Best analytical practices for estimating consumer reliance on basal food sources using bulk stable isotopes

Mattia Ghilardi[1,*], Renato A. Morais[1], Simon J. Brandl[2], Jordan M. Casey[2], Alexandre Mercière[1], Fabien Morat[1], Nina M. D. Schiettekatte[3], Mohsen Kayal[4,5], Yves Letourneur[6], Valeriano Parravicini[1]

[1]PSL Université Paris, CRIOBE, EPHE-CNRS-UPVD, 58 Avenue Paul Alduy, 66100 Perpignan, France
[2]Department of Marine Science, The University of Texas at Austin, Marine Science Institute, 750 Channel View Drive, 78373 Port Aransas, Texas, USA
[3]MARE – Marine and Environmental Sciences Centre & ARNET – Aquatic Research Infrastructure Network Associated Laboratory, Laboratório Marítimo da Guia, Faculdade de Ciências, Universidade de Lisboa, Cascais, Portugal
[4]UMR Entropie (IRD-IFREMER-UR-UNC), BP A5, 98848 Nouméa, New Caledonia
[5]Station Marine d'Endoume, Marseille, France
[6]UMR Entropie (IRD-IFREMER-UR-UNC), EMR SantEco, Université de la Nouvelle Calédonie, BP R4, 98851 Nouméa cedex, New Caledonia
[*]Corresponding author: mattia.ghilardi91@gmail.com

Abstract

1. Bulk stable isotopes have been utilised by ecologists for decades to trace the assimilation of basal food sources across trophic positions. However, there is a lack of guidance on the post-laboratory analytical workflow, leading to inconsistencies in how isotopic data are processed and interpreted.

2. We provide a comprehensive description of this necessary post-laboratory analytical workflow, including the mathematical rationale underlying the estimation of consumer reliance on basal sources. Using simulated data, we test the sensitivity of estimated basal source proportions to the number of isotopic baselines used for the estimation of consumer trophic position and to varying sample sizes of sources and consumers. Finally, we apply the workflow to a coral reef food web to quantify fish reliance on benthic and pelagic primary production.

3. Our results indicate that estimated basal source proportions are highly sensitive to bias in the estimation of consumer trophic position, suggesting that trophic position estimation requires careful consideration of the number and nature of selected isotopic baselines. We provide an R function to quantify this bias. By contrast, the precision and uncertainty—but not the accuracy—of these proportions are influenced by the sample sizes of consumers and sources. Thus, small sample sizes can still yield valid estimations and can be included. The case study further highlights the sensitivity of estimated proportions to the selection of

trophic discrimination factors (TDFs), and we suggest performing a sensitivity analysis with alternative TDFs.

4. While compound-specific stable isotope analysis is gaining prominence, bulk stable isotope analysis remains a cornerstone for field ecologists due to its high throughput, cost-effectiveness and availability of data collected over the last 20 years. By providing a detailed analytical framework and identifying future directions for model development, we offer a robust analytical foundation for future isotopic studies.

Keywords: mixing models, trophic position, trophic discrimination factors, food web analysis, methodological framework, resource assimilation, isotope ecology

1 Introduction

Tracing energy and matter across food webs has been a central endeavour since Lindeman's (1942) seminal study on trophodynamics. Declining cost for the analysis of biotracers (i.e., biological markers such as stable isotopes and fatty acids; Appendix S1) and the advancement of statistical methods (e.g., quantitative fatty acid signature analysis and stable isotope mixing models [SIMMs] used to estimate proportional source contributions to consumers; Appendix S1) have facilitated their application in ecological research, including in large-scale monitoring efforts conducted by non-specialists (e.g., Bird et al., 2018; Lorrain et al., 2020; Pethybridge et al., 2018). Consequently, standardised guidance on biotracer data and modelling procedures has become essential. While foundational work has established best practices for the general use of SIMMs in food web studies (e.g., Phillips et al., 2014), there remains a lack of clear guidance regarding the post-laboratory analytical workflow for using bulk stable isotopes to estimate consumer reliance on basal food sources (i.e., primary producers and particulate organic matter) across trophic positions (i.e., a continuous measure of an organism's position in a food web; Appendix S1).

Over the past few decades, stable isotope analysis (SIA; the measurement of the relative isotopic abundance of elements within a sample; Appendix S1) has become a primary tool for the study of bulk tissues in animal ecology (Gannes et al., 1997; Martínez del Rio et al., 2009). The standardisation of laboratory protocols has led to the vast accumulation of experimental and field data (Martínez del Rio et al., 2009), facilitating the development of Bayesian SIMMs to reconstruct animal diets and trace basal source contributions (e.g., Ferger et al.,

2013; Garcia et al., 2018; Hilting et al., 2013; McMahon et al., 2021; Moustaka et al., 2024; Reid et al., 2008; Santos et al., 2020; Solomon et al., 2011; Zapata-Hernández et al., 2021). While other more accurate biotracer methods such as compound-specific isotope analysis are increasingly used to study animal diets (Nielsen et al., 2018; Pickett et al., 2024), their application is constrained by high financial costs, time-consuming and highly-specialised sample processing (Vane et al., 2025); therefore, bulk SIA remains the mainstream approach. However, accurately interpreting bulk SIA data requires addressing several confounding factors, such as isotopic variability at the base of the food web and trophic discrimination factor (TDF; the difference in isotopic values between consumer tissues and their diet; Appendix S1), as isotopic values are influenced by both the consumer's environment and its physiological processing of nutrients (Martínez del Rio et al., 2009).

Estimating the contribution of basal sources to secondary or higher-order consumers using bulk SIA and SIMMs requires consumer signatures to be corrected for total trophic discrimination between them and the food web base (Reid et al., 2008). This necessitates accurate estimates of consumer trophic positions and TDFs of individual trophic steps, introducing multiple sources of error. While total discrimination can be directly calculated for simplified food webs by summing the TDFs of each trophic link (Pickett et al., 2024; Reid et al., 2008), the standard approach for more complex food webs is to multiply an average TDF per trophic step by the number of trophic steps between consumers and basal sources (i.e., their difference in trophic position) (Phillips et al., 2014). This approach assumes that discrimination remains constant across trophic levels—an assumption rarely verified. In addition, a thorough evaluation of its mathematical rationale and sensitivity to trophic position estimation is currently lacking. Most studies address this by using either literature-derived trophic position values (e.g., Hilting et al., 2013) or estimates derived from one or two isotopic baselines (i.e., the isotopic composition of a basal component of a food web; Appendix S1), such as primary consumers (e.g., Zapata-Hernández et al., 2021) or the mean δ^{15}N of all primary producers (e.g., Ferger et al., 2013; Garcia et al., 2018; Moustaka et al., 2024). However, the impact of these methodological choices on the resulting source proportions is frequently unknown, highlighting the need for a detailed evaluation of the analytical workflow.

Here, we describe a comprehensive workflow to estimate consumer reliance on basal food sources using bulk SIA of carbon and nitrogen (and potentially other elements such as sulfur or hydrogen) and provide the mathematical rationale underpinning this approach. Using simulated data, we: (1) test the accuracy of the analytical workflow when using one or two baselines to estimate consumer trophic position; and (2) investigate how the estimated proportional contributions of sources are affected by consumer or source sample sizes. Finally, we illustrate this analytical workflow using a case study of a coral reef food web in Mo'orea, French Polynesia, estimating the reliance of 194 coral reef fish species on pelagic (external) and benthic (internal) primary production.

2 Materials and Methods

2.1 Analytical workflow and mathematical rationale

Estimating consumer reliance on basal food sources using bulk SIA relies on a core mathematical principle: the stable isotope value of a mixture (e.g., consumer tissue) is determined by the proportional contributions of its sources and their respective stable isotope values. This general SIMM is expressed as:

$$y_j = \sum_{k=1}^{K} p_k \mu_{jk}, \tag{1}$$

where y_j is the stable isotope value of the mixture for the j^{th} isotope, μ_{jk} is the mean stable isotope value of the k^{th} source (of K total sources), and p_k is the proportional contribution of the k^{th} source to the mixture, such that $\sum_{k=1}^{K} p_k = 1$.

To apply this model to animal diet and food web analysis, the mean isotopic values of food sources must be corrected for trophic discrimination:

$$y_j = \sum_{k=1}^{K} p_k (\mu_{jk} + \Lambda_{jk}), \tag{2}$$

where Λ_{jk} is the total TDF accumulated between the consumer and the k^{th} source for the j^{th} isotope.

Directly measuring Λ_{jk} is generally impractical for higher-order consumers (TP > 2), as it requires feeding experiments spanning multiple trophic links. This framework resolves this limitation via a three-step analytical workflow: (1) estimating consumer trophic position relative to isotopic baselines; (2) mathematically rescaling consumer isotopic values to the food web base to remove accumulated discrimination; and (3) fitting Bayesian SIMMs to the rescaled data to estimate proportional contributions of basal sources (Figure 1).

2.1.1 Preliminary steps: Data preparation

1. *$\delta^{13}C$ lipid correction.* Prior to analysys, carbon stable isotope data must be evaluated for lipid-derived bias. Lipids are significantly depleted in ^{13}C compared to proteins and carbohydrates (DeNiro & Epstein, 1977; McConnaughey & McRoy, 1979), meaning high variability in lipid content among organisms and tissues can confound source proportion estimations (Kiljunen et al., 2006). Animal tissue samples (consumers and baselines) with high lipid content, typically indicated by C:N ratios >3.5 (Post et al., 2007), should be corrected either through chemical lipid extraction during sample preparation or via mathematical normalisation of δ^{13}C values using established equations based on C:N ratios (e.g., Skinner et al., 2016 for fish tissues). For plant samples (basal sources), extraction or normalisation should be performed on samples with >40% carbon, and normalisation should be based on percent carbon rather than the C:N ratio (Post et al., 2007). Correcting carbon signatures from the outset prevents lipid-induced noise from propagating through the analytical workflow.

2. *Selection of basal sources.* The number of basal sources (K) must be carefully managed. While all relevant sources should ideally be included—an implicit but rarely met assumption in any mixing model—exceeding the $n + 1$ limit (where n is the number of stable isotopes) creates an underdetermined system lacking an exact analytical solution. Bayesian SIMMs, however, provide probabilistic solutions for underdetermined systems, but often yield broad, multimodal posterior distributions because multiple, disparate source combinations can produce identical consumer signatures (Osada et al., 2021; Phillips et al., 2014; Stock et al., 2018). To improve model inference, re-

searchers may group sources *a priori* based on isotopic similarity, thereby reducing the number of possible analytical solutions, or aggregate their estimated proportions *a posteriori* based on functional similarity, potentially yielding more constrained results (see Phillips et al., 2005; Phillips et al., 2014; Stock et al., 2018).

3. *Selection of isotope-specific TDFs (λ_j).* This framework assumes constant discrimination across trophic levels, meaning total discrimination (Λ_{jk}) is the product of the TDF of a single trophic step (λ_{jk}) and the difference in trophic position between the consumer (TP_c) and the k^{th} source (TP_k) (Phillips et al., 2014):

$$\Lambda_{jk} = \lambda_{jk}(\mathrm{TP}_c - \mathrm{TP}_k). \tag{3}$$

Because source-specific values (λ_{jk}) are generally unknown for primary producers, they are replaced by a common isotope-specific TDF (λ_j) obtained from meta-analyses (e.g., 0.39 ± 1.3‰ for δ^{13}C and 3.4 ± 0.98‰ for δ^{15}N, Post, 2002), ideally matched to tissue and taxonomic groups. When all k sources represent the base of the food web ($\mathrm{TP}_k = 1$), the equation simplifies to:

$$\Lambda_{jk} = \Lambda_j = \lambda_j(\mathrm{TP}_c - 1). \tag{4}$$

While discrimination varies across individual feeding links (Canseco et al., 2022; Caut et al., 2009; Stephens et al., 2023), using an average stepwise TDF (λ_j) provides a consistent discrimination across multiple trophic transfers and is considered a robust approximation when integrated across the diverse pathways of an entire food web (Post, 2002). Substituting Equation 4 into Equation 2 yields the foundational workflow equation:

$$y_j = \sum_{k=1}^{K} p_k(\mu_{jk} + \lambda_j(\mathrm{TP}_c - 1)). \tag{5}$$

4. *Selection of isotopic baselines.* Rather than primary producers, long-living primary consumers (e.g., grazers, filter feeders) should be used as isotopic baselines to integrate baseline temporal variations (Cabana & Rasmussen, 1996; Post, 2002; Vander

Zanden & Rasmussen, 1999, 2001). The suitability of these baselines should be verified by estimating their basal source contributions via SIMMs. While this verification is often completed *a posteriori*, it provides a critical diagnostic check on whether the chosen baselines align with the intended trophic pathways. If a baseline consumer, assumed to represent a specific pathway (e.g., benthic), is found to derive most of its energy from an alternative pathway (e.g., pelagic), its baseline status is violated. When consumers derive energy from multiple discrete pathways (e.g., benthic-pelagic, terrestrial-aquatic), multiple baseline consumers are required (Post, 2002).

2.1.2 Step 1: Estimating consumer trophic position (TP_c)

To solve Equation 5, consumer trophic position (TP_c) must be estimated relative to either one or two isotopic baselines using models derived from the general mixing model (Equation 2):

- *One-baseline model.* When a single baseline consumer represents the food web base, TP_c can be estimated using a single isotope, typically $\delta^{15}N$ (Cabana & Rasmussen, 1996; Post, 2002). In this framework, Equation 2 becomes

$$\delta^{15}N_c = \delta^{15}N_b + \lambda_N(TP_c - TP_b), \tag{6}$$

  which, rearranged to solve for TP_c, gives

$$TP_c = TP_b + \frac{\delta^{15}N_c - \delta^{15}N_b}{\lambda_N}, \tag{7}$$

  where TP_b represents the trophic position of the baseline and λ_N is the isotope-specific TDF for $\delta^{15}N$.

- *Two-baseline model.* In systems with two distinct trophic pathways, a dual-baseline, dual-isotope mixing model ($\delta^{13}C$ and $\delta^{15}N$) solves for both TP_c and the proportional contribution of the two baselines that typically have an identical trophic position ($TP_{b1} = TP_{b2} = TP_b$) (Quezada-Romegialli et al., 2018):

$$\delta^{15}N_c = p\delta^{15}N_{b1} + (1-p)\delta^{15}N_{b2} + \lambda_N(TP_c - TP_b) \tag{8}$$

$$\delta^{13}\mathrm{C}_c = p\delta^{13}\mathrm{C}_{b1} + (1-p)\delta^{13}\mathrm{C}_{b2} + \lambda_{\mathrm{C}}(\mathrm{TP}_c - \mathrm{TP}_b), \tag{9}$$

Under this model, TP_c is estimated by accounting for the isotopic signatures of both baselines:

$$\mathrm{TP}_c = \mathrm{TP}_b + \frac{\delta^{15}\mathrm{N}_c - p\delta^{15}\mathrm{N}_{b1} - (1-p)\delta^{15}\mathrm{N}_{b2}}{\lambda_{\mathrm{N}}}, \tag{10}$$

where the proportional contribution of the first baseline (p) can be expressed as

$$p = \frac{\delta^{13}\mathrm{C}_c - \delta^{13}\mathrm{C}_{b2} - \lambda_{\mathrm{C}}(\mathrm{TP}_c - \mathrm{TP}_b)}{\delta^{13}\mathrm{C}_{b1} - \delta^{13}\mathrm{C}_{b2}}, \tag{11}$$

and λ_{C} is the isotope-specific TDF for $\delta^{13}\mathrm{C}$.

These models are the most frequently used across the literature (Kjeldgaard et al., 2021), and traditional point-estimate calculations have recently been replaced by Bayesian models that propagate parameter uncertainty into the TP_c estimation (e.g., the R package *tRophicPosition*, Quezada-Romegialli et al., 2018), providing a more modern and robust approach.

While these equations rely on constant discrimination, an alternative framework accounting for systematic decreases in TDF values at higher trophic levels (Hussey et al., 2014) is detailed in Appendix S2.

2.1.3 Step 2: Isotopic rescaling of consumers

Once TP_c is estimated, the raw isotopic signature of each individual consumer (y_j) is mathematically rescaled to the food web base (y_j^*) by rearranging Equation 5:

$$y_j^* = y_j - \lambda_j(\mathrm{TP}_c - 1) = \sum_{k=1}^{K} p_k \mu_{jk}. \tag{12}$$

This transformation isolates y_j^*, representing the consumer's signature as if it fed directly at the base of its food web (TP = 1), thus stripping away the confounding effect of accumulated trophic discrimination. Standard Bayesian SIMMs cannot process individual-level uncertainty in consumer signatures during source contribution estimation, preventing uncertainty in trophic position and discrimination from propagating into the rescaled signatures (y_j^*). Consequently, the mean of the TDF (λ_j) and the mode of the estimated TP_c posterior

distribution are used as point estimates for this step. The mode is selected because it provides a more robust measure of central tendency than the mean or median when consumer sample sizes are small and the posterior distribution exhibits non-normal geometry.

2.1.4 Evaluation step: Mixing model validation

Before implementing the final mixing models, the rescaled consumer values (y_j^*) and raw basal source signatures must be plotted together in isospace to ensure that the consumer signatures fall within the boundaries of the mixing polygon defined by the basal sources, so that mathematically valid source contribution solutions exist (Phillips et al., 2014). If consumers fall consistently outside, researchers must systematically re-evaluate the preliminary steps to identify missing basal sources or more appropriate baselines and TDFs. Conversely, if only isolated individuals fall outside the mixing polygon, they should be flagged as potential outliers. Because extreme values can cause model convergence issues despite the integrated error structures of Bayesian SIMMs that account for natural isotopic variation, a probabilistic framework (e.g., Smith et al., 2013) can be employed to quantify the probability that an individual's isotopic signature is explained by the model. An extremely low probability indicates that the individual cannot be modelled by the sources at hand. Rather than serving as an automatic rule for exclusion, this metric provides a diagnostic check; the ultimate decision whether to exclude the outliers and continue the inference, collect additional basal source data to expand the polygon, or abort the analysis entirely rests with the researcher. Regardless of the initial results, performing sensitivity analyses with alternative baselines or TDFs remains good practice. Additionally, computing the normalised surface area of the mixing polygon (Brett, 2014) using the *calc_area()* function in R package *MixSIAR* (Stock et al., 2018) may be useful to detect weakly informative model geometry caused by poor source separation or high source variance.

2.1.5 Step 3: Estimating basal source proportions (p_k)

Finally, a Bayesian SIMM is fitted to the rescaled consumer values (y_j^*) and the isotopic signatures of basal sources to estimate the proportional contributions (p_k). Because all accumulated trophic discrimination was accounted for and removed in step 2, the TDFs

within the mixing model must be explicitly set to zero (mean = 0, SD = 0).

Models can initially be run using a "uninformative", generalist prior implying that all source combinations are equally likely (i.e., a symmetric Dirichlet distribution where all hyperparameters $\alpha_k = 1$). If the model exhibits poor convergence or yields biologically unrealistic results, informative priors should be constructed using expert system knowledge of the system or literature estimates. For example, for specialist consumers with clear trophic affinities (e.g., grazing herbivores, or planktivores), it is sensible to assign a higher prior weight (α_k) to the basal source representing the primary pathway, while decreasing the weights of alternative sources. To prevent these priors from overriding the actual stable isotope data, the Dirichlet hyperparameters should be rescaled so that their sum equals the total number of sources (K, Stock et al., 2018).

These models can be fitted using Markov Chain Monte Carlo (MCMC) or Fixed Form Variational Bayes algorithms across several R packages such as *MixSIAR* (Stock et al., 2018), *simmr* (Govan et al., 2023), or *cosimmr* (Govan et al., 2024). Among these, MixSIAR is currently the most feature-rich, allowing for the inclusion of covariates, hierarchical structures, and diverse error formulations (Stock et al., 2018).

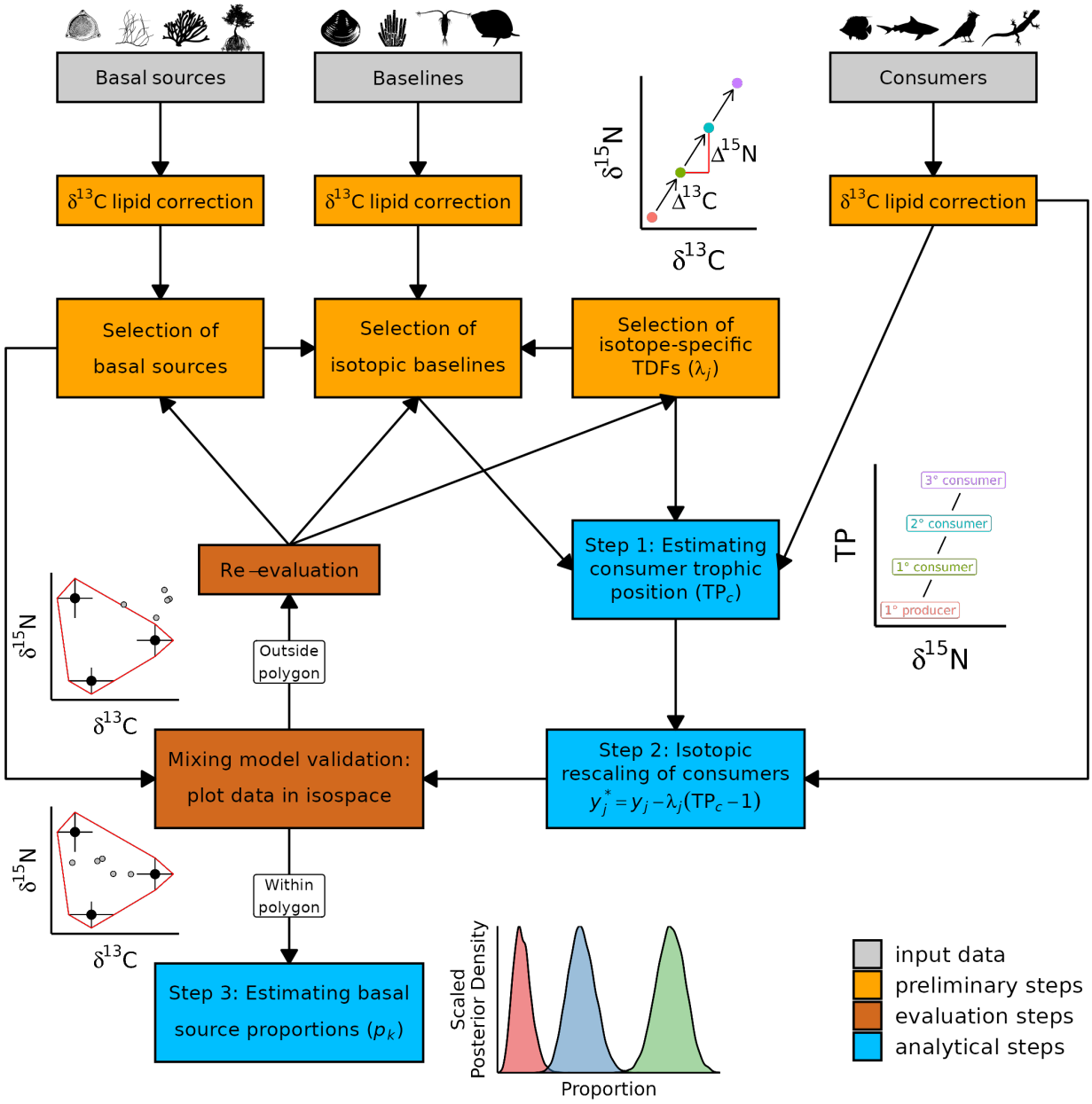

Figure 1: Flowchart for the estimation of consumer reliance on basal sources, made with the R package *ggflowchart* (Rennie, 2023). The depicted workflow is valid for both assumptions of constant and narrowing discrimination across trophic levels (see Appendix S2). Silhouettes from https://www.phylopic.org/ and contributed by Emily Jane McTavish, Rachel Wooliver, Guillaume Dera, Pablo Amador Crespo, Marina Vingiani, Alexandra Hahn, perevolotsky, John Hlavin, Kai Caspar, and Jose Carlos Arenas-Monroy.

2.2 Simulations

To evaluate the performance of the proposed workflow across different ecological scenarios, we conducted simulations using food webs where the "true" consumer trophic position (TP_c) and basal source proportions (p_k) were known. Specifically, we used these simulated food webs to: (1) test the accuracy of the workflow in estimating the proportional contribution of basal sources using one or two isotopic baselines for the estimation of consumer trophic position, and (2) investigate how consumer and source sample sizes influence the precision, uncertainty, and accuracy of the estimated proportions. These simulations allow for a quantitative assessment of the workflow's sensitivity to baseline configuration and sampling effort.

2.2.1 Simulation design and data generation

We carried out all simulations for a two-isotope system (δ^{13}C and δ^{15}N) consisting of three basal sources, one or two isotopic baselines, and one consumer. The data generation process described below was implemented using the custom R function *simulate_food_web()* (available in Appendix S3).

For each isotope, source standard deviations were generated from a uniform distribution, $U(0.5, 1.5)$. Source means were randomly sampled without replacement from a vector of six equally spaced values—twice the number of sources—between -10 and 10 (i.e., -10, -6, -2, 2, 6, 10), avoiding collinearity between isotopes (Schwarcz, 1991). This sampling strategy ensures that sources have isotopically distinct signatures (Gannes et al., 1998), which cannot be assumed when sampling from a probability distribution as in previous simulation studies (Semmens et al., 2009; Stock & Semmens, 2016).

Source proportions for baseline(s) and consumer were generated from Dirichlet distributions. A flat $Dirichlet(1, 1, 1)$ was used for the consumer. For the one-baseline simulations, a symmetric $Dirichlet(10, 10, 10)$ was used to represent the common practice of using the mean isotopic signature of all basal sources as a baseline (e.g., Ferger et al., 2013; Garcia et al., 2018; Moustaka et al., 2024). For the two-baseline simulations, we used asymmetric Dirichlet distributions to ensure that the baselines represented two distinct trophic pathways (see Appendix S4: Figure S1). To satisfy the mixing models assumption that all sources con-

tribute to the consumer (i.e., all sources included will have an estimated relative contribution exceeding zero, even if some do not significantly contribute to consumer nutrition), we constrained all source proportions to a minimum of 5% following Stock & Semmens (2016). This was achieved with a linear transformation commonly used in Beta and Dirichlet regressions (Douma & Weedon, 2019; Maier, 2014; Smithson & Verkuilen, 2006):

$$p^* = \frac{p(n-1) + \frac{1}{K}}{n}, \tag{13}$$

where p is the source proportion, K is the number of sources, and $n = \frac{1/K}{0.05} = 6.67$ for three sources.

Baseline and consumer isotopic data were generated following Stock & Semmens (2016). In practice, for each individual baseline and consumer, the number of samples taken from each source was simulated using a multinomial distribution defined by the simulated proportions and a total number of source items eaten (i.e., total consumption). The normal distribution of each source, defined by its mean and standard deviation, was repeatedly sampled based on these sample sizes. Isotopic values were calculated as the means of all source samples plus a residual error from $N(0,\, 1)$.

To reflect the reduction in isotopic variance at higher trophic levels due to time averaging (O'Reilly et al., 2002), we set a higher consumption for consumers than for baselines (40 vs. 20). These values serve as mathematical control for the variance of the simulated isotopic data and have no direct biological scaling (Stock & Semmens, 2016).

Finally, the total trophic discrimination, $\lambda_j(\mathrm{TP}_c - 1)$, was added to the generated isotopic values of consumers and baselines. Carbon and nitrogen TDFs (λ_j) were taken from Post (2002). Baseline trophic position was fixed at 2 (i.e., primary consumers), so their total discrimination equals λ_j. For consumers, a population-level mean TP_c was sampled from a truncated normal distribution, $\mathit{TNormal}(\mu,\, \sigma,\, 2,\, 5)$. By setting the truncation bounds at 2 and 5, the population mean was strictly constrained within this realistic ecological range. The hyper-parameters of this distribution, $\mu \sim U(2.5, 4.5)$ and $\sigma \sim U(0.1, 0.3)$, represent the trophic distribution of a higher taxonomic rank (e.g., family). Individual-level TP_c values were then generated by sampling from this population mean with an additional standard

deviation of 0.05 to account for intra-population variability.

2.2.2 Modelling and software settings

We used the R packages *tRophicPosition* (Quezada-Romegialli et al., 2018) for TP_c estimation and *MixSIAR* (Stock et al., 2018) for estimating basal source proportions. tRophicPosition models were run for four chains, each with 4,000 iterations and a 2,000-iteration burn-in. MixSIAR models were run for three parallel chains—utilising modified source code to allow chain parallelisation—each with 100,000 iterations, a 50,000-iteration burn-in and a thinning rate of 50. Unless otherwise specified, all MixSIAR models utilised a multiplicative error structure (Residual × Process error) to integrate both unexplained consumer isotopic variability (residual error) and variance due to the sampling process (process error) (Stock & Semmens, 2016). We used default, weakly informative priors for all simulations and checked model convergence using Gelman-Rubin statistics (Gelman & Rubin, 1992).

2.2.3 One vs two baselines

To compare the performance of the workflow under different baseline configurations, we conducted simulations using 100 "true" food web structures (Appendix S4: Figure S2). For each simulation, the workflow was applied twice to the same underlying data: once using a one-baseline configuration and once using a two-baseline configuration. This paired design ensured that the food webs were identical in terms of basal sources and consumer data, such that any differences in the resulting estimates could be attributed specifically to the baseline configuration. All food web components comprised of 10 observations ("samples").

Workflow performance across the resulting estimations (n = 200) was evaluated based on two metrics: 1) the maximum absolute error between estimated and known true source proportions (bound between 0 and 1), and 2) the coverage of the true source proportions within the 95% credible interval (CI). We then fitted a Bayesian distributional multilevel beta regression using the R package *brms* (Bürkner, 2017) to test for differences in both the mean and precision of the maximum absolute error between the one-baseline and two-baseline approaches. Furthermore, we used this model to investigate the extent to which the absolute error in TP_c estimation propagated through the workflow to affect the final source

proportion estimates (see Appendix S4: Section 4.1 for model details).

2.2.4 Sensitivity to consumer sample size

To investigate the workflow's sensitivity to consumer sampling effort, we used a single, fixed simulated food web with a two-baseline configuration (Appendix S4: Figure S3) as the reference scenario for all tests. Holding the source sample sizes constant at n = 10, we simulated 100 datasets for each of five consumer sample sizes (n = 1, 3, 5, 10, 20). This approach ensured that the underlying food web structure remained identical, allowing us to isolate the effect of sampling effort on the workflow's performance.

For cases where n = 1, a process-only error structure was required because residual variance cannot be estimated from a single data point (Stock & Semmens, 2016); for all other sample sizes, the standard multiplicative error structure was maintained. We evaluated the effect of consumer sample size on two metrics: (1) the median and (2) the 95% CI width of the estimated source proportions. To model these relationships, we fitted two Bayesian distributional non-linear power models using the R package *brms* (Bürkner, 2017) (see Appendix S4: Section 4.2 for model details). While n = 1 cases were excluded from the power models due to the differing error structure, they were retained in the overall analysis to provide a heuristic indication of bias and uncertainty in data-poor scenarios.

2.2.5 Sensitivity to source sample size

An identical approach was used to assess sensitivity to source sampling effort. Using the same fixed, two-baseline food web (Appendix S4: Figure S3), we simulated 100 datasets for each of five source sample sizes (n = 1, 3, 5, 10, 20), while holding the consumer sample size constant at n = 10. As with the consumer sample size simulations, we evaluated the effect of source sample size on the median and 95% CI width of the estimated source proportions using Bayesian distributional non-linear power models (Appendix S4: Section 4.2). All n = 1 cases were excluded from the power models because source variance cannot be estimated from a single value. These cases were nonetheless evaluated to determine the extent of bias and uncertainty introduced when only a single representative value is available per source—a limitation that can occur in field-based studies due to logistical constraints or when analysing

existing datasets where only a single mean or sample is available.

2.3 Case study: Mo'orea coral reef fishes

To demonstrate the empirical application and robustness of our three-step analytical workflow, we applied it to an extensive stable isotope database of coral reef fishes and their basal food web components from Mo'orea, French Polynesia. The dataset contains stable isotope compositions (δ^{13}C and δ^{15}N) for a total of 1,375 individual fishes spanning 194 species and 45 families collected across lagoon and reef slope habitats between 2016 and 2019. To characterise the base of the underlying food web, representative samples of major macroalgal groups, benthic cyanobacteria, oceanic particulate organic matter (POM), and invertebrate primary consumers (serving as isotopic baselines) were collected from the same habitats in 2016. Initial Bayesian linear regressions confirmed that while minor interannual isotopic variations occurred in some species, the direction of these variations was inconsistent across taxa, justifying the aggregation and analysis of the entire multi-year dataset. A detailed description of sample collection, laboratory processing, and statistical analysis is provided in Appendix S5: Section 1.

Collections were performed using permit number 681/MCE/ENV delivered by the Government of French Polynesia.

The analysis followed all sequential steps of the proposed workflow:

1. *Data preparation and parameter selection.* To eliminate lipid-derived bias, δ^{13}C data were mathematically normalised based on individual C:N ratios, applying taxon-specific models for fish and invertebrate samples. For the basal sources, five initial categories (benthic cyanobacteria, green, brown, and red algae, and oceanic POM) were evaluated in isospace. Brown and green algae exhibited substantial isotopic overlap and were grouped a priori, reducing the system to four distinct basal sources. To evaluate workflow sensitivity to parameter selection, we selected two distinct sets of widely used literature TDFs: general, non-tissue-specific values from Post (2002) and muscle tissue-specific values from McCutchan Jr et al. (2003). Finally, two long-lived primary consumers—benthic grazing gastropods (*Monetaria*

spp.) and filter-feeding bivalves (*Limaria fragilis*)—were selected as isotopic baselines. Their suitability was verified via a Bayesian SIMM fitted with *MixSIAR* (Stock et al., 2018), which confirmed their high fidelity to their assumed pathways (substantial pelagic reliance for bivalves and predominant benthic reliance for gastropods).

2. *Steps 1 and 2: Trophic position and rescaling.* To estimate trophic position for each fish species, we fitted Bayesian two-baseline models using *tRophicPosition* (Quezada-Romegialli et al., 2018) using the validated baselines and selected TDFs. Then, the raw isotopic signatures of individual fish were rescaled to the food web base using the estimated trophic position modes and mean TDF values, removing the accumulated trophic discrimination. Because these steps were executed independently for non-tissue-specific and muscle tissue-specific TDFs, two separate rescaled fish datasets were generated for all subsequent steps.

3. *Evaluation step.* Both rescaled fish datasets were plotted in bi-dimensional isospace to determine whether consumers correctly fell within the mixing polygon. While the vast majority of individuals fell within the polygon boundaries, isolated individuals falling outside were evaluated using the probabilistic framework of Smith et al. (2013). Observations with an explanatory probability <1% (11 individuals for non-tissue-specific and 23 for muscle tissue-specific TDFs, representing 0.8% and 1.7% of the total observations, respectively) were flagged and removed to prevent model convergence failures. Furthermore, the calculated normalised surface area of the mixing polygon was 35.5 SD^2, confirming a well-resolved source geometry (Brett, 2014).

4. *Step 3: Basal source proportions.* To accommodate their complex ecological and evolutionary structure, the rescaled consumer datasets were partitioned into eight distinct trophic guilds capturing the primary energetic pathways of the reef (Appendix S6). Within each guild, we fitted alternative Bayesian SIMMs using a process-only error structure with *MixSIAR* (Stock et al., 2018) to evaluate the relative effects of body size and evolutionary history. For each guild, a suite of four models of descending complexity was executed: (1) a full model incorporating taxonomy as a nested random effect (species within family) and fish total length as a continuous covariate;

(2) a taxonomy-only model; (3) a length-only model; and (4) a null model. Models were fit using informative Dirichlet priors constructed from expert system knowledge (Parravicini et al., 2020), maintaining a hyperparameter sum equal to the number of sources ($K = 4$). Final model selection was performed through leave-one-out cross validation (Vehtari et al., 2017) and stacking weights (Yao et al., 2018). Estimated proportions for all algal sources (green-brown combined and red algae) were aggregated *a posteriori* to yield an overarching algal contribution. Finally, species-specific median source proportions estimated with non-tissue-specific and muscle tissue-specific TDFs were directly compared using Bayesian multivariate linear regressions fitted with *brms* (Bürkner, 2017) to provide a comprehensive check on the sensitivity of our ecological conclusions.

2.4 Computational environment

Simulations and analyses were performed in R version 4.4.2 (R Core Team, 2024), using *targets* version 1.6.0 (Landau, 2021) as pipeline tool to manage the project workflow, *renv* version 1.0.5 (Ushey & Wickham, 2024) for package management and *crew* version 0.10.2 (Landau, 2024) for parallel computing. Bayesian SIMMs were fitted with *tRophicPosition* version 0.8.0 (Quezada-Romegialli et al., 2018) and *MixSIAR* latest development version (commit 037978f) (Stock et al., 2018) in JAGS (Just Another Gibbs Sampler) version 4.3.2 (Plummer, 2003), while Bayesian regression models were fitted with *brms* version 2.21.0 (Bürkner, 2017) using CmdStan version 2.35.0 (Gabry et al., 2024; Stan Development Team, 2024) as backend.

3 Results

3.1 Simulations

3.1.1 One vs two baselines

The workflow performed similarly irrespective of the baseline configuration. The 95% CI of estimated source proportions included 59% and 54% of the true source proportions (n =

300, 100 food webs $\times$ 3 sources) for the one-baseline and the two-baseline configurations, respectively. Maximum absolute errors in estimated source proportions ranged between 0.01 and 0.63, although 84% and 85% of all estimated source proportions had an absolute error of <0.2 for the one-baseline and the two-baseline configurations, respectively (Figure 2A and Appendix S4: Figure S5A). The maximum absolute error did not differ in either the mean ($P_{One\text{-}baseline>Two\text{-}baselines} = 0.16$) or precision ($P_{One\text{-}baseline>Two\text{-}baselines} = 0.13$) between the two approaches (Appendix S4: Figure S5B). Furthermore, the maximum absolute error in estimated source proportions was strongly, positively related to the absolute error in estimated consumer trophic position (TP_c, Figure 2A), with no difference in the slope between the two approaches ($\beta = 0.74[0.65, 0.86]$ for one baseline and $\beta = 0.77[0.69, 0.86]$ for two baselines, respectively, with $P_{One\text{-}baseline>Two\text{-}baselines} = 0.33$). As demonstrated in Appendix S7, the error in estimated TP_c is proportional to the difference in TP-corrected $\delta^{15}N$ (i.e., at TP = 1) between the consumer and the baseline(s) (Figure 2B), which results from different contributions of the sources to the consumer and baseline(s).

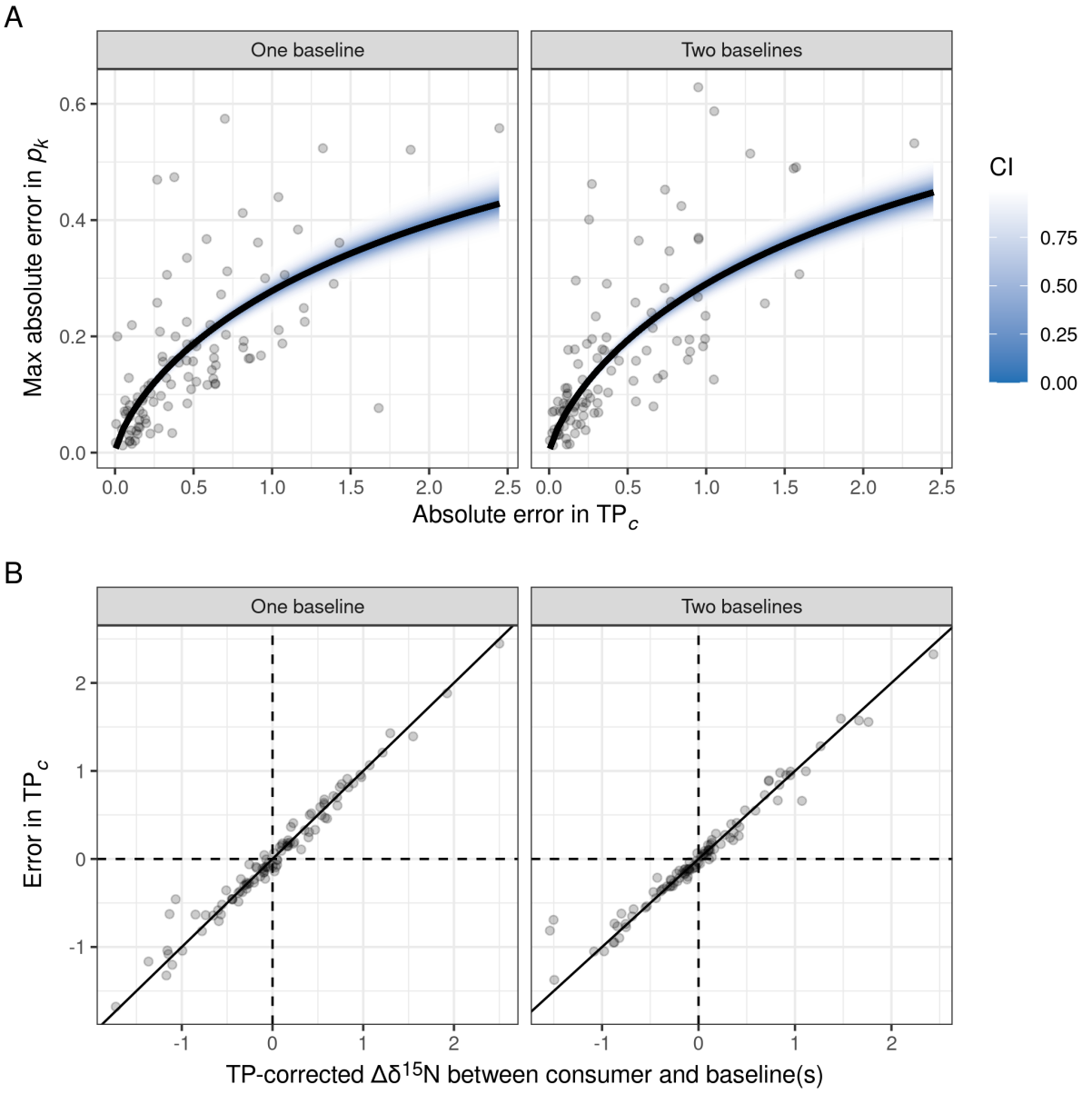

Figure 2: A) Relationship between the maximum absolute error in estimated source proportions (p_k) and the absolute error in estimated consumer trophic position (TP_c) using one or two isotopic baselines. Semi-transparent points are raw data, while solid lines and ribbons are model predictions that represent the median and credible interval (CI), respectively. B) Relationship between the error in estimated TP_c and the TP-corrected difference in $\delta^{15}N$ between the consumer and the baseline(s) for the one-baseline and two-baseline approaches. See appendix S7 for mathematical derivation. Semi-transparent points are raw data, while solid lines represent a 1:1 relationship.

3.1.2 Sensitivity to consumer sample size

In this food web (Appendix S4: Figure S3), the true source contributions to the simulated consumer were relatively well recovered with small absolute errors (<0.1) in the estimates (Figure 3A). Consumer sample size affected both the estimated median and the 95% CI width of source proportions (Figure 3A-B). However, it only had a weak effect on the mean of the estimated medians (i.e., accuracy), while the mean of the estimated 95% CI widths (i.e., uncertainty) strongly decreased with increasing sample size (Figure 3A-B and C-D, panel μ). In addition, the variance of both estimates (i.e., precision) decreased with increasing sample size (Figure 3C-D, panel σ). Note that the slightly different point estimates obtained with a consumer sample size of one compared to larger sample sizes are likely due to different error structures in the mixing model (i.e., Process-only vs Residual $\times$ Process).

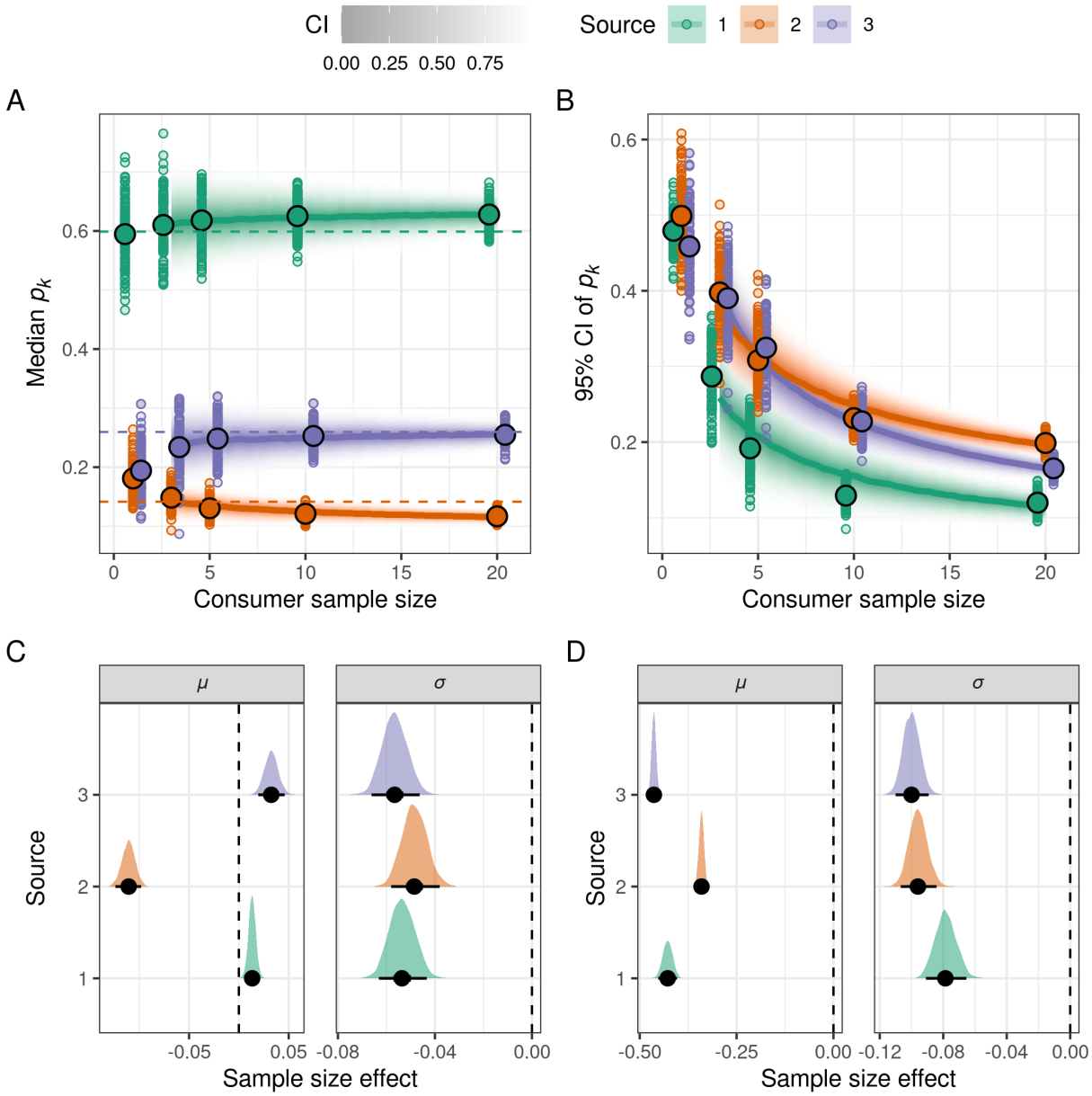

Figure 3: A-B) Relationships between the median and 95% credible interval (CI) width of source proportions (p_k) and consumer sample size. Small, coloured points are raw data, while large points represent the means. These are horizontally adjusted for clarity. Solid lines and ribbons are model predictions and represent the median and CI, respectively. Coloured dashed lines in A represent true p_k. C-D) Effect of consumer sample size on the mean (μ) and standard deviation (σ) of the median and 95% CI width of p_k, respectively. Black dots and intervals represent the median, 50 and 95% CI, respectively, while the coloured area represents the posterior density. Some intervals are too small to be seen. Vertical dashed lines depict 0 (i.e., no effect).

3.1.3 Sensitivity to source sample size

Source sample size did not affect the mean of the estimated median source proportions (i.e., accuracy) (Figure 4A and C, panel μ), but a larger source sample size reduced the mean of the estimated 95% CI width of source proportions (Figure 4B and D, panel μ), although to a smaller extent compared to the consumer sample size. Furthermore, a larger source sample size reduced the variance of both estimates (i.e., precision) (Figure 4C-D, panel σ).

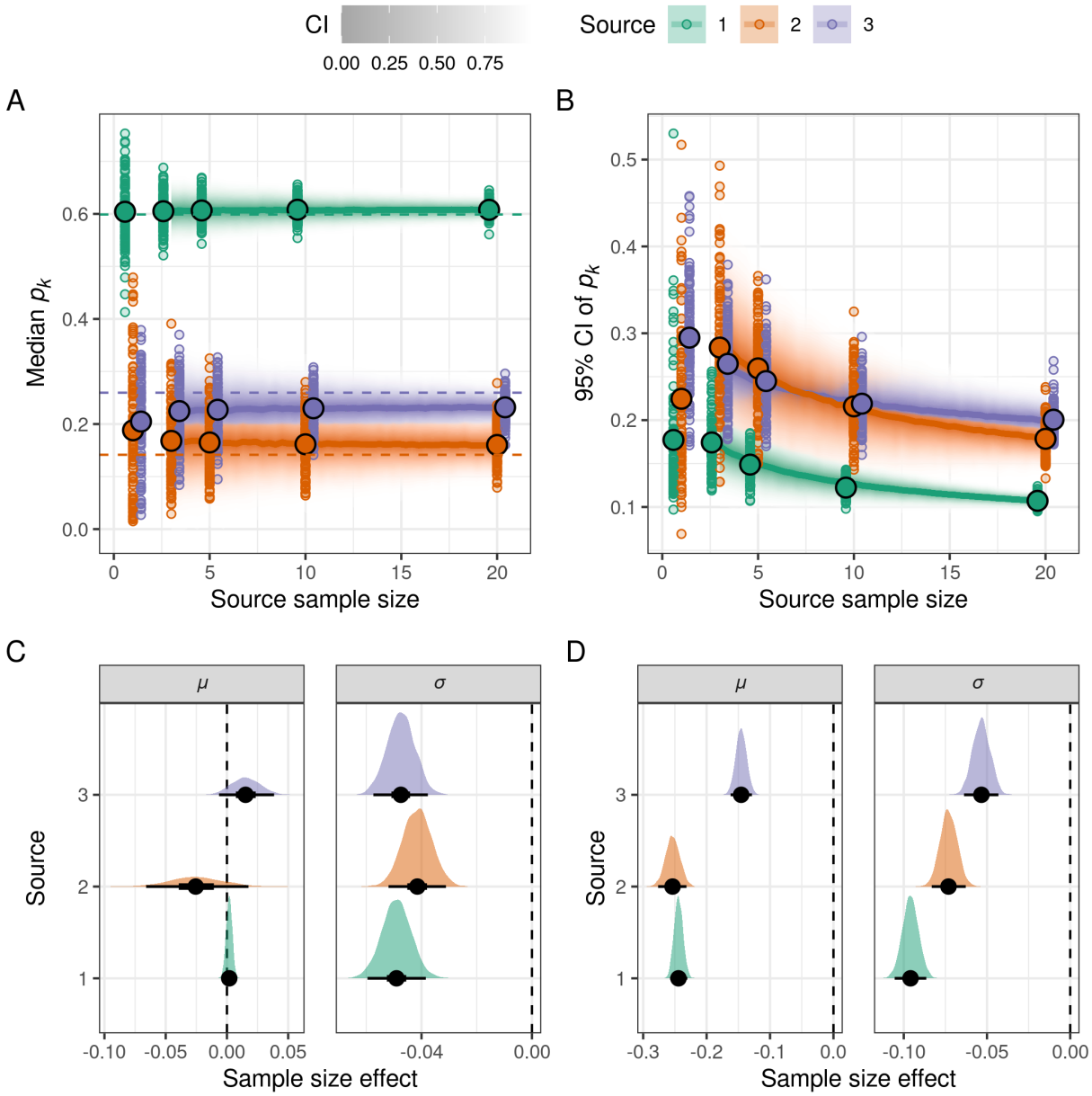

Figure 4: A-B) Relationships between the median and 95% credible interval (CI) width of source proportions (p_k) and source sample size. Small, coloured points are raw data, while large points represent the means. These are horizontally adjusted for clarity. Solid lines and ribbons are model predictions and represent the median and CI, respectively. Coloured dashed lines in A represent true p_k. C-D) Effect of source sample size on the mean (μ) and standard deviation (σ) of the median and 95% CI width of p_k, respectively. Black dots and intervals represent the median, 50 and 95% CI, respectively, while the coloured area represents the posterior density. Some intervals are too small to be seen. Vertical dashed lines depict 0 (i.e., no effect).

3.2 Case study

The selected models consistently included the full structure (including taxonomy and fish total length) or only taxonomy (Appendix S5: Tables S5-S6). The estimated relative contribution of basal sources to coral reef fishes in Mo'orea were strongly correlated between the non-tissue-specific and muscle tissue-specific TDFs, with POM and algae (brown, green, and red algae combined) providing the largest overall contributions to similar extents and cyanobacteria providing the smallest contributions, regardless of the TDFs used (Appendix S5: Figure S18A). However, the contribution of cyanobacteria was consistently higher using non-tissue-specific TDFs than muscle tissue-specific ones, while the contributions of POM were lower and those of algae showed no significant difference (Appendix S5: Figure S18B). These differences were more pronounced for some trophic guilds than others (Appendix S5: Figure S18C). Furthermore, the probability of fish falling within the mixing polygon was similar when isotopic signatures were rescaled using non-tissue-specific or muscle tissue-specific TDFs (Appendix S5: Figure S19), although fewer observations fell outside the mixing polygon using non-tissue-specific TDFs than muscle tissue-specific ones (Appendix S5: Figures S13-S14). Therefore, both TDFs appear appropriate for our study. We present the results obtained with non-tissue-specific TDFs here and report those obtained with muscle tissue-specific TDFs in Appendix S5.

On average, algae (brown, green, and red algae combined) were the dominant source for herbivores and detritivores, benthic and sessile invertivores, and macrocarnivores; however, for sessile invertivores and macrocarnivores, the contribution of algae was only slightly higher than that of POM (Figure 5). POM, however, was by far the dominant source for omnivores and planktivores, contributing over 50% and highlighting their high reliance on pelagic primary production. Cyanobacteria, despite having an overarching low contribution, were the dominant source for microphages, indicating a high reliance on benthic microbial primary production. These results are consistent with the known trophic ecology of these fishes (Clements et al., 2017; McMahon et al., 2016; Parravicini et al., 2020). However, the relative contribution of basal sources varied with fish body size for all trophic guilds except benthic omnivores (Appendix S5: Figure S20), as well as across species and families for

all trophic guilds (Appendix S5: Figures S21-S28). Taxonomic variation was more marked for some guilds (e.g., microphages) and less for others (e.g., planktivores). All results were consistent using muscle tissue-specific TDFs (Appendix S5: Figures S29-S38), although the estimated proportions differed in absolute values.

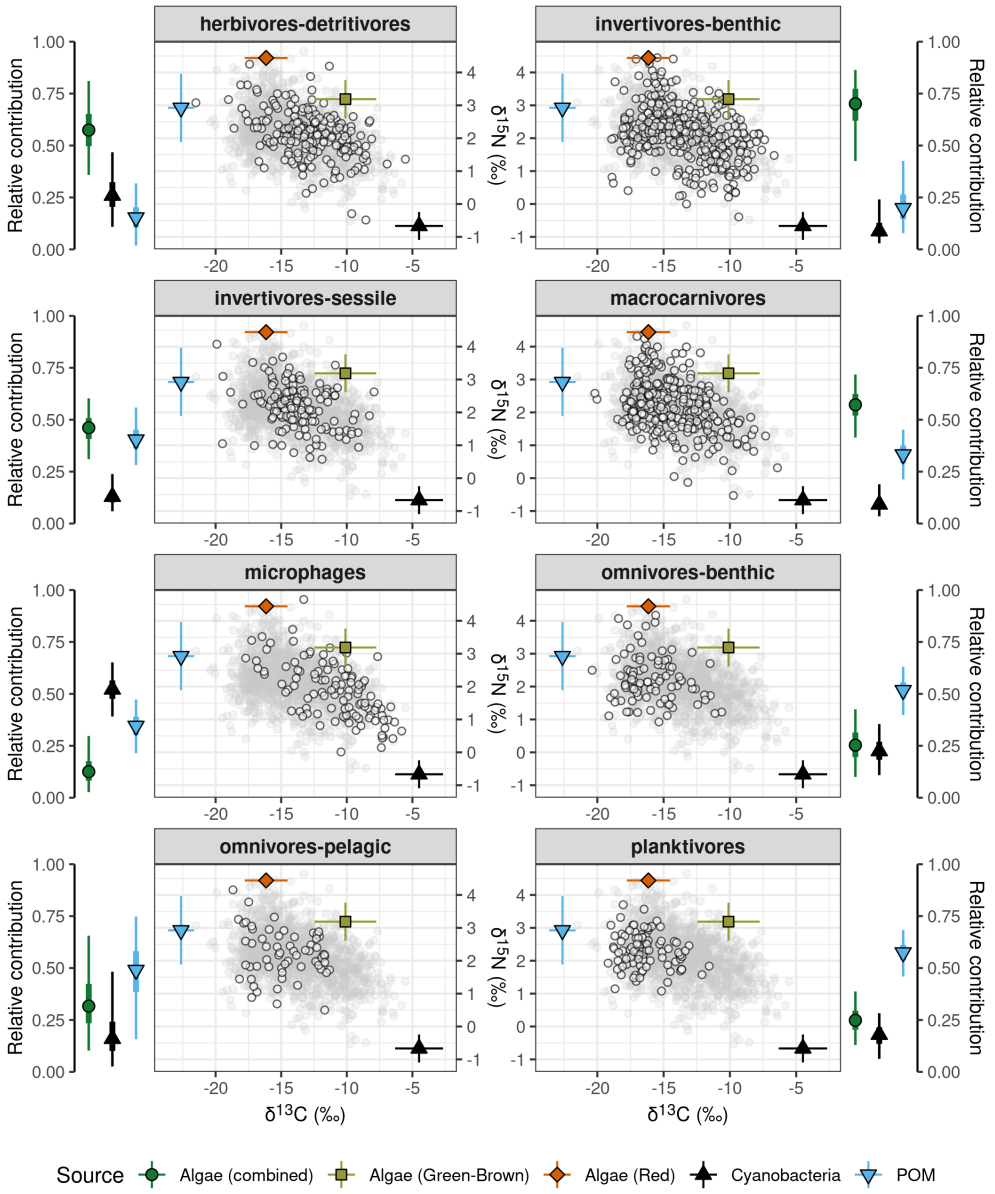

Figure 5: Isospace of reef fish species from Mo'orea, French Polynesia, and relative contribution of basal sources to each trophic guild. Within the isospace, grey, semitransparent dots represent individual fish isotopic signatures rescaled using non-tissue-specific trophic discrimination factors (Post, 2002), white dots represent observations belonging to each trophic guild, while coloured dots and lines represent means (± 1SD) isotopic signatures of basal sources. Dots and intervals in side plots represent the median, 50 and 95% credible intervals, respectively, of the estimated relative contribution of basal sources to each trophic guild. Note that the contributions of green-brown algae and red algae were aggregated *a posteriori* to provide an overarching algal contribution. POM = particulate organic matter.

4 Discussion

We present a comprehensive description of the analytical workflow for estimating consumer reliance on basal food sources using bulk SIA. To our knowledge, existing guidance on this approach is scarce and fragmented. In particular, only brief mentiones exist about correcting consumer isotopic values for total trophic discrimination by multiplying the TDF of a single trophic step by the number of trophic steps between the consumer and basal sources (Phillips et al., 2014). Here we make this rationale explicit by describing each step required to derive consumer reliance on basal sources from the bulk stable isotope data. We also test how relaxing some sampling and parameter assumptions influence the results, providing a clear mathematical and empirical foundation for best practices in the analysis of bulk isotope data.

We show that the method used to estimate consumer trophic position influences the results (Letourneur et al., 2024), but not to the extent that we anticipated. In particular, we focus on using existing models (Post, 2002) that require a single isotopic baseline, typically using the average δ^{15}N of basal sources (e.g., Ferger et al., 2013; Garcia et al., 2018; Moustaka et al., 2024) to represent an average trophic pathway, or two isotopic baselines representing two distinct trophic pathways (e.g., McMahon et al., 2021). Our simulations show that estimates of source-specific reliance for consumers are largely robust to the choice between these two approaches. However, different outcomes could occur on a case-by-case basis, particularly depending on the geometry of the mixing polygon and the position of the consumers in isospace. Indeed, using the average δ^{15}N of the basal sources as the sole isotopic baseline will only yield an accurate estimate of trophic position for consumers that uniformly rely on basal sources or primarily on a basal source with an average δ^{15}N value. However, as suggested by Equation S10 in Appendix S7, it will overestimate the trophic position of consumers that primarily rely on a basal source with a high δ^{15}N value and underestimate the trophic position of consumers that primarily rely on a basal source with a low δ^{15}N value. Furthermore, this error will be greater as the range in δ^{15}N of the sources increases. On the other hand, using two isotopic baselines will yield an accurate estimate of trophic position for consumers that primarily rely on one or more basal sources represented

by these baselines, but it may yield an inaccurate estimate for consumers that primarily rely on a basal source not represented by either of the baselines.

Nevertheless, our simulations highlight the high sensitivity of the estimated basal source proportions to error in estimating consumer trophic position, indicating that trophic position estimation is a critical step in the workflow. Therefore, while we cannot directly recommend whether to use one or two baselines, we provide an R function (*estimate_tp_error()*, Appendix S8) that implements the equations derived in Appendix S7, which ecologists can use to assess bias in estimating trophic position for hypothetical consumers, given the mean isotopic signature of the sources and baselines. We recommend publishing the expected bias to allow the reader to independently assess the accuracy of the results.

Applying this function to our case study, we show that the absolute bias in estimating consumer trophic position is consistently below 0.5 for a set of consumers that primarily rely on one, two, or three basal sources, or equally on all four basal sources (Appendix S5: Tables S7-S8). Furthermore, according to our model (Figure 2), the expected absolute error in the estimated source proportions is consistently less than 0.2, supporting our results.

Our simulations also show that consumer and source sample sizes only have a minimal effect on the accuracy of estimated proportions, but larger sample sizes increase the precision of estimates and reduce their uncertainty. Whenever possible, a minimum sample size of ten for each consumer and basal source is recommended. However, under financial constraints or when using pre-existing data, smaller sample sizes may still provide accurate results at the expense of greater uncertainty, while broadening the study scope by allowing for the analysis of a greater number of taxa, locations, or time points.

Finally, our case study highlights the sensitivity of the estimated basal source proportions to the choice of isotope-specific TDFs. It is therefore essential to perform a sensitivity analysis using alternative TDFs to increase the robustness of the results and support TDF choices.

To facilitate the application of the proposed analytical workflow, we also provide a step-by-step guide in R using a simulated food web in a determined system characterised by three basal sources and two isotopes (Appendix S9). In this vignette, we demonstrate that accurate results can be obtained for 20 simulated species belonging to five families, even with sample sizes as small as one.

4.1 Limitations and future directions

Our proposed analytical workflow has the following limitations:

1. Currently available software packages only allow estimation of consumer trophic position using one or two isotopic baselines, which results in a quantifiable error in trophic position estimates when the studied system has more than two basal sources (Appendix S7). This is typical in complex ecosystems and affects estimates of basal source proportions (Figure 2).

2. Uncertainty in TDFs and trophic position cannot be propagated into rescaled consumer values, as current Bayesian SIMMs do not account for measurement error in individual-level consumer isotopic signatures. This affects the estimated uncertainty in basal source proportions. An alternative approach involves fitting SIMMs according to Equation 5, thus using consumer data at the original scale and incorporating total discrimination and its uncertainty into the model. However, consumer taxa exhibit distinct total discrimination depending on their trophic position. Therefore, a separate SIMM needs to be run for each consumer (e.g., Zapata-Hernández et al., 2021) since it is not currently possible to specify TDFs per factor. While this approach may be preferable in studies involving one or a few taxa, it becomes unrealistic across a larger number of taxa, as in our case study. It can also lead to inefficient inferences with greater uncertainties in source proportions (see Stock et al., 2018), and it limits the ability to obtain large-scale information (e.g., at the level of trophic guilds or higher taxonomic ranks).

3. Compounding these structural and software constraints, our framework relies on the assumption of constant trophic discrimination. This is a notable limitation, as it is widely recognised that TDFs vary substantially across taxa, diet and tissue types, and trophic groups (Canseco et al., 2022; Caut et al., 2009; Stephens et al., 2023). While alternative frameworks, such as the "narrowing discrimination" approach (Hussey et al., 2014), offer a compelling way to model variable TDFs across multiple trophic levels, their application remains severely constrained by data availability and software

limitations. Currently, the only available parameters for narrowing discrimination frameworks are restricted to nitrogen isotopes and fish muscle tissue from the foundational study that proposed the framework (Hussey et al., 2014). To address this limitation and facilitate the adoption of this approach, we have mathematically integrated the narrowing discrimination framework into our proposed workflow in Appendix S2, providing the full required equations for its implementation. However, applying this advanced theoretical framework to our empirical case study was unfeasible due to the lack of derived parameter data for other stable isotopes (such as carbon), as well as the lack of existing software packages capable of fitting these complex models. Because the total trophic discrimination accumulated between basal sources and high-level consumers depends inherently on consumer trophic positions—which must be estimated first to rescale consumer isotopic signatures—the constant TDF approach remains the only viable method at present. We validated the robustness of this choice through a sensitivity analysis on the Mo'orea case study, re-running our complete workflow under alternative constant TDFs (general non-tissue-specific vs muscle tissue-specific values). The final estimated basal source proportions exhibited limited sensitivity to these parameter changes, demonstrating that evaluating model sensitivity via alternative constant TDFs remains a critical, reliable safeguard until software capabilities and empirical meta-analyses expand.

To address these issues, SIMMs that can explicitly handle multiple trophic steps are desirable (Heikkinen et al., 2022). However, these can only be applied to simple, well-known food webs. For more complex food webs, which constitute the majority of cases, new models based on Equation 2 are needed. These would extend existing models (Equations 6-11), allowing for the estimation of trophic position along with basal source proportions in situations with more than two basal sources and, possibly, modelling both trophic position and source proportions as a function of covariates. However, to avoid underdetermined systems, these models require the same number of stable isotopes as sources (i.e., a third isotope would be needed for a system with three sources).

Beyond providing a robust analytical framework, our case study demonstrates how this integrated workflow can be applied to large, complex datasets to extract ecological insights

at multiple taxonomic and functional levels. By bridging existing methodological gaps, this framework unlocks the potential to leverage vast, underutilised repositories of stable isotope data across the published literature (e.g., Berto et al., 2025; Boulêtreau et al., 2025; Shipley et al., 2024), enabling researchers to generate novel meta-analyses and synthesise energy pathways at broad macroecological scales. Furthermore, these findings could be scaled directly to the community level to quantify energy and nutrient fluxes across environmental gradients, map functional biogeography, and predict how ecosystem functioning will respond to changes in primary production sources. Ultimately, this workflow offers a scalable and accessible standard that enhances the comparative power of stable isotope ecology across diverse aquatic and terrestrial biomes.

Author Contributions

Mattia Ghilardi, Renato A. Morais and Valeriano Parravicini conceived and designed the study. Simon J. Brandl, Jordan M. Casey, Alexandre Merciere, Fabien Morat, Nina M. D. Schiettekatte and Valeriano Parravicini collected the data for the case study. Mattia Ghilardi developed the workflow and simulations, performed the analysis and wrote the first draft of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Acknowledgements

We thank Samuel Degregori, Titouan Roncin, Gabrielle Martineau, Kailey Bissell, Calvin Quigley, Tommy Norin, Jérémy Carlot, and Pierrick Harnay for assistance with field collections and fish dissections. We thank Jason Vii for support with stable isotope preparation. This work was funded by LabEx "Corail" (projects Nectar and METAREEF; VP, MK, YL), the BNP Paribas Foundation (Reef Services Project; VP), and a Make Our Planet Great Again Postdoctoral Grant (mopga-pdf-0000000144; JMC).

Data and Code Availability statement

Code to reproduce results and figures is available on GitHub for both the simulations (https://github.com/mattiaghilardi/SIMMmultitroph_simulations) and the case study (https://github.com/mattiaghilardi/SIMMmultitroph_case_study), along with the Mo'orea stable isotope dataset.

References

Berto, D., Fanelli, E., Vizzini, S., Rampazzo, F., Da Ros, Z., Noventa, S., Fortibuoni, T., Antonini, C., Cilluffo, G., Signa, G., Premici, A., Bardelli, R., & Raicevich, S. (2025). ISOMED - a stable ISOtope database of MEDiterranean marine food web components. *Scientific Data*, *12*(1), 1768. https://doi.org/10.1038/s41597-025-05981-y

Bird, C. S., Veríssimo, A., Magozzi, S., Abrantes, K. G., Aguilar, A., Al-Reasi, H., Barnett, A., Bethea, D. M., Biais, G., Borrell, A., Bouchoucha, M., Boyle, M., Brooks, E. J., Brunnschweiler, J., Bustamante, P., Carlisle, A., Catarino, D., Caut, S., Cherel, Y., … Trueman, C. N. (2018). A global perspective on the trophic geography of sharks. *Nature Ecology & Evolution*, *2*(2), 299–305. https://doi.org/10.1038/s41559-017-0432-z

Boulêtreau, S., Vagnon, C., Comte, L., Sagouis, A., Pool, T., Stiling, R. R., Harrod, C., South, J., McIntosh, A. R., Perga, M.-E., Sánchez-Hernández, J., Roussel, J.-M., Tunney, T. D., Jackson, M., Olden, J. D., & Cucherousset, J. (2025). IsoFresh: A global stable isotope database of freshwater food webs. *Knowledge & Management of Aquatic Ecosystems*, (426), 15. https://doi.org/10.1051/kmae/2025010

Brett, M. T. (2014). Resource polygon geometry predicts bayesian stable isotope mixing model bias. *Marine Ecology Progress Series*, *514*, 1–12. https://doi.org/10.3354/meps11017

Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Cabana, G., & Rasmussen, J. B. (1996). Comparison of aquatic food chains using nitrogen isotopes. *Proceedings of the National Academy of Sciences*, *93*(20), 10844–10847. https:

679 //doi.org/10.1073/pnas.93.20.10844
680 Canseco, J. A., Niklitschek, E. J., & Harrod, C. (2022). Variability in δ^{13}C and δ^{15}N
681 trophic discrimination factors for teleost fishes: A meta-analysis of temperature and
682 dietary effects. *Reviews in Fish Biology and Fisheries*, *32*(2), 313–329. https://doi.org/
683 10.1007/s11160-021-09689-1
684 Caut, S., Angulo, E., & Courchamp, F. (2009). Variation in discrimination factors (Δ^{15}N and
685 Δ^{13}C): The effect of diet isotopic values and applications for diet reconstruction. *Journal*
686 *of Applied Ecology*, *46*(2), 443–453. https://doi.org/10.1111/j.1365-2664.2009.01620.x
687 Clements, K. D., German, D. P., Piché, J., Tribollet, A., & Choat, J. H. (2017). Integrat-
688 ing ecological roles and trophic diversification on coral reefs: Multiple lines of evidence
689 identify parrotfishes as microphages. *Biological Journal of the Linnean Society*, *120*(4),
690 729–751. https://doi.org/10.1111/bij.12914
691 DeNiro, M. J., & Epstein, S. (1977). Mechanism of carbon isotope fractionation associ-
692 ated with lipid synthesis. *Science*, *197*(4300), 261–263. https://doi.org/10.1126/science.
693 327543
694 Douma, J. C., & Weedon, J. T. (2019). Analysing continuous proportions in ecology and
695 evolution: A practical introduction to beta and dirichlet regression. *Methods in Ecology*
696 *and Evolution*, *10*(9), 1412–1430. https://doi.org/10.1111/2041-210X.13234
697 Ferger, S. W., Böhning-Gaese, K., Wilcke, W., Oelmann, Y., & Schleuning, M. (2013).
698 Distinct carbon sources indicate strong differentiation between tropical forest and farm-
699 land bird communities. *Oecologia*, *171*(2), 473–486. https://doi.org/10.1007/s00442-012-
700 2422-9
701 Gabry, J., Češnovar, R., Johnson, A., & Bronder, S. (2024). *Cmdstanr: R interface to*
702 *'CmdStan'*. https://mc-stan.org/cmdstanr/
703 Gannes, L. Z., Martínez del Rio, C., & Koch, P. (1998). Natural abundance variations
704 in stable isotopes and their potential uses in animal physiological ecology. *Comparative*
705 *Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *119*(3), 725–737.
706 https://doi.org/10.1016/S1095-6433(98)01016-2
707 Gannes, L. Z., O'Brien, D. M., & Martínez del Rio, C. (1997). Stable isotopes in animal
708 ecology: Assumptions, caveats, and a call for more laboratory experiments. *Ecology*,

78(4), 1271–1276. https://doi.org/10.2307/2265878

Garcia, A. F. S., Santos, M. L., Garcia, A. M., & Vieira, J. P. (2018). Changes in food web structure of fish assemblages along a river-to-ocean transect of a coastal subtropical system. *Marine and Freshwater Research*, *70*(3), 402–416. https://doi.org/10.1071/MF18212

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Govan, E., Jackson, A. L., Bearhop, S., Inger, R., Stock, B. C., Semmens, B. X., Ward, E. J., & Parnell, A. C. (2024). *Cosimmr: An r package for fast fitting of stable isotope mixing models with covariates* (arXiv:2408.17230). arXiv. https://doi.org/10.48550/arXiv.2408.17230

Govan, E., Jackson, A. L., Inger, R., Bearhop, S., & Parnell, A. C. (2023). *Simmr: A package for fitting stable isotope mixing models in r* (arXiv:2306.07817). arXiv. https://doi.org/10.48550/arXiv.2306.07817

Heikkinen, R., Hämäläinen, H., Kiljunen, M., Kärkkäinen, S., Schilder, J., & Jones, R. I. (2022). A bayesian stable isotope mixing model for coping with multiple isotopes, multiple trophic steps and small sample sizes. *Methods in Ecology and Evolution*, *13*(11), 2586–2602. https://doi.org/10.1111/2041-210X.13989

Hilting, A. K., Currin, C. A., & Kosaki, R. K. (2013). Evidence for benthic primary production support of an apex predator–dominated coral reef food web. *Marine Biology*, *160*(7), 1681–1695. https://doi.org/10.1007/s00227-013-2220-x

Hussey, N. E., MacNeil, M. A., McMeans, B. C., Olin, J. A., Dudley, S. F. J., Cliff, G., Wintner, S. P., Fennessy, S. T., & Fisk, A. T. (2014). Rescaling the trophic structure of marine food webs. *Ecology Letters*, *17*(2), 239–250. https://doi.org/10.1111/ele.12226

Kiljunen, M., Grey, J., Sinisalo, T., Harrod, C., Immonen, H., & Jones, R. I. (2006). A revised model for lipid-normalizing 13C values from aquatic organisms, with implications for isotope mixing models. *Journal of Applied Ecology*, *43*(6), 1213–1222. https://doi.org/10.1111/j.1365-2664.2006.01224.x

Kjeldgaard, M. K., Hewlett, J. A., & Eubanks, M. D. (2021). Widespread variation in stable isotope trophic position estimates: Patterns, causes, and potential consequences.

Ecological Monographs, *91*(3), e01451. https://doi.org/10.1002/ecm.1451

Landau, W. M. (2021). The targets r package: A dynamic make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software*, *6*(57), 2959. https://doi.org/10.21105/joss.02959

Landau, W. M. (2024). *Crew: A distributed worker launcher framework*. https://CRAN.R-project.org/package=crew

Letourneur, Y., Fey, P., Dierking, J., Galzin, R., & Parravicini, V. (2024). Challenging trophic position assessments in complex ecosystems: Calculation method, choice of baseline, trophic enrichment factors, season and feeding guild do matter: A case study from marquesas islands coral reefs. *Ecology and Evolution*, *14*(7), e11620. https://doi.org/10.1002/ece3.11620

Lindeman, R. L. (1942). The trophic-dynamic aspect of ecology. *Ecology*, *23*(4), 399–417. https://doi.org/10.2307/1930126

Lorrain, A., Pethybridge, H., Cassar, N., Receveur, A., Allain, V., Bodin, N., Bopp, L., Choy, C. A., Duffy, L., Fry, B., Goñi, N., Graham, B. S., Hobday, A. J., Logan, J. M., Ménard, F., Menkes, C. E., Olson, R. J., Pagendam, D. E., Point, D., … Young, J. W. (2020). Trends in tuna carbon isotopes suggest global changes in pelagic phytoplankton communities. *Global Change Biology*, *26*(2), 458–470. https://doi.org/10.1111/gcb.14858

Maier, M. J. (2014). *DirichletReg: Dirichlet regression for compositional data in r* (Research Report Series / Department of Statistics and Mathematics 125). WU Vienna University of Economics; Business. https://research.wu.ac.at/ws/portalfiles/portal/17761231/Report125.pdf

Martínez del Rio, C., Wolf, N., Carleton, S. A., & Gannes, L. Z. (2009). Isotopic ecology ten years after a call for more laboratory experiments. *Biological Reviews*, *84*(1), 91–111. https://doi.org/10.1111/j.1469-185X.2008.00064.x

McConnaughey, T., & McRoy, C. P. (1979). Food-web structure and the fractionation of carbon isotopes in the bering sea. *Marine Biology*, *53*(3), 257–262. https://doi.org/10.1007/BF00952434

McCutchan Jr, J. H., Lewis Jr, W. M., Kendall, C., & McGrath, C. C. (2003). Variation in trophic shift for stable isotope ratios of carbon, nitrogen, and sulfur. *Oikos*, *102*(2),

378–390. https://doi.org/10.1034/j.1600-0706.2003.12098.x

McMahon, K. W., Ambrose, W. G., Reynolds, M. J., Johnson, B. J., Whiting, A., & Clough, L. M. (2021). Arctic lagoon and nearshore food webs: Relative contributions of terrestrial organic matter, phytoplankton, and phytobenthos vary with consumer foraging dynamics. *Estuarine, Coastal and Shelf Science*, *257*, 107388. https://doi.org/10.1016/j.ecss.2021.107388

McMahon, K. W., Thorrold, S. R., Houghton, L. A., & Berumen, M. L. (2016). Tracing carbon flow through coral reef food webs using a compound-specific stable isotope approach. *Oecologia*, *180*(3), 809–821. https://doi.org/10.1007/s00442-015-3475-3

Moustaka, M., Bassett, T. J., Beltran, L., Cuttler, M. V. W., Evans, R. D., Gorman, D., Grimaldi, C. M., Gruber, R. K., Hyndes, G. A., Kendrick, G. A., Travaglione, N., & Wilson, S. K. (2024). Suspended particulate organic matter supports mesopredatory fish across a tropical seascape. *Ecosystems*, *27*(7), 918–936. https://doi.org/10.1007/s10021-024-00929-6

Nielsen, J. M., Clare, E. L., Hayden, B., Brett, M. T., & Kratina, P. (2018). Diet tracing in ecology: Method comparison and selection. *Methods in Ecology and Evolution*, *9*(2), 278–291. https://doi.org/10.1111/2041-210X.12869

O'Reilly, C. M., Hecky, R. E., Cohen, A. S., & Plisnier, P.-D. (2002). Interpreting stable isotopes in food webs: Recognizing the role of time averaging at different trophic levels. *Limnology and Oceanography*, *47*(1), 306–309. https://doi.org/10.4319/lo.2002.47.1.0306

Osada, Y., Matsubayashi, J., & Tayasu, I. (2021). Diagnosing underdetermination in stable isotope mixing models. *PLOS ONE*, *16*(10), e0257818. https://doi.org/10.1371/journal.pone.0257818

Parravicini, V., Casey, J. M., Schiettekatte, N. M. D., Brandl, S. J., Pozas-Schacre, C., Carlot, J., Edgar, G. J., Graham, N. A. J., Harmelin-Vivien, M., Kulbicki, M., Strona, G., & Stuart-Smith, R. D. (2020). Delineating reef fish trophic guilds with global gut content data synthesis and phylogeny. *PLOS Biology*, *18*(12), e3000702. https://doi.org/10.1371/journal.pbio.3000702

Pethybridge, H., Choy, C. A., Logan, J. M., Allain, V., Lorrain, A., Bodin, N., Somes, C. J., Young, J., Ménard, F., Langlais, C., Duffy, L., Hobday, A. J., Kuhnert, P., Fry, B.,

Menkes, C., & Olson, R. J. (2018). A global meta-analysis of marine predator nitrogen stable isotopes: Relationships between trophic structure and environmental conditions. *Global Ecology and Biogeography*, *27*(9), 1043–1055. https://doi.org/10.1111/geb.12763

Phillips, D. L., Inger, R., Bearhop, S., Jackson, A. L., Moore, J. W., Parnell, A. C., Semmens, B. X., & Ward, E. J. (2014). Best practices for use of stable isotope mixing models in food-web studies. *Canadian Journal of Zoology*, *92*(10), 823–835. https://doi.org/10.1139/cjz-2014-0127

Phillips, D. L., Newsome, S. D., & Gregg, J. W. (2005). Combining sources in stable isotope mixing models: Alternative methods. *Oecologia*, *144*(4), 520–527. https://doi.org/10.1007/s00442-004-1816-8

Pickett, P. J., Dwyer, G. K., Macqueen, A., Holt, G., Halliday, B. T., Barton, J. L., & Lester, R. E. (2024). Using biotracer techniques to uncover consumer diets: A comparison of stable isotopes, fatty acids, and amino acids. *Ecosphere*, *15*(2), e4767. https://doi.org/10.1002/ecs2.4767

Plummer, M. (2003). JAGS: A program for analysis of bayesian graphical models using gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*.

Post, D. M. (2002). Using stable isotopes to estimate trophic position: Models, methods, and assumptions. *Ecology*, *83*(3), 703–718. https://doi.org/10.1890/0012-9658(2002)083%5B0703:USITET%5D2.0.CO;2

Post, D. M., Layman, C. A., Arrington, D. A., Takimoto, G., Quattrochi, J., & Montaña, C. G. (2007). Getting to the fat of the matter: Models, methods and assumptions for dealing with lipids in stable isotope analyses. *Oecologia*, *152*(1), 179–189. https://doi.org/10.1007/s00442-006-0630-x

Quezada-Romegialli, C., Jackson, A. L., Hayden, B., Kahilainen, K. K., Lopes, C., & Harrod, C. (2018). tRophicPosition, an r package for the bayesian estimation of trophic position from consumer stable isotope ratios. *Methods in Ecology and Evolution*, *9*(6), 1592–1599. https://doi.org/10.1111/2041-210X.13009

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reid, D. J., Quinn, G. P., Lake, P. S., & Reich, P. (2008). Terrestrial detritus supports the food webs in lowland intermittent streams of south-eastern australia: A stable isotope study. *Freshwater Biology*, *53*(10), 2036–2050. https://doi.org/10.1111/j.1365-2427.2008.02025.x

Rennie, N. (2023). *Ggflowchart: Flowcharts with 'ggplot2'*. https://nrennie.github.io/ggflowchart/

Santos, E. P., Condini, M. V., Santos, A. C. A., Alvarez, H. M., Moraes, L. E. de, Garcia, A. F. S., & Garcia, A. M. (2020). Spatio-temporal changes in basal food source assimilation by fish assemblages in a large tropical bay in the SW atlantic ocean. *Estuaries and Coasts*, *43*(4), 894–908. https://doi.org/10.1007/s12237-020-00716-1

Schwarcz, H. P. (1991). Some theoretical aspects of isotope paleodiet studies. *Journal of Archaeological Science*, *18*(3), 261–275. https://doi.org/10.1016/0305-4403(91)90065-W

Semmens, B. X., Ward, E. J., Moore, J. W., & Darimont, C. T. (2009). Quantifying inter- and intra-population niche variability using hierarchical bayesian stable isotope mixing models. *PLOS ONE*, *4*(7), e6187. https://doi.org/10.1371/journal.pone.0006187

Shipley, O. N., Dabrowski, A. J., Bowen, G. J., Hayden, B., Pauli, J. N., Jordan, C., Anderson, L., Bailey, A., Bataille, C. P., Cicero, C., Close, H. G., Cook, C., Cook, J. A., Desai, A. R., Evaristo, J., Filley, T. R., France, C. A. M., Jackson, A. L., Kim, S. L., … Newsome, S. D. (2024). Design, development, and implementation of IsoBank: A centralized repository for isotopic data. *PLOS ONE*, *19*(9), e0295662. https://doi.org/10.1371/journal.pone.0295662

Skinner, M. M., Martin, A. A., & Moore, B. C. (2016). Is lipid correction necessary in the stable isotope analysis of fish tissues? *Rapid Communications in Mass Spectrometry*, *30*(7), 881–889. https://doi.org/10.1002/rcm.7480

Smith, J. A., Mazumder, D., Suthers, I. M., & Taylor, M. D. (2013). To fit or not to fit: Evaluating stable isotope mixing models using simulated mixing polygons. *Methods in Ecology and Evolution*, *4*(7), 612–618. https://doi.org/10.1111/2041-210X.12048

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, *11*(1), 54–71. https://doi.org/10.1037/1082-989X.11.1.54

Solomon, C. T., Carpenter, S. R., Clayton, M. K., Cole, J. J., Coloso, J. J., Pace, M. L., Vander Zanden, M. J., & Weidel, B. C. (2011). Terrestrial, benthic, and pelagic resource use in lakes: Results from a three-isotope bayesian mixing model. *Ecology*, *92*(5), 1115–1125. https://doi.org/10.1890/10-1185.1

Stan Development Team. (2024). *Stan reference manual*. https://mc-stan.org

Stephens, R. B., Shipley, O. N., & Moll, R. J. (2023). Meta-analysis and critical review of trophic discrimination factors (Δ^{13}C and Δ^{15}N): Importance of tissue, trophic level and diet source. *Functional Ecology*, *37*(9), 2535–2548. https://doi.org/10.1111/1365-2435.14403

Stock, B. C., Jackson, A. L., Ward, E. J., Parnell, A. C., Phillips, D. L., & Semmens, B. X. (2018). Analyzing mixing systems using a new generation of bayesian tracer mixing models. *PeerJ*, *6*, e5096. https://doi.org/10.7717/peerj.5096

Stock, B. C., & Semmens, B. X. (2016). Unifying error structures in commonly used biotracer mixing models. *Ecology*, *97*(10), 2562–2569. https://doi.org/10.1002/ecy.1517

Ushey, K., & Wickham, H. (2024). *Renv: Project environments*. https://CRAN.R-project.org/package=renv

Vander Zanden, M. J., & Rasmussen, J. B. (1999). Primary consumer δ^{13}C and δ^{15}N and the trophic position of aquatic consumers. *Ecology*, *80*(4), 1395–1404. https://doi.org/10.1890/0012-9658(1999)080%5B1395:PCCANA%5D2.0.CO;2

Vander Zanden, M. J., & Rasmussen, J. B. (2001). Variation in δ^{15}N and δ^{13}C trophic fractionation: Implications for aquatic food web studies. *Limnology and Oceanography*, *46*(8), 2061–2066. https://doi.org/10.4319/lo.2001.46.8.2061

Vane, K., Cobain, M. R. D., & Larsen, T. (2025). The power and pitfalls of amino acid carbon stable isotopes for tracing origin and use of basal resources in food webs. *Ecological Monographs*, *95*(1), e1647. https://doi.org/10.1002/ecm.1647

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. (2018). Using stacking to average bayesian predictive distributions (with discussion). *Bayesian Analysis*, *13*(3), 917–1007. https:

//doi.org/10.1214/17-BA1091

Zapata-Hernández, G., Sellanes, J., Letourneur, Y., Harrod, C., Morales, N. A., Plaza, P., Meerhoff, E., Yannicelli, B., Carrasco, S. A., Hinojosa, I., & Gaymer, C. F. (2021). Tracing trophic pathways through the marine ecosystem of rapa nui (easter island). *Aquatic Conservation: Marine and Freshwater Ecosystems*, *31*(2), 304–323. https://doi.org/10.1002/aqc.3500