research Participation Program at the Office for Research and Development, Center for Environmental Measurement and Modelling, Atlantic Coastal Environmental Sciences Division, U.S. Environmental Protection Agency (U.S. EPA), administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and U.S. EPA. The views expressed in this presentation are those of the authors and do not necessarily represent the views or polices of the U.S. EPA. This article has been peer reviewed and approved for publication consistent with USGS Fundamental Science Practices (https://pubs.usgs.gov/circ/1367/). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Abstract

RATIONALE : In stable isotope mass spectrometry, isotope values are normalized to internationally recognized reference scales using certified reference materials and working standards. Numerous techniques exist for performing this normalization, but these methodologies need to be experimentally assessed to compare their impact on reproducibility of isotope results.

METHODS : We tested normalization methods by the number of standards used, their matrix, their isotope range, and whether normalization required extrapolating beyond the isotope range. Using 8 certified reference materials and 5 working standards on a ThermoFinnigan Delta-V IRMS and Elementar VisION IRMS for nitrogen and carbon isotope composition via solid combustion with an elemental analyzer, we computed every possible isotope normalization (n=6272). Additionally, we assessed how sample matrix impacted linearity effects on both instruments.

RESULTS : Normalizations composed of three or four reference materials had better performance than one-point and two-point methods, especially when the normalization was matrix-mixed or extrapolated, and normalizations with an isotope range greater than 15under these conditions. Normalizations that were matrix-matched and were not extrapolated exhibited the highest accuracy. Linearity effects were found to exceed instrument precision by two orders of magnitude irrespective of sample matrix and were not predicted by reference gas diagnostics.

CONCLUSIONS : To maximize interlaboratory comparability of isotope results, operators of EAIRMS systems should use at least 3 calibration standards to construct their normalizations, select standards with a large isotope range to avoid extrapolation, and match the matrix of their standards to their samples to the best extent possible.

Stable isotope mass spectrometry is a growing tool across disciplines, including in biology where the nitrogen (N) and carbon (C) isotope composition of solid samples are routinely used for investigating a variety of physiological, ecological, and biogeochemical questions[1]. In recent decades, the broad applications of stable isotopes combined with faster and more accessible instrumentation have led to a large growth in the utilization and publication of stable isotope measurements as they relate to the natural sciences[2]. Despite the growing importance of this tool, immense variations in analytical methodology exist within the field, leading to poor interlaboratory comparability[3]and complicating the interpretation and reproducibility of scientific studies that use stable isotopes[4].

Continuous flow stable isotope mass spectrometry[5] via solid combustion[6,7] is the predominant analytical technique used to quantify the C and N isotope composition of solid samples for biological applications[2]. Samples are delivered to the isotope ratio mass spectrometer (IRMS) in a gaseous form through a peripheral elemental analyzer (EA) and the isotope composition is determined relative to a working gas that is injected sequentially with the sample gas[8,9]. This methodology allows for high precision measurements, but to facilitate interlaboratory comparison the results must be normalized to internationally referenced isotope scales[10-12]. Contemporary normalization methods call for standard reference materials to be processed identically to the unknown samples using the "identical treatment" principle[5,13], and analyzed in tandem with unknown samples[14-16], a normalization curve is then computed with a least-squares linear regression[17]. As the number of available certified standards has increased in recent decades[18-20], operators now have a multitude of certified standards to choose from when performing their isotope normalizations in addition to their own working standards.

Unfortunately, the reproducibility of normalization techniques has not advanced at the same pace as the expanding use of stable isotopes. Although modeling work has suggested that normalization error generally decreases with the number of standards[21], stable isotope laboratories vary immensely in the number of standards used due in part to limited experimental assessments of related normalization error. Furthermore,

users of elemental analyzer isotope ratio mass spectrometry (EAIRMS) for biological applications often analyze sample matrixes for which certified reference materials do not exist (e.g., sediment[22], marine algae[23], samples imbedded on glass fiber filters[24]), but the impact of mixing organic matrixes between samples and standards on normalization accuracy is unknown. Whether standards should be selected to maximize isotope range or closely bracket the unknown samples is another decision that, absent of an experimental assessment across a variety of normalization methods, is left to anecdotal procedures that may vary between laboratories. Some analyses, such as those that incorporate N tracer as part of the methodology[25], may require extrapolation beyond the range of the normalization curve – again with unknown consequences. Ultimately, readers of studies that incorporate stable isotopes are left to determine the reproducibility of the study for themselves – assuming the particulars of the normalization are even included in the methods. This matter becomes increasingly more difficult in the era of big data and meta-analysis, as multiple studies using multiple methods are integrated and consequently compared directly.

Here, we aim to better quantify best normalizations practices for EAIRMS analysis using an experimental assessment of a variety of normalization methods using certified reference materials analyzed on two instrument systems in two laboratories. We specifically assess how the accuracy of the normalization is impacted by the number of standards, their matrix relative to the unknown samples, their isotope range, and whether the normalization requires extrapolation beyond the isotope range. We further investigate how instrument linearity effects are impacted by sample matrix through the first known interlaboratory comparison of EAIRMS linearity effects. By developing a refined understanding of EAIRMS best practices, we hope to better facilitate the reproducibility of biological stable isotope applications.

Number of standards used to construct the normalization

Although a variety of methodologies have been used to normalize results from isotope ratio mass spectroscopy[8,15], this work focuses on the most common methods used in continuous flow applications: one-point anchoring, two-point linear normalization, or multipoint linear normalization [2,17]. Typically, the IRMS software will perform a one-point anchoring relative to the working gas according to the following equations:

$$\delta_{\text{raw(sample)}} = \frac{R_{\text{measured(sample)}}}{R_{\text{measured(wg)}}} - 1 \qquad \text{Eq. 1}$$

$$\delta_{\text{true(sample)}} = \delta_{\text{raw(sample)}} + \delta_{\text{true(wg)}} + \left(\delta_{\text{raw(sample)}} \times \delta_{\text{true(wg)}}\right) \qquad \text{Eq. 2}$$

where $\delta_{\text{raw(sample)}}$ is the measured isotope composition ($R_{measured}$) of the sample relative to the working gas (wg) and$\delta_{\text{true(wg)}}$ is the actual isotope composition of the working gas specified by the operator. Per convention, the relative difference of isotope ratios (δ) for N and C are expressed in parts per thousand (Although it is possible to quantify$\delta_{\text{true(wg)}}$ and thus calculate the true isotope composition of the sample($\delta_{\text{true(sample)}}$) via a one-point anchoring using the working gas alone, identical treatment principle dictates replacing Eq. 2 with replicate analyses of combusted reference material according to the following equation:

$$\delta_{\text{true(sample)}} = \delta_{\text{raw(sample)}} + \left(\frac{\sum \delta_{\text{true(std)}} - \sum \delta_{\text{raw(std)}}}{n}\right) \qquad \text{Eq. 3}$$

where $\delta_{\text{raw(std)}}$ is the isotope composition of the standard relative to the reference gas (Eq. 1) and$\delta_{\text{true(std)}}$ is the certified isotope composition of the standard. When two or more standards are used for normalization as recommended by the International Atomic Energy Agency (IAEA)[11,27],$\delta_{\text{true(sample)}}$ can be calculated using a least squares regression between$\delta_{\text{true(std)}}$ and$\delta_{\text{raw(std)}}$ according to the following equations:

$$m = \left(\frac{n \sum (\delta_{\text{raw(std)}} \times \delta_{\text{true(std)}}) - \sum \delta_{\text{raw(std)}} \sum \delta_{\text{true(std)}}}{n \sum \left(\delta_{\text{raw(std)}}\right)^2 - \left(\sum \delta_{\text{raw(std)}}\right)^2} \right) \quad \text{Eq. 4}$$

$$b = \left(\frac{\sum \delta_{\text{true(std)}} - m \sum \delta_{\text{raw(std)}}}{n} \right) \quad \text{Eq. 5}$$

$$\delta_{\text{true(sample)}} = \delta_{\text{raw(sample)}} \times m + b \quad \text{Eq. 6}$$

where m is the expansion coefficient, b is the intercept, and n is the number of measurements of standards. Past work investigating C isotopes has shown that normalizations using two or more standards substantially reduce normalization errors compared to one-point anchoring[11,17] because the expansion coefficient is rarely equal to one when multipoint normalizations are used[21], but Eq. 3 assumes an expansion coefficient of one. The addition of supplementary standards to produce a three- or four-point normalization allows a coefficient of determination (r^2) to be calculated and theoretically reduces the extrapolation of errors stemming from isotope heterogeneity of the standards, although experimental assessments[17] and Monte Carlo simulations[21] of this hypothesis have so far been limited to C. Furthermore, past assessments of how many standards should be used in a normalization have not concurrently considered the effects of scenarios such as matrix mixing and extrapolation, which may require additional standards to ensure acceptable results.

Finally, we note that single or multipoint anchoring can also be accomplished by analyzing one or multiple isotope standards to calculate an intercept (Eq. 5) and applying an expansion coefficient (m) calculated from a previous analysis of multiple standards to conduct a pseudo-linear normalization (Eq. 6) assuming that the expansion coefficient is stable over time[17]. This normalization method is susceptible to changes in instrument performance over days or weeks and with different instrument tunings and is thus not investigated in this work. Here, we perform one-point anchoring and two-point, three-point, and four-point linear normalizations for N and C on two EAIRMS systems in two laboratories to experimentally assess their accuracy and interlaboratory comparability.

1.2 Standard matrix, normalization isotope range, and instrument linearity

The impact of sample matrix and isotope range of the standards on normalization accuracy have not been as thoroughly investigated as the number of standards. Past work has identified matrix effects between organic and inorganic standards for N[28], but laboratories that process biological samples frequently analyze a variety of organic sample matrixes (e.g., muscle tissue[29], plants[30], soils[22]) that have different preparation techniques and combustion properties. Although proper application of the "identical treatment" protocol would suggest matching the matrix of the sample and of the standard[31], some sample matrixes, such as sediment and particulate organic matter collected on glass fiber filters, do not have corresponding certified reference materials available for purchase, which thwarts matrix matching. The effect of matrix on the isotope results of organic samples complicates the interpretation of biological stable isotopes[32,33]and requires additional experimental assessment.

The isotope range of the normalization curve, and how it may affect the accuracy of results, presents another area of uncertainty. Laboratories may use standards that have very different isotope compositions, thus generating a normalization curve with a large isotope range (i.e., >40samples by using normalization curves with a comparatively small isotope range. Furthermore, when a sample has an isotope composition that exceeds the range of normalization standards, extrapolation beyond the normalization curve may be required. Past work on C isotopes has suggested that tightly bracketing unknown samples may marginally improve accuracy if the normalization is not extrapolated[21], but this investigation was constrained to two-point C normalizations. In this study, we explore how the matrix and isotope range of standards impacts the accuracy of one-point, two-point, three-point, and four-point normalizations for N and C.

The impact of instrument linearity and subsequent linearity corrections is a final aspect of stable isotope analyses that may hinder the reproducibility of biological studies. Linearity, the mass-dependent change in reported isotope composition from an IRMS, is anecdotally well known and yet has not been subjected to an

experimental interlaboratory comparison[34]. This characteristic of IRMS systems is of particular importance for biological applications because there can be large variations in the organic matter content of the samples, particularly for sediments and particulate organic matter. Because the amount of N_2 and CO_2 that enters the IRMS is proportional to the amount of organic matter in biological samples, instrument linearity can become the most important determinant of analysis precision for samples with high variations in organic matter content. Operators of EAIRMS systems frequently develop their own linearity corrections, if they correct for the phenomenon at all, but the magnitude and variability of instrument linearity between facilities is unknown. Whether linearity is dependent on the matrix being analyzed or can be predicted from reference gas diagnostics are additional areas of uncertainty. We seek to better understand how linearity affects EAIRMS isotope results for biological applications by performing reference gas diagnostics and replicate analyses of working standards of different organic matter compositions and sample weight at two laboratories.

Methods

1.

Stable isotope analysis

Eight certified reference materials and five in-house working standards were analyzed at the University of New Mexico Center for Stable Isotopes (UNM-CSI) and the United States Environmental Protection Agency, Atlantic Coastal Environmental Sciences Division (U.S. EPA). These standards encompassed a $\delta\ ^{15}$N range of -4.52to +37.83-1.17included in Table 1. For the normalization analysis, the eight certified reference materials were weighed to between 0.15 and 3.3 mg depending on the N content of the sample matrix to provide ˜0.05 mg of N. Dilutions were then computed for each matrix to provide ˜0.04 mg C to the IRMS. Each standard was run 4 times at each facility and were run in an alternating order to mitigate the effects of instrument drift. To quantify instrument linearity, the five working standards were weighed to between 0.09 and 40 mg and analyzed using a constant matrix-specific C dilution to produce a range of 0.013-0.16 mg of N and 0.009-0.10 mg of C across all working standards. At both facilities, the elemental analyzer was configured to inject a longer O_2 dosing for standards with plant matrixes (USGS90, USGS91, CSI Blue Grama, and CSI Green Chile) and soil matrixes (CSI Soil). Analyses at UNM-CSI were performed on a Costech 4010 combustion EA paired with a ThermoFinnigan Delta-V continuous-flow IRMS (by a Conflo IV), while analyses at the U.S. EPA were performed on an Elementar Vario Isotope Select EA paired with an Elementar VisION continuous-flow IRMS. The pooled standard deviation ($\pm1\sigma$) of the certified reference materials were $\pm$ 0.115 for$\delta\ ^{15}$N and $\pm$0.044 for$\delta\ ^{13}$C at the U.S. EPA, and $\pm$ 0.084 for$\delta\ ^{15}$N and $\pm$0.069 for$\delta\ ^{13}$C at UNM-CSI. Pooled standard deviations of each certified reference standard are shown in Table 1.

Table 1 : Certified isotope composition and associated uncertainty ($\pm1\sigma$) of certified standard reference materials used in this study.

Name	Certified δ^{15}N (	Certified δ^{13}C (	Observed δ^{15}N ($\pm1\sigma$)	Observed δ^{13}C ($\pm1\sigma$)	Matrix	Matrix Classification
IAEA600[11]	+1.0 ± 0.2	–27.771 ± 0.043	±0.063	±0.046	Caffeine	High Organic
USGS40[18]	–4.52 ± 0.06	–26.39 ± 0.04	±0.041	±0.067	L-glutamic acid	High Organic
USGS61[19]	–2.87 ± 0.04	–35.05 ± 0.04	±0.027	±0.042	Caffeine	High Organic
USGS63[19]	+37.83 ± 0.06	–1.17 ± 0.04	±0.052	±0.120	Caffeine	High Organic
USGS88[20]	+14.96 ± 0.14	–16.06 ± 0.07	±0.046	±0.092	Collagen	High Organic

Name	Certified $\delta^{15}N$ (	Certified $\delta^{13}C$ (	Observed $\delta^{15}N$ ($\pm1\sigma$)	Observed $\delta^{13}C$ ($\pm1\sigma$)	Matrix	Matrix Classification
USGS89[20]	+6.25 ± 0.12	–18.13 ± 0.11	±0.142	±0.094	Collagen	High Organic
USGS90[20]	+8.84 ± 0.17	–13.75 ± 0.06	±0.052	±0.210	Flour	Plant
USGS91[20]	+1.78 ± 0.12	– 28.28 ± 0.08	±0.048	±0.111	Flour	Plant
CSI Blue Grama					Plant tissue	Plant
CSI Casein					Protein	High Organic
CSI Tuna					Muscle tissue	High Organic
CSI Chile					Plant tissue	Plant
CSI Soil					Sediment	Sediment

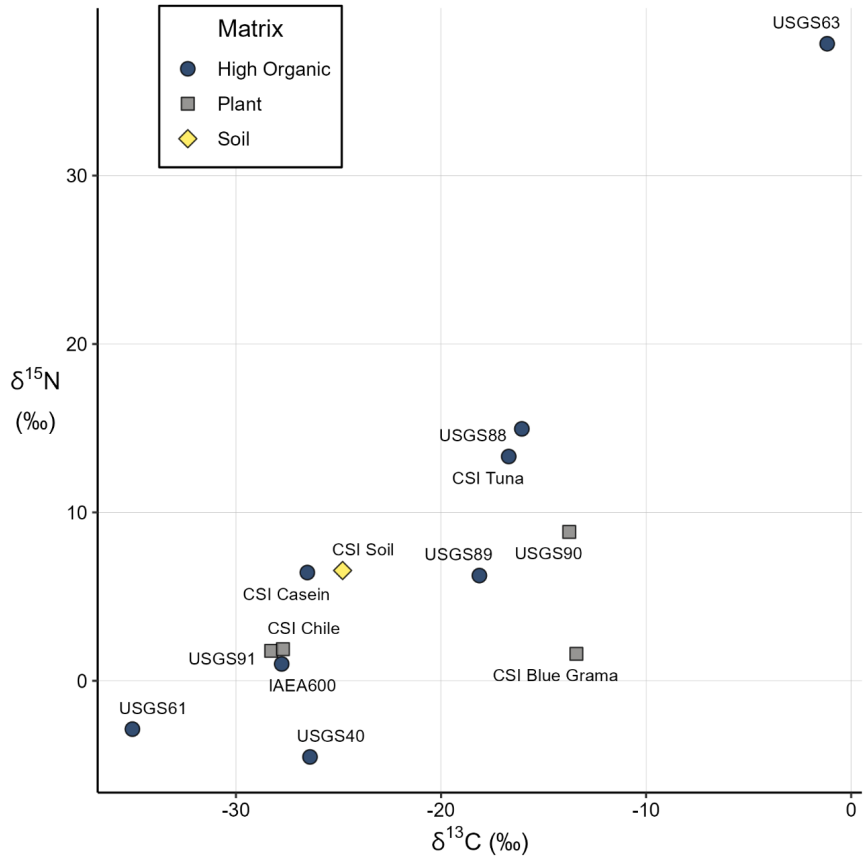

Fig. 1 : Isotope range of certified reference materials and working standards used in this study[11,18-20]

2.2 Statistical analyses

To test normalization accuracy, we only used the 8 certified reference materials as calibration standards, while the laboratory working standards were used to test linearity in the two instruments. The isotope composition of the certified and working standards were normalized to the working gas (Eq. 1) in the vendor IRMS software (Isodat and lyticOS for the Thermo Delta-V and Elementar VisION, respectively) and then exported in a tabular format; all subsequent normalizations and analyses (Eq. 3-5) were performed in R version 4.2.1[35]. For each normalization, two certified reference materials were designated as quality controls. Quality controls were excluded from the normalization calculation, and the performance of the normalization was assessed using the average observed isotope composition of the quality controls relative to their expected value[36]. For each combination of quality controls (28 unique combinations), all possible one-point, two-point, three-point, and four-point combinations of the remaining certified reference materials were determined for a total of 1568 combinations. These remaining certified reference materials were used as calibration standards. Using those combinations of calibration standards and quality controls, one-point anchoring and multipoint linear normalizations were calculated for each element (C and N) and facility (2) for a total of 6272 normalizations. Two-point normalizations composed of IAEA 600 and USGS 91 were excluded from subsequent data analysis and visualization because the small isotope range between those standards (<1precluded an accurate calculation of a realistic expansion coefficient. Although these standards could be used for a two-point anchoring using an expansion coefficient derived from a different multipoint normalization, assessing that method is beyond the scope of this study.

To assess how instrument accuracy was impacted by the selection of standards and quality controls, the normalizations were characterized according to their isotope range, the matrix of the standards relative to the quality controls, and whether the normalization was extrapolated. The isotope range of each normalization was calculated for each element as the difference between the maximum and minimum expected isotope composition of the calibration standards used in that normalization. If the expected isotope value of both quality controls fell outside the isotope range of the calibration standards, then the normalization was classified as an "extrapolation". One-point normalizations, which have an isotope range of zero, were classified as an "extrapolation" if the single calibration standard was not bracketed by the two quality controls. Finally, the matrix of each standard was classified as high organic (i.e., protein, caffeine, collagen, L-glutamic acid) or plant (i.e., plant tissue, flour). If the matrix of the calibration standards matched the matrix of the quality controls than the normalization was classified as "matrix matched", while if the matrix of the calibration standards and quality controls were different (e.g., high organic standards used to normalize plant quality controls), then the normalization was classified as "matrix mixed". If both the quality controls and the calibration standards were composed of a combination of plants and high organics, then the normalization was classified as "both mixed."

The significance of differences between different normalization methodologies, facilities, matrixes, and extrapolation status were assessed using Kruskal-Wallis testing with Dunns post-hoc testing after the assumption of normality was rejected with Shapiro-Wilks's testing[35].

Results

1.

Normalization methodology comparison

The impact of the number of calibration standards on normalization errors were considered for two subsets of the data: normalizations that were matrix-matched and bounded, and normalizations that were matrix-mixed and extrapolated. The former set of conditions were hypothesized to perform better than the latter. When normalizations were matrix-matched and bounded, no significant difference for either element was observed between one-point, two-point, three-point, and four-point normalizations, although the variance of two-point normalizations is higher than other methods (Fig. 2A). When the analysis was constrained to

normalizations that were matrix-mixed and extrapolated, the number of standards used imparted significant differences on the accuracy of the normalization (Fig. 2B). Two-point N normalizations (median error = 0.232error = 0.119(median error = 0.118four-point normalizations (median error = 0.070, n = 170, $p <$ 0.0001). Furthermore, three-point normalizations had significantly higher error than four-point normalizations ($p = 0.021$). Similarly, two-point C normalizations (median = 0.308exhibited significantly higher error than one-point (median error = 0.209$p <$ 0.0001), and four-point normalizations (median error = 0.148, $p <$ 0.0001).

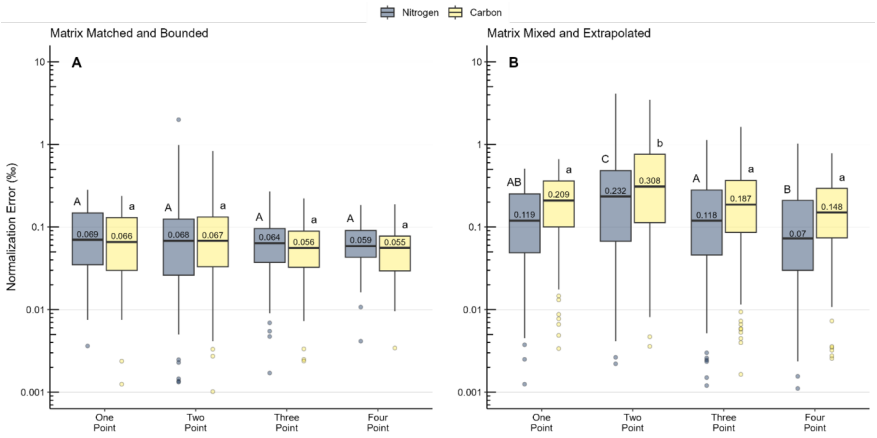

Fig 2 : Boxplots of the accuracy of one-point, two-point, and multipoint normalizations that are matrix matched and bounded (A) and matrix mixed/both mixed and extrapolated (B). Significance is shown in the compact letter display, where any two methods with the same letter are not significantly different as per Dunn's post-hoc testing. Nitrogen (uppercase) and carbon (lowercase) letters are not compared with each other.

Effect of matrix and isotope range on normalization accuracy

To isolate the effect of matrix-matching on normalization errors, we present results from only three-point normalizations. When the results are constrained to bounded normalizations, the matrix of the standards relative to the quality controls has a significant effect for both N and C (Fig. 3). Nitrogen normalizations that have calibration standards composed of a different matrix than the quality controls (median error = 0.104matrixed-matched normalizations (median error = 0.064$<$ 0.0001) and normalizations where the standards and quality controls are both composed of a combination of matrixes (median error = 0.058normalizations (median error = 0.165errors than normalizations composed of matched matrixes (median error = 0.056error = 0.121normalizations with both-mixed matrixes have significantly greater errors than matched-matrix normalizations ($p < 0.0001$).

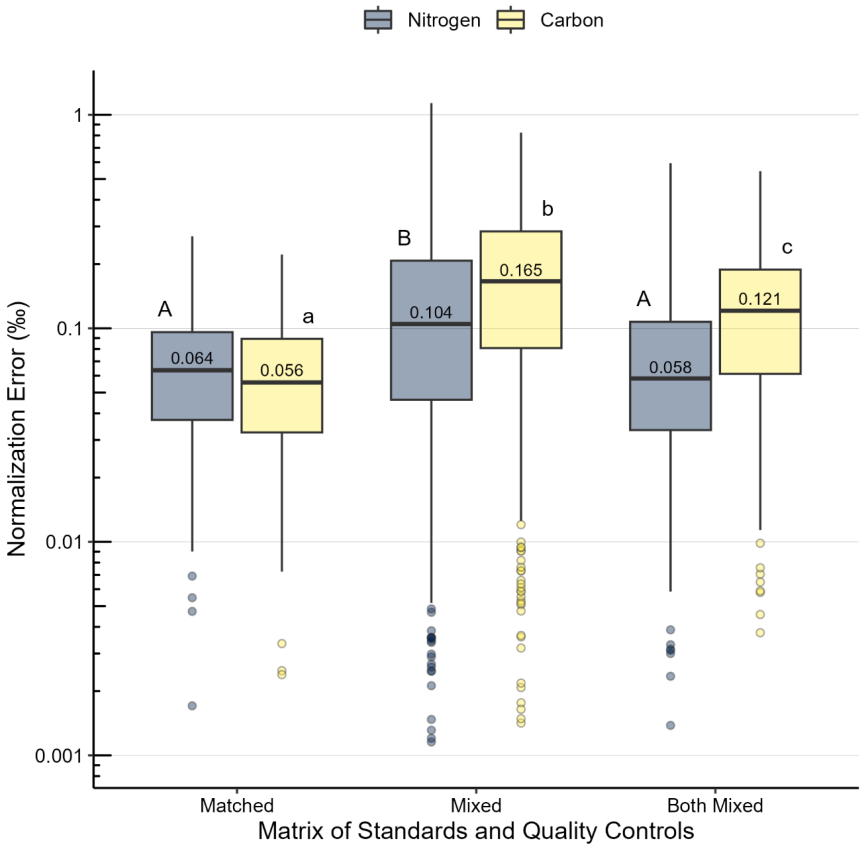

Fig. 3 : Boxplots of the accuracy of bounded three-point normalizations based on whether the matrixes of the standards and the quality controls are the same (matched) or mixed. If the three standards and the two quality controls are both composed of a combination of matrixes, the normalization is classified as "both mixed." Significance is shown in the compact letter display, where any two matrixes with the same letter are not significantly different as per Dunn's post-hoc testing. Nitrogen (uppercase) and carbon (lowercase) letters are not compared with each other.

To better understand how extrapolating outside of the normalization curve contributes to normalization errors, we assessed the impact of extrapolation on three-point normalizations that were matrix-matched (Fig. 4). Three-point normalizations that had at least one quality control within the isotope range of the normalization (i.e., "bounded", n = 102) had decreased normalization errors for N and C (median error = 0.064extrapolated normalizations (n = 30) for N (median error = 0.116$p <$ 0.0001).The isotope ranges of the two-point and three-point normalizations were then binned into three categories for statistical assessment: less than 15When only bounded matrix-matched normalizations are considered, neither two-point or three-point normalizations exhibit a significant relationship between isotope range and normalization error (Fig. 5A, 5B), with median normalization errors less than 0.1data is constrained to extrapolated matrix-mixed normalizations, normalizations with a range less than 15errors for both two-point and three-point normalizations (Fig. 5C, 5D). Under these conditions, three-point N normalizations with an isotope range less than 15(median error = 0.171to 30a range of 30< 0.0001). Similarly, three-point C normalizations with an isotope range less than 15errors (median error = 0.219to 30normalizations that were extrapolated and matrix-mixed also fared worse: N normalizations with a range less than 15higher normalization errors (median error = 0.26230range of 30and C normalizations with an isotope range less than 15significantly higher errors (median error = 0.360with a range between 15$p <$ 0.0001). No C normalizations that were

extrapolated and matrix-mixed had a range greater than 30

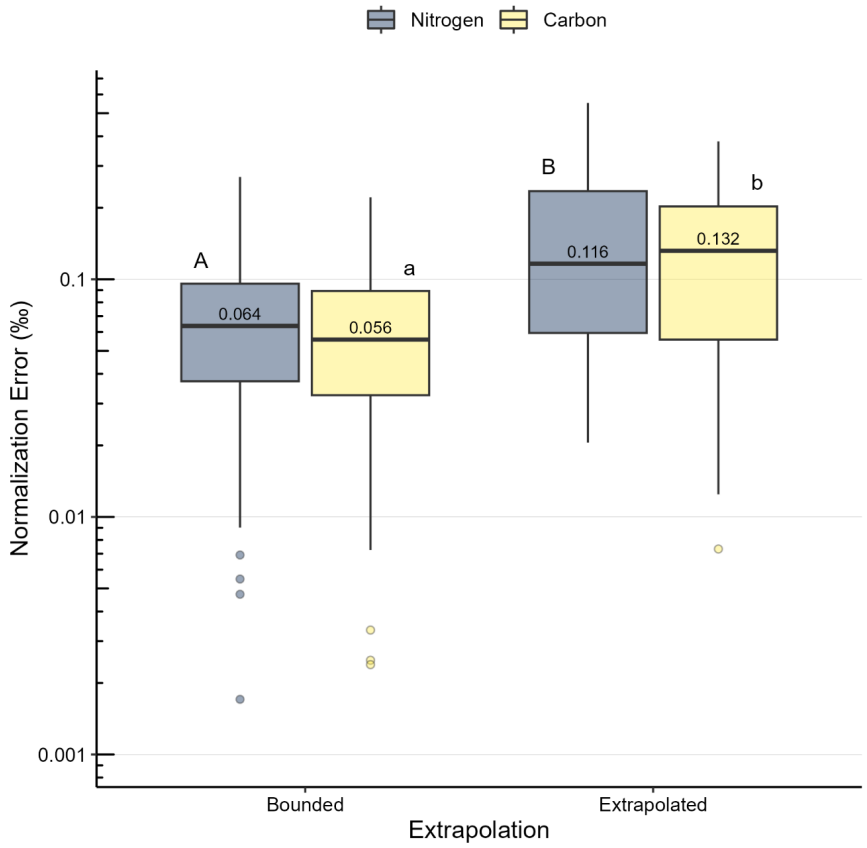

Fig. 4 : Boxplots of the accuracy of matrix-matched three-point normalizations based on whether at least one of the quality control standards falls within the isotope range of the standards (bounded) or if both quality control standards fall outside the isotope range of the standards (extrapolated). Significance is shown in the compact letter display, where any two extrapolation status with the same letter are not significantly different as per Dunn's post-hoc testing. Nitrogen (uppercase) and carbon (lowercase) letters are not compared with each other.

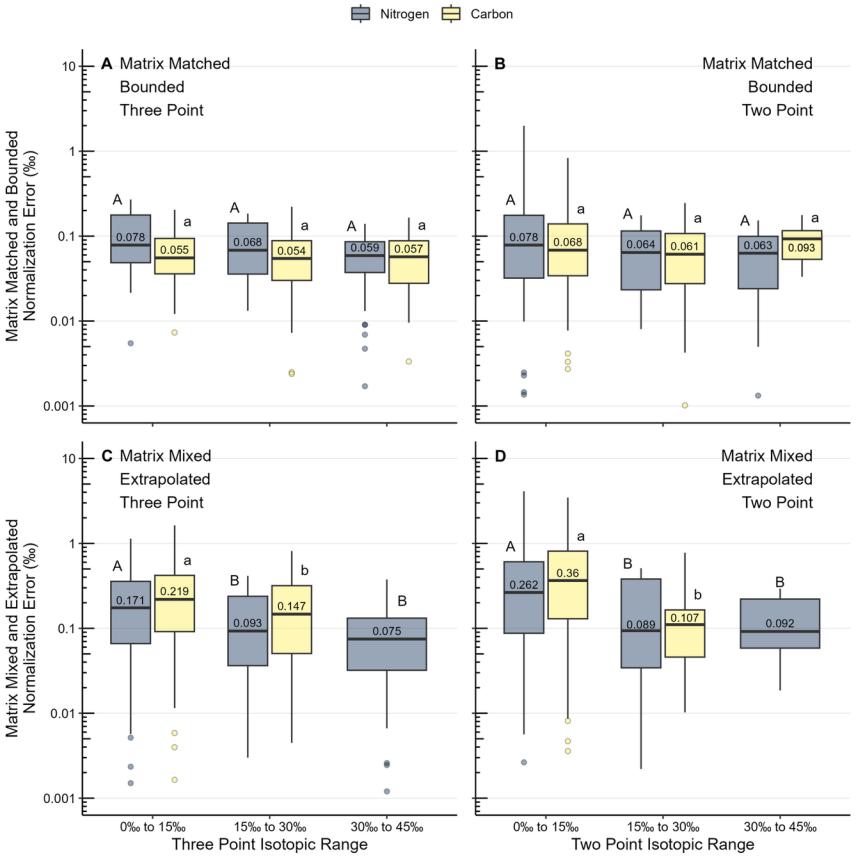

Fig. 5 : Normalization error of multipoint normalizations based on the isotope range of the standards. A: matrix matched and bounded three-point normalizations do not exhibit significant differences with isotope range; B: matrix matched and bounded two-point normalizations do not exhibit significant differences with isotope range; C: matrix mixed and extrapolated three-point normalizations have significantly higher normalization errors when the isotope range is less than 15mixed and extrapolated two-point normalizations have significantly higher normalization errors when the isotope range is less than 15Significance is shown in the compact letter display, where any two isotope ranges with the same letter are not significantly different as per Dunn's post-hoc testing. Nitrogen (uppercase) and carbon (lowercase) letters are not compared with each other.

Interlaboratory comparison of instrument linearity

To assess the linearity effect across laboratories and instrument manufacturers, five different working standards were each analyzed across a range of sample weights (n=10) while the sample dilution was held constant, thus producing a range of peak amplitudes. Both instruments exhibited a non-linear deviation in reportedδ ^{15}N as beam amplitude decreased (Fig. 6) – at U.S. EPA this deviation occurred at amplitudes below 6nA, and at UNM CSI it occurred below 2V (peak amplitude is measured in units of current (nA) on Elementar instruments and voltage (V) on ThermoFinnigan instruments). The linearity effect of reportedδ ^{15}N occurred regardless of sample matrix and resulted in substantial (>1.5^{15}N at both facilities. Instrument linearity effects for δ ^{13}C were not evident at UNM CSI across the observed range of peak amplitudes. At U.S. EPA, instrument linearity effects for δ ^{13}C were not observed below 20nA, but linearity effects were observed for beam amplitudes above 20nA (Fig. 6).

Within 48 hours of analyzing the solid samples, diagnostic reference gas linearity tests were performed for N and C at U.S. EPA and for N at UNM CSI with the same tuning and configuration. Both facilities displayed a small reference gas linearity effect for N across the tested range of peak amplitudes (Fig. 6), with a total isotope range of 0.20at UNM CSI and U.S. EPA, respectively. At the U.S. EPA, the reference gas C linearity diagnostics suggested a consistent inverse relationship between reported $\delta\ ^{13}C$ and beam height across the tested range of peak amplitudes (2-12nA), with a total isotope range of 0.33

To investigate whether the amount of combustion in the elemental analyzer varies as a function of sample weight, and thus influences the linearity results, we assessed how the ratio between peak amplitude and sample weight varied as a function of sample weight (Figure S1). Combustion effects appeared to be matrix-dependent, with high-organic matrixes exhibiting a higher variation in the ratio between peak amplitude and sample weight than other matrixes.

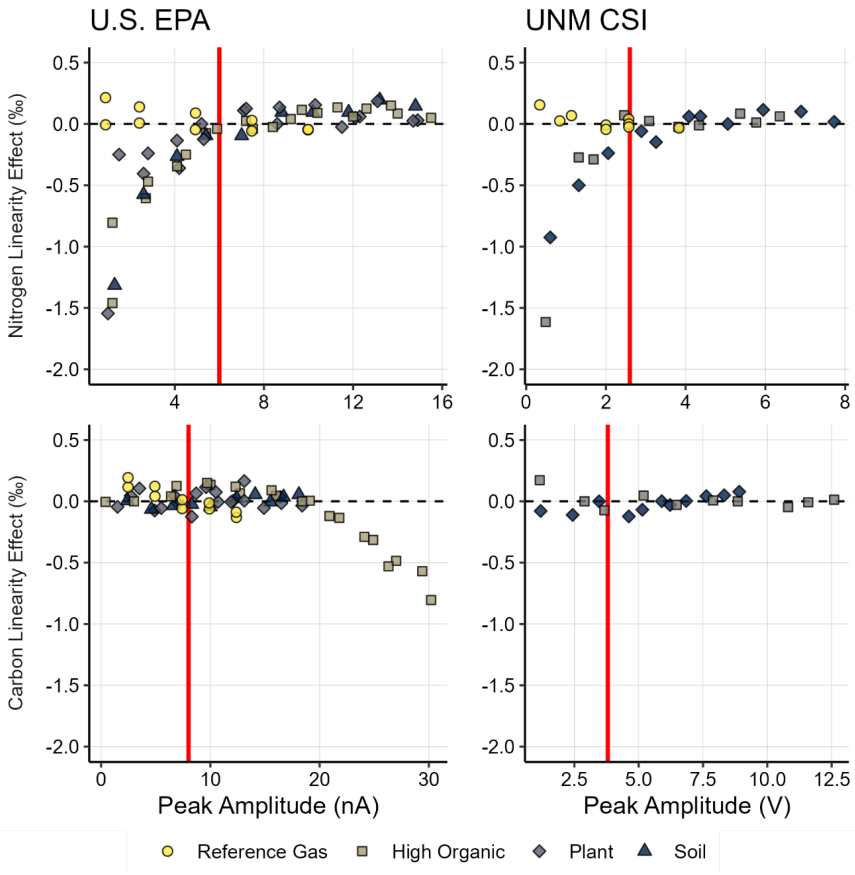

Fig. 6 : Linearity effects across sample matrixes at the U.S. EPA (left) and UNM CSI (right). The linearity effect is shown relative to the median observed isotope composition of each working standard. Vertical red lines indicate nominal peak amplitudes that were targeted for tuning and sample peak heights. Note that units of peak amplitude vary between instruments.

Discussion

1.

Two-point isotope normalizations are insufficient for EAIRMS

In this study, we conducted a total of 6272 normalizations for N and C at two laboratories using 8 certified isotope reference materials. Past work has found that one-point normalizations have larger normalization errors than two-point, three-point, and four-point normalizations[11,17,27], and our results further suggest that normalization accuracy generally improves with the number of standards. Regardless of whether we test bounded, matrix-matched normalizations or extrapolated, matrix-mixed normalizations, three-point and four-point normalizations exhibit better accuracy than one-point and two-point normalizations (Fig. 2). The lowest range of normalization errors was consistently found for three-point and four-point normalizations, when samples were bounded within the range of calibration standards, and when the sample matrix matched between samples and calibration standards (Fig 2A).

The dramatic reduction in two-point normalization accuracy when the normalization was extrapolated and matrix mixed (Fig. 2B) was surprising given that foundational literature suggests that two-point normalizations are sufficient for the normalization of stable isotope results[17,21,37]. Extrapolating beyond the normalization and mixing matrixes between the samples and standards are expected to increase normalization errors, regardless of how many standards are used. However, the median error of two-point normalizations conducted under these abnormal conditions were 64%-230% greater than corresponding three-point and four-point normalizations, with the error of some two-point normalizations exceeding 1

The sensitivity of two-point normalizations to matrix effects and extrapolation are evident in its derivation process (Eq 3-5). In a two-point normalization, the sensitivity of m to inaccuracies in$\delta_{\rho\alpha\omega(\sigma\tau\delta)}$ is a function of the isotope range of the two standards – as the isotope range of the standards decreases,m becomes more sensitive to isotope variability of the standards[21], including those due to matrix effects. Indeed, we find that two-point normalizations have the highest $m$variability (Fig. 7A), and that the variability of m is inversely related to the isotope range of the standards (Fig. 7B). The effect of$m$ on $\delta_{\tau\rho\upsilon\epsilon(\sigma\alpha\mu\pi\lambda\epsilon)}$ also varies as a function of the isotope range between the standards and the samples: as the difference between $\delta_{\rho\alpha\omega(\sigma\tau\delta)}$ and$\delta_{\rho\alpha\omega(\sigma\alpha\mu\pi\lambda\epsilon)}$ increases, errors due to an incorrect m are compounded. Changes in$\delta_{\rho\alpha\omega(\sigma\tau\delta)}$ due to matrix effects will manifest as inaccuracy when the matrix of the standards and the samples are mixed, which would subsequently be amplified by the effects of extrapolation. Thus, two-point normalizations that are extrapolated and matrix-mixed have poor performance – particularly with a small isotope range between the two standards (Fig. 5D). Overall, this work questions whether two-point linear normalizations are sufficient for biological applications of EAIRMS – in our study, many two-point normalizations were inferior to one-point normalizations because the actual $m$was close to 1, the assumed slope for 1-point normalization. Notably, three-point and four-point normalizations are much more resilient to the effects of extrapolation and matrix mixing (Fig. 2B), even when the isotope range was less than 15users of EAIRMS systems use at least 3 calibration standards to compose their normalization curve.

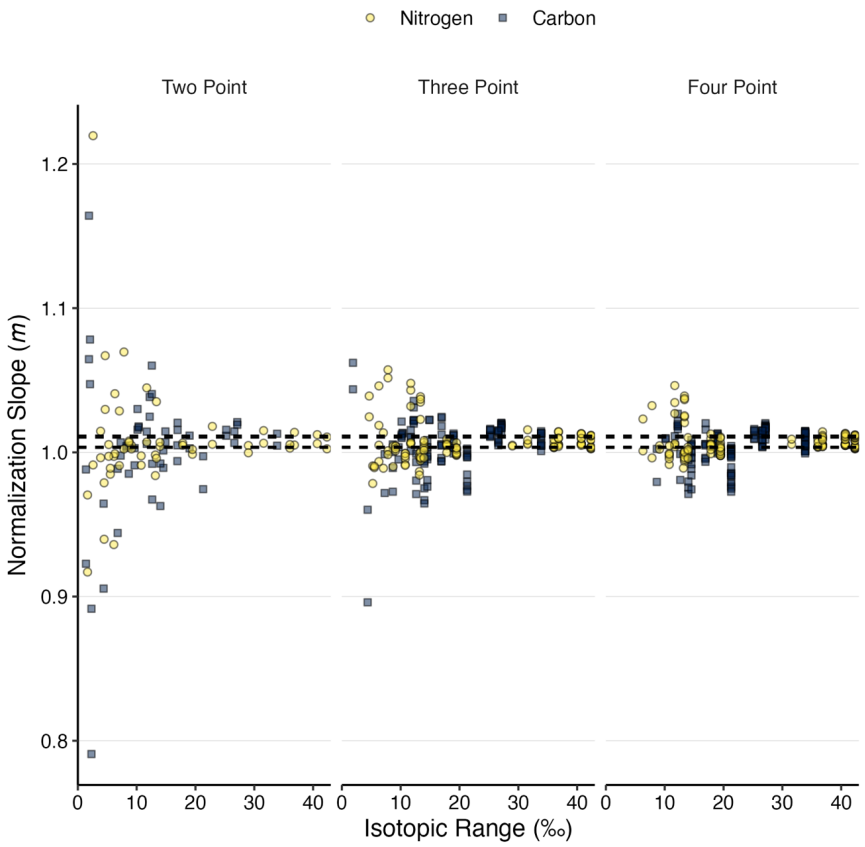

Fig. 7 : The slope of all multipoint normalizations by isotope range of the standards. Dashed horizontal lines indicate the slopes for each facility derived from an 8-point normalization composed of all certified reference standards.

Normalizations are significantly impacted by matrix and extrapolation effects

Even when results are constrained to three-point normalizations, the matrix of the standard relative to the quality controls has a significant effect on the accuracy of the normalization (Fig. 3). Bounded normalizations where the matrixes of the standards were mixed relative to the quality controls exhibited median errors 63%-195% greater than those where the matrixes were matched. The mechanism behind this matrix effect is unclear, particularly because the isotope composition of the plant-based standards were quantified using high organic certified reference materials[20]. In accordance with typical EAIRMS usage, the plant-based standards were weighed to a higher mass and analyzed with a higher oxygen dosing and a greater sample dilution for C, and thus it is possible that these instrumentation factors are contributing to the matrix effect observed in this study. Although past work has suggested that matrix matching between organic and inorganic samples would reduce normalization errors[28], this study posits that matrix effects should be considered even within organic samples. In studies where matrix-matched certified reference materials do not exist (e.g., sediment; glass fiber filters), the effects of matrix-mixing should be considered as a source of imprecision that will not be reflected in the variance of the standards alone.

Similarly, this study shows that extrapolating beyond the isotope range of the standards– not an uncommon occurrence when analyzing a wide variety of biological samples – increases median normalization errors by 81%-135%, even when the analysis is constrained to matrix-matched normalizations (Fig. 4). Normalizations with a smaller isotope range are more likely to require extrapolation, but the lack of significant relationship

between isotope range and normalization error for bounded, matrix-matched, three-point normalizations (Fig. 5) indicates that extrapolation is the primary factor driving inaccuracy. Overall, our results provide experimental evidence to support the emerging consensus[37] that calibration standards for EAIRMS should have isotope values that span the full natural range of the measured elements, regardless of the isotope values of the samples being analyzed.

Instrument linearity cannot be predicted by reference gas diagnostics alone.

Although reference gas linearity diagnostics are anecdotally used as a means of assessing instrument performance, we quantified instrument linearity using replicate analyses of working standards across a range of sample weights. At both facilities, a large linearity effect was observed in the reported isotope composition of the working standards that was not reflected by reference gas linearity diagnostics, which inject pulses of reference gas into the IRMS to assess the linearity effect. The linearity effect does not correspond to incomplete combustion in the elemental analyzer (Fig. S1), suggesting that other factors, such as a N blank introduced during sample preparation, are driving our observations. These results show that, while reference gas linearity diagnostics may be useful for confirming normal instrument operation, they are not representative of the linearity effect that will be observed when analyzing solid samples. We suggest running replicate measurements of a solid standard at varying sample weights to determine the peak amplitudes at which linearity has an effect and applying a linearity correction curve when necessary. In our study, N linearity was observed at peak amplitudes corresponding to ˜0.06 mg of N, meaning that instrument linearity is of particular importance for samples with a low N content such as sediment and plant tissue. The high magnitude of the linearity effect observed in both facilities across a range of sample matrixes reinforces the importance of this instrument effect for any facility measuring stable isotopes in biological samples.

Conclusion – best practices for biological applications of EAIRMS

Through an experimental assessment of isotope normalizations across a variety of sample matrixes and isotope ranges, we assessed how the number, matrix, and isotope range of the calibration standards effected normalization error on two EAIRMS systems. In the first known assessment of normalization methods for both N and C, we found that three-point and four-point normalizations have the lowest normalization errors and are most resilient to the effects of sample matrix effects and extrapolation.

Past work has identified two-point normalizations as a common and acceptable means of normalizing isotope results[2,17]. In contrast, some of the observed two-point normalizations had deviations more than 1three-point and four-point normalizations. We posit that two-point normalizations are vulnerable to the effects of matrix-mixing and extrapolation and should not be considered sufficient for EAIRMS normalization in biological applications. Although normalizations using at least three standards were more resilient to these factors, we found that mixing the matrix between the samples and the standards and extrapolating outside of the curve reduced accuracy regardless of how many standards were analyzed. No significant impact was observed with the isotope range of three-point normalization but maximizing the isotope range will reduce the likelihood of extrapolation. Thus, we recommend users of EAIRMS systems normalize their results using at least three calibration standards that span a large isotope range and are matrix matched with the samples being analyzed, and to include at least one additional independent quality control standard.

In our interlaboratory comparison of instrument linearity, we found that linearity error was substantial, especially for N, regardless of instrument or sample matrix. Diagnostic reference gas linearity testing was unable to reproduce the observed linearity effect, suggesting that the reference gas is not a useful predictor of real-world instrument linearity. Although reference gas diagnostic testing may be a beneficial tool for assessing nominal instrument operation, the linearity response of the instrument should be assessed experimentally by analyzing solid standards across a range of masses.

References

1. Brand WA, Coplen TB. Stable isotope deltas: tiny, yet robust signatures in nature. *Isotopes in Environmental and Health Studies.* 2012;48(3):393-409.

2. Jardine TD, Cunjak RA. Analytical error in stable isotope ecology. *Oecologia.* 2005;144(4):528-533.

3. Carter JF, Fry B. Ensuring the reliability of stable isotope ratio data—beyond the principle of identical treatment. *Analytical and Bioanalytical Chemistry.* 2013;405(9):2799-2814.

4. Skrzypek G, Allison CE, Böhlke JK, et al. Minimum requirements for publishing hydrogen, carbon, nitrogen, oxygen and sulfur stable-isotope delta results (IUPAC Technical Report). *Pure and Applied Chemistry.* 2022.

5. Brenna JT, Corso TN, Tobias HJ, Caimi RJ. High-precision continuous-flow isotope ratio mass spectrometry. *Mass Spectrometry Reviews.* 1997;16(5):227-258.

6. Preston T, Owens NJP. Interfacing an automatic elemental analyser with an isotope ratio mass sepectrometer: the potential for fully automated total nitrogen and nitrogen-15 analysis. *Analyst.*1983;108(1289):971-977.

7. Werner RA, Bruch BA, Brand WA. ConFlo III – an interface for high precision δ13C and δ15N analysis with an extended dynamic range. *Rapid Communications in Mass Spectrometry.* 1999;13(13):1237-1241.

8. Werner RA, Brand WA. Referencing strategies and techniques in stable isotope ratio analysis. *Rapid Communications in Mass Spectrometry.* 2001;15(7):501-519.

9. Santrock J, Studley SA, Hayes JM. Isotopic analyses based on the mass spectra of carbon dioxide. *Analytical Chemistry.*1985;57(7):1444-1448.

10. Craig H. Isotopic standards for carbon and oxygen and correction factors for mass-spectrometric analysis of carbon dioxide. *Geochimica et Cosmochimica Acta.* 1957;12(1):133-149.

11. Coplen TB, Brand WA, Gehre M, et al. New Guidelines for δ13C Measurements. *Analytical Chemistry.* 2006;78(7):2439-2441.

12. Hélie J-F, Adamowicz-Walczak A, Middlestead P, Chartrand MMG, Mester Z, Meija J. Discontinuity in the Realization of the Vienna Peedee Belemnite Carbon Isotope Ratio Scale. *Analytical Chemistry.*2021;93(31):10740-10743.

13. Brand WA. High Precision Isotope Ratio Monitoring Techniques in Mass Spectrometry. *Journal of Mass Spectrometry.* 1996;31(3):225-235.

14. De Bièvre P, De Laeter JR, Peiser HS, Reed WP. Reference materials by isotope ratio mass spectrometry. *Mass Spectrometry Reviews.*1993;12(3):143-172.

15. Merritt DA, Brand WA, Hayes JM. Isotope-ratio-monitoring gas chromatography-mass spectrometry: methods for isotopic calibration. *Organic Geochemistry.* 1994;21(6):573-583.

16. Gröning M. Chapter 40 - International Stable Isotope Reference Materials. In: de Groot PA, ed. *Handbook of Stable Isotope Analytical Techniques.* Amsterdam: Elsevier; 2004:874-906.

17. Paul D, Skrzypek G, Fórizs I. Normalization of measured stable isotopic compositions to isotope reference scales – a review. *Rapid Communications in Mass Spectrometry.* 2007;21(18):3006-3014.

18. Qi H, Coplen TB, Geilmann H, Brand WA, Böhlke JK. Two new organic reference materials for δ13C and δ15N measurements and a new value for the δ13C of NBS 22 oil. *Rapid Communications in Mass Spectrometry.* 2003;17(22):2483-2487.

19. Schimmelmann A, Qi H, Coplen TB, et al. Organic Reference Materials for Hydrogen, Carbon, and Nitrogen Stable Isotope-Ratio Measurements: Caffeines, n-Alkanes, Fatty Acid Methyl Esters, Glycines, l-Valines, Polyethylenes, and Oils. *Analytical Chemistry.*2016;88(8):4294-4302.

20. Schimmelmann A, Qi H, Dunn PJH, et al. Food Matrix Reference Materials for Hydrogen, Carbon, Nitrogen, Oxygen, and Sulfur Stable Isotope-Ratio Measurements: Collagens, Flours, Honeys, and Vegetable Oils. *Journal of Agricultural and Food Chemistry.*2020;68(39):10852-10864.

21. Skrzypek G, Sadler R, Paul D. Error propagation in normalization of stable isotope data: a Monte Carlo analysis. *Rapid Communications in Mass Spectrometry.* 2010;24(18):2697-2705.

22. Oczkowski A, Santos E, Gray A, et al. Tracking the Dynamic Ecological History of a Tropical Urban Estuary as it Responds to Human Pressures. *Ecosystems.* 2020;23(2):231-245.

23. Balint SJ, Joyce E, Pennino S, Oczkowski A, McKinney R, Hastings MG. Identifying Sources and Impacts of Precipitation-Derived Nitrogen in Narragansett Bay, RI. *Estuaries and Coasts.* 2021.

24. Oczkowski A, Kreakie B, McKinney RA, Prezioso J. Patterns in Stable Isotope Values of Nitrogen and Carbon in Particulate Matter from the Northwest Atlantic Continental Shelf, from the Gulf of Maine to Cape Hatteras. *Frontiers in Marine Science.* 2016;3(252).

25. Slawyk G, Raimbault P. Simple procedure for simultaneous recovery of dissolved inorganic and organic nitrogen in 15N-tracer experiments and improving the isotopic mass balance. *Marine Ecology Progress Series.* 1995;124:289-299.

26. Coplen TB. Guidelines and recommended terms for expression of stable-isotope-ratio and gas-ratio measurement results. *Rapid Communications in Mass Spectrometry.* 2011;25(17):2538-2560.

27. Coplen TB, Brand WA, Gehre M, et al. After two decades a second anchor for the VPDB δ13C scale. *Rapid Communications in Mass Spectrometry.* 2006;20(21):3165-3166.

28. Gentile N, Rossi MJ, Delémont O, Siegwolf RTW. δ15N measurement of organic and inorganic substances by EA-IRMS: a speciation-dependent procedure. *Analytical and Bioanalytical Chemistry.*2013;405(1):159-176.

29. Hedgespeth ML, Taylor DL, Balint S, Schwartz M, Cantwell MG. Ecological characteristics impact PFAS concentrations in a U.S. North Atlantic food web. *Science of The Total Environment.*2023;880:163302.

30. Krause JR, Gannon ME, Oczkowski AJ, et al. Tidal Flushing Rather Than Non-Point Source Nitrogen Pollution Drives Nutrient Dynamics in A Putatively Eutrophic Estuary. *Water.* 2023;15(1):15.

31. Meier-Augenstein W, Schimmelmann A. A guide for proper utilisation of stable isotope reference materials. *Isotopes in Environmental and Health Studies.* 2019;55(2):113-128.

32. Bond AL, Hobson KA. Reporting Stable-Isotope Ratios in Ecology: Recommended Terminology, Guidelines and Best Practices. *Waterbirds.* 2012;35(2):324-331, 328.

33. Gentile N, Besson L, Pazos D, Delémont O, Esseiva P. On the use of IRMS in forensic science: Proposals for a methodological approach. *Forensic Science International.* 2011;212(1):260-271.

34. Ohlsson EAK, Wallmark HP. Novel calibration with correction for drift and non-linear response for continuous flow isotope ratio mass spectrometry applied to the determination of δ15N, total nitrogen, δ13C and total carbon in biological material+. *Analyst.*1999;124(4):571-577.

35. R Core Team. A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.* 2022.

36. Brooks JR, Rugh WD, Werner RA. Tree-Ring Stable Isotope Measurements: The Role of Quality Assurance and Quality Control to Ensure High Quality Data. In: Siegwolf RTW, Brooks JR, Roden J, Saurer M, eds. *Stable Isotopes in Tree Rings: Inferring Physiological, Climatic and Environmental Responses.* Cham: Springer International Publishing; 2022:191-213.

37. Skrzypek G. Normalization procedures and reference material selection in stable HCNOS isotope analyses: an overview. *Analytical and Bioanalytical Chemistry.* 2013;405(9):2815-2823.