Received: 13 June 2025 | Accepted: 28 November 2025
DOI: 10.1111/2041-210x.70225

BRITISH ECOLOGICAL SOCIETY Methods in Ecology and Evolution

RESEARCH ARTICLE

A global estimator of C and N isotope baselines for fresh waters

Marie-Elodie Perga[1] | **Stéphanie Bouletreau**[2] | **Julien Cucherousset**[2] | **Chris Harrod**[3] | **Angus McIntosh**[4] | **Julian D. Olden**[5] | **Chloé Vagnon**[6] | **Tim Jardine**[7]

[1]Faculté des Géosciences et Environnement, Université de Lausanne, Lausanne, Switzerland; [2]Université de Toulouse, Toulouse INP, CNRS, IRD, CRBE, Toulouse, France; [3]Scottish Centre for Ecology and the Natural Environment, School of Biodiversity, One Health and Veterinary Medicine, University of Glasgow, Glasgow, UK; [4]School of Biological Sciences, University of Canterbury, Christchurch, New Zealand; [5]School of Aquatic and Fishery Sciences, University of Washington, Seattle, USA; [6]Fondation pour la Recherche sur la Biodiversité—Centre de Synthèse et d'Analyse sur la Biodiversité (FRB-CESAB), Montpellier, France and [7]School of Environment and Sustainability, Toxicology Centre, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

Correspondence
Marie-Elodie Perga
Email: marie-elodie.perga@unil.ch

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung; Fondation pour la Recherche sur la Biodiversite; Fondation Herbette; INEE CNRS

Handling Editor: David Soto

Abstract

1. Baselines are the pebbles in the shoes of isotope ecologists. The extreme variability of the isotope composition of resources at the base of food webs governs the spatial differences of consumers' isotope composition, so that isotope-inferred trophic properties can be compared across ecosystems only after correction for baseline effects. However, acquiring comparable and reproducible isotopic baselines in different lakes and rivers has been so challenging that many isotope datasets lack baseline values. Global estimates of C and N isotopic baselines would considerably expand the scope of large-scale isotope analyses in ecology.
2. Cross-referencing the global freshwater isotope database ISOFRESH (>800 sites across five continents) with a set of environmental attributes describing hydrology, physiography, climate, land use, soils and anthropogenic influences, we built data-driven models that predict C and N stable isotope compositions of benthic and pelagic/open-water baseline organisms for lakes and rivers with an error within 11%–13% of the overall range of values observed worldwide.
3. We then applied the models globally to predict spatial patterns in isotope baselines. We showed simulated baselines accurately preserve patterns in across-site variability for most of the common isotope-derived trophic metrics computed for freshwater fish at the population and community levels.
4. We conclude with guidance on the best use of such predictive baseline models, highlighting their usefulness for filling in gaps in meta-ecological analyses that test regional or global drivers of food web structure, but caution against substituting them for measured values in local-scale studies.

KEYWORDS
baselines, food webs, isotope-based trophic metrics, stable-isotope studies, trophic ecology

1 | INTRODUCTION

The analysis of natural abundance stable isotopes of carbon (C) and nitrogen (N) is widely used across a diversity of scientific fields, and it is now a pivotal tool in trophic ecology (Majdi et al., 2018). The method's basic principle is the transmissibility of isotope compositions from the assimilated food to the consumer's tissues, after accounting for a relatively predictable heavy-isotope enrichment at each trophic level (De Niro & Epstein, 1978, 1981). As such, a consumer's C and N isotope composition captures a time- and space-integrated signal of its dietary sources and the number of trophic steps from the primary producer to the focal organism (Layman et al., 2012; Post, 2002). Stable isotope techniques can infer continuous trophic positions, links and weighted biomass flows in food webs in a highly time- and cost-efficient manner (Layman et al., 2012; Post, 2002).

Stable isotope analysis has long been regarded as holding great promise to more efficiently test classical ecological theories and enable large-scale cross-ecosystem comparisons of food web structures (Cabana & Rasmussen, 1996; Kling et al., 1992; Post et al., 2000; Vander Zanden et al., 1999). The democratization of stable isotope analyses in trophic ecology in recent decades has indeed generated synthetic isotope datasets (Bodin et al., 2021; Hixon et al., 2024), but their value at continental or global scales is yet to be fully realized. Cross-ecosystem isotope studies have improved knowledge of environmental or human factors driving food web structure (Layman et al., 2007; Lejeune et al., 2024; Post et al., 2000; Vander Zanden et al., 1999; Wainright et al., 2021), but such studies remain limited at larger spatial extents due to a host of persistent challenges that complicate comparative analyses (Stiling et al., 2023; Vander Zanden & Fetzer, 2007).

Transforming raw C and N isotope data into meaningful and comparable trophic metrics for ecological inquiry (i.e. species resource use or trophic position, trophic diversity or food chain length) has been a significant impediment to broad cross-ecosystem comparisons of food web structure and properties (Kjeldgaard et al., 2021; Vander Zanden & Fetzer, 2007). Core to this challenge is the simple fact that the same primary resource at the base of the food web can exhibit considerable variability in isotope composition across space and study sites. For instance, Gu (2009) and Gu et al. (2011) reported seston δ^{13}C and δ^{15}N spanning >30 ‰ and 15 ‰ isotopic ranges, respectively, across lakes worldwide. This causes across-ecosystem variation in consumer δ^{13}C and δ^{15}N values to arise largely, if not primarily, from differing C and N isotope values of the primary resources, which are also referred to as isotopic baselines or endmembers. Variation in consumer trophic position or habitat use may only act as a secondary determinant of observed differences in δ^{13}C and δ^{15}N (Arnoldi et al., 2024; Cabana & Rasmussen, 1996). Comparing isotope-inferred trophic properties across ecosystems requires reproducible metrics corrected for such baseline effects (Hoeinghaus & Zeug, 2008; Post, 2002; Vander Zanden & Fetzer, 2007; Vander Zanden & Rasmussen, 1999).

Baselines are therefore the pebbles in the shoes of isotope ecologists. Because isotopic baselines should reflect the isotope composition of available primary producers over temporal and spatial scales that match those of the consumers (Heuvel et al., 2023), long-lived primary consumers have usually been targeted for baselines. For instance, bivalves and snails have typically been standards for assessing pelagic and benthic baselines, respectively, in lake studies (Cabana & Rasmussen, 1996; Post, 2002). However, such organisms are not ubiquitous across ecosystems of interest, and ecosystems may lack other species of functional equivalence (Barrus et al., 2024; Jardine et al., 2014; Kristensen et al., 2016). Past attempts to circumvent this problem have led to vastly different short-lived primary consumers (e.g. herbivorous insects or zooplankton) and primary producers (periphyton, seston) being used as baselines (Jepsen & Winemiller, 2002; Kling et al., 1992), resulting in the nature, turnover time and trophic level of organisms used as baselines being highly heterogeneous between systems and sites. Challenges with acquiring comparable isotopic baselines are only further compounded by many isotope datasets containing incomplete or unreproducible baseline values. For instance, in the global lake isotope dataset collated by Stiling et al. (2023), pelagic and benthic baselines were missing for 80% of the sites. Finding a way to overcome these baseline challenges would lead to faster progress by facilitating large-scale comparisons.

Global estimates of C and N isotopic baselines hold the key to leveraging the considerable power of global-scale isotope analyses in ecology. We address this challenge by exploring whether C and N isotopic baselines can be robustly predicted for freshwater ecosystems based on their environmental characteristics. This objective is premised on spatial variation of isotopic baselines stemming from local environmentally driven differences in C and N biogeochemical processes. For instance, spatial variations in primary producer δ^{13}C values are mainly driven by the isotopic composition of the dissolved inorganic carbon source and the extent of isotopic fractionation during photosynthesis, themselves depending on local environmental conditions such as temperature, CO_2 concentrations and biological communities (Finlay, 2004; Goericke & Fry, 1994; Hecky & Hesslein, 1995; Laws et al., 1997). Such premises underpin the mechanistic or statistical models developed to predict the local isotopic composition of terrestrial plants or marine plankton as a function of observed local and/or extra-local environmental variables (Espinasse et al., 2022; Magozzi et al., 2017; St John Glew et al., 2021). These 'isoscape' models (Bowen, 2010; Trueman et al., 2017; Yoshikawa et al., 2024) offer surrogates for local isotopic baselines in marine and terrestrial studies (Mullineaux et al., 2022; St John Glew et al., 2021). Similar links between local environmental conditions and aquatic primary producer isotope composition (Cabana & Rasmussen, 1996; Perga & Gerdeaux, 2004) suggest strong potential for the development of large-scale predictive models for C and N isotope baselines in freshwater environments.

The ISOFRESH database, that is, a global database of biological δ^{13}C and δ^{15}N values from >800 inland waters across the world (Bouletreau et al., 2025), provides a new opportunity to develop a baseline isotopic model at the global scale for freshwaters. Cross-referencing ISOFRESH with a set of environmental attributes describing hydrology, physiography, climate, land use, soils and anthropogenic influences, we built data-driven statistical models that predict C and N stable isotope compositions of benthic and pelagic/open-water baselines for lakes and rivers. We then applied the models to predict global spatial patterns in isotope baselines, testing whether simulated baselines accurately preserved patterns in across-site variability of common isotope-derived trophic metrics computed for freshwater fishes at the population and community levels (Jackson et al., 2011; Layman et al., 2007). We conclude with guidance on the best use of predictive baseline models, highlighting their utility for filling in gaps in meta-ecological analyses that test regional or global drivers of food web structure. This guidance includes cautionary notes regarding their substitution for measured values in individual food webs.

2 | MATERIALS AND METHODS

2.1 | Isotope dataset

The ISOFRESH's version used in this study contains species-level fish isotopic data from 509 lentic and 363 lotic sites and the accompanying data for non-fish organisms, retrieved from 285 published studies (Bouletreau et al., 2025). Non-fish organisms varied in taxonomy and aggregation levels across studies (micro and macro-algae, biofilms and macroinvertebrates ranging from sponges to molluscs and arthropods). They were classified based on their trophic level (i.e. primary producers, primary consumers or carnivores) from the source information. For the baseline prediction model, we filtered the original dataset to retain only non-fish entries. Primary producers were regarded as potential isotopic baselines, with seston attributed to pelagic habitats (or open-water for streams–rivers thereafter referred to as pelagic) and benthic mats, biofilms, periphyton, macrophytes to benthic habitats. All zooplankton species except ostracods were considered as proxies for pelagic baselines, as were bivalve molluscs. Other macroinvertebrates were classified as benthic primary consumers, or not attributed when taxonomic resolution was too coarse. Non-fish entries that did not qualify for any of these categories were removed.

Of the lotic and lentic sites in the original database, 77% and 69%, respectively, included at least one isotopic baseline, resulting in 279 lotic and 353 lentic unique site entries in the baseline database. Lotic and lentic sites were distributed across all continents except Antarctica (Figure 1a). In lentic sites, pelagic and benthic baselines were equally frequent (86% vs. 83%, respectively), and both baselines were reported for 69% of the lentic sites. Benthic baselines were measured in all lotic sites, but estimates for pelagic baselines were present for only 30% of sites (Figure 1b–e). The taxa used as baselines in the different studies reported included considerable taxonomic diversity. Macroinvertebrate taxa not only included (in decreasing order of occurrence) mayflies, gastropods, midges, bivalves, caddisflies and shrimps, but also, in some cases, sponges. Even for primary consumer baselines, the level of taxonomic reporting varied considerably, from the species name up to a high level of aggregation (i.e. 'benthic macro-invertebrates'). For primary producers, baselines included (in decreasing order of occurrence) biofilms scraped from plants or rocks, macrophyte samples and seston. Baseline δ^{13}C values varied across a 30‰ (−41‰ to −11‰) and 26‰ (−40‰ to −14‰) range for the lentic and lotic datasets, respectively. The range of δ^{15}N baselines across sites was narrower, with a minimum value of −1‰ and a maximum value of 17‰ in lentic sites and minimum/maximum values of −2‰/18‰ in lotic sites. Baselines were more frequently estimated from primary consumers (60% of baselines for lotic and 74% for lentic sites) than from primary producers.

2.2 | Environmental dataset

We considered a suite of environmental attributes, including variables related to basin physiography, hydrology, climate, land use, soils and anthropogenic influences. Variables included: (1) surface elevation (m a.s.l.); (2) lake surface area (km^2); (3) river surface area (km^2) based on estimated channel and bankfull width (Lehner & Grill, 2013); (4) modelled long-term (1971–2000) average annual 'naturalized' discharge (m^3.s^{-1}) according to the global integrated water model WaterGap v2.2 (Döll et al., 2003); (5) Strahler stream order; (6) long-term (1950–2000) average annual air temperature (°C); and (7) average total annual precipitation (mm) according to WorldClim (Hijmans et al., 2005); (8) modelled average organic carbon content in soils of the upstream catchment (Hengl et al., 2014); (9) human population count (2010); and (10) human footprint index (2009) representing a composite measure of the direct and indirect human pressures on the land's surface within a 3 km vicinity of the lake or in the reach catchment (Venter et al., 2016). Different variable combinations were used for river versus lake models. Data were sourced from HydroATLAS v. 1.0.1, specifically RiverATLAS (Linke et al., 2019) and LakeATLAS (Lehner et al., 2022). The database included an extensive range of sizes and elevations, with sites within forested or relatively undeveloped catchments up to highly urbanized and agricultural catchments (Table 1). Predictors were checked for outliers (i.e. capped when unrealistic extremes were detected) and log-transformed to reduce skewness when necessary.

2.3 | Predictive models

Because we not only expected non-linear relationships, but also aimed at preserving some interpretability, we used Generalized Additive Models (GAMs) to predict baselines from the selected set of predictors (Hastie & Tibshirani, 1987). Models were specified as:

$$Y = b + f_1(X_1) + f_2(X_2)\ \dots\ + f_n(X_n) \tag{1}$$

where Y is the response variable ($\delta^{13}C_{lotic}$, $\delta^{13}C_{lentic}$, $\delta^{15}N_{lotic}$ or $\delta^{15}N_{lentic}$), b the intercept, f_i the smoothing functions (thin-plate

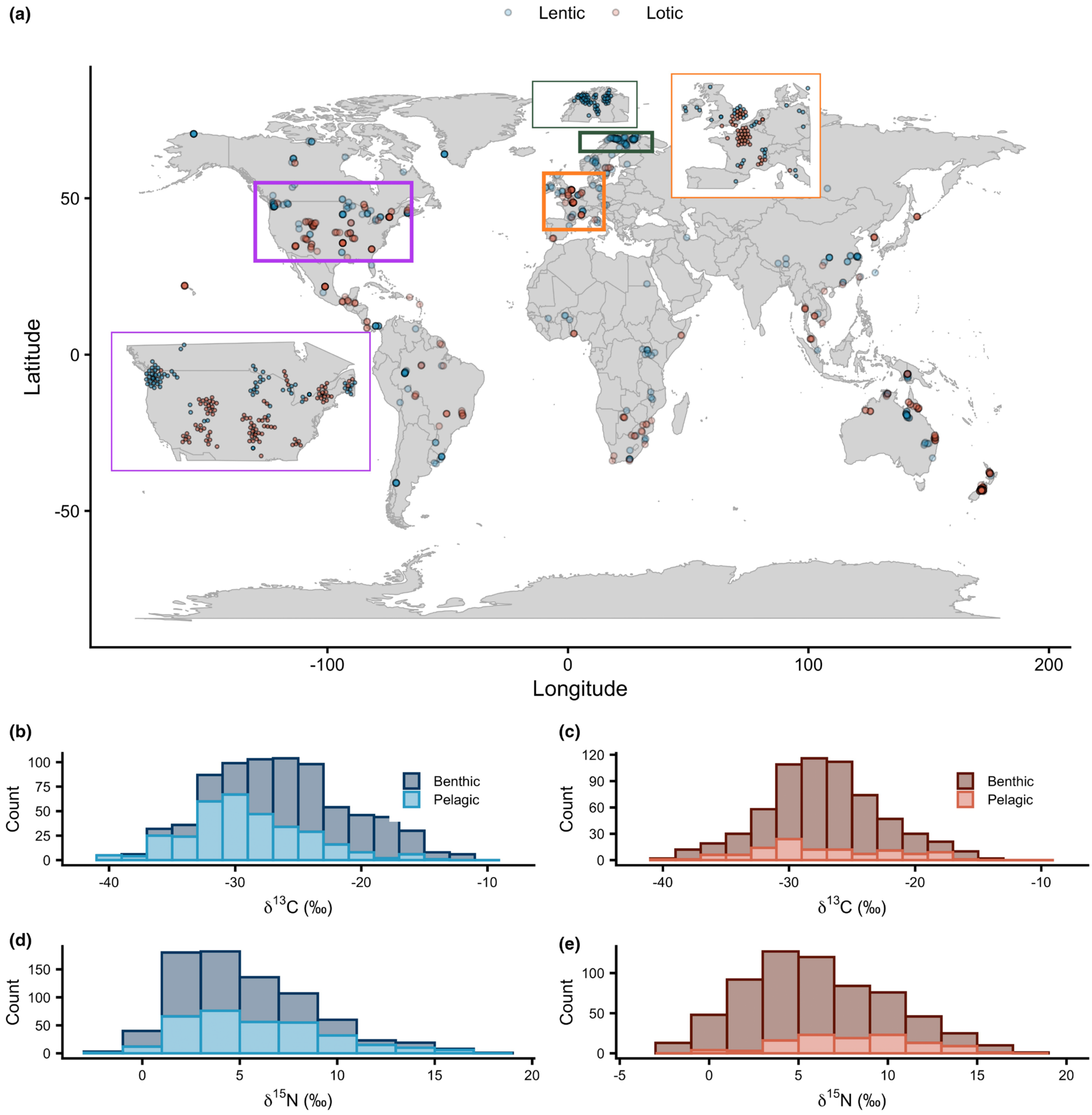

FIGURE 1 The available global (a) dataset of $\delta^{13}C$ (b and c) and $\delta^{15}N$ (d and e) values for pelagic and benthic baselines in lentic (blue) and lotic (red) ecosystems.

regression splines) and X_i the covariates. Predictive models were run separately for lotic and lentic systems as available covariates varied between ecosystem types. The full datasets were randomly split into training (80%) and test (20%) sub-datasets. The trophic level of the organism used as a baseline (TL, i.e. primary producers or consumers) and its trophic habitat (HAB, i.e. pelagic or benthic) were included as fixed, factorial terms in the GAMs, as was the Strahler order (STRA) for lotic sites. The most parsimonious models were selected using a double penalty approach, using the Akaike Information Criterion, with concurvity control for predictors (Ramsay et al., 2003). Model optimization was conducted manually, starting with backward selection from the full model without interactions, used as a benchmark for explained deviance. Interactions were introduced once variables with significant predictive power were selected. Only interactions between continuous and factorial terms were included, in order to minimize the models' effective degrees of freedom. Because our purpose was predictive rather than explanatory, the wiggliness of each smoother was set to maximize likelihood through its *K*-basis dimensions. However, we controlled the risk for overfitting, due to excessive wiggliness or too many variables interactions, and for instability, due to concurvity,

TABLE 1 Observed range of values (min-max) for the covariates used as predictors within the lentic and lotic datasets. Blank cells in the table indicate variables not used in a model.

Covariates (abbreviation, unit)	Lentic sites	Lotic sites
Latitude (*lat*, °)	−41.1–70.6	−43.6–61.2
Longitude (*lon*, °)	−155.5–175.9	−159.4–175.4
Elevation (*ele*, m asl)	−1.0–4724	1–2844
Lake Area (*area*, km^2)	0.1–81,843	
River area along the reach segment (*ria*, ha)		0.2–796.0
Annual average discharge (*dis*, m^3.s^{-1})		0.009–15,177
Strahler stream order (*STRA*, class)		1–9
Annual average air temperature (*tmp*, °C)	−13.6–28.5	−4.1–27.9
Annual accumulated precipitation (*pre*, mm)	0–4620	205–4952
Average organic carbon content in soils of the upstream catchment (*soc*, tons.ha^{-1})	8–246	6–139
Human population count (*pop*, thousands within 3 km vicinity of the lake or in the reach catchment)	0–3946	0.0–42.3
Human footprint (*hft* dimensionless, average within 3 km vicinity of the lake or in the reach catchment) for the year 2009	0–389	0–445

by comparing the prediction accuracy for the optimal models between the training and test sub-datasets. All models were validated by checking a posteriori the homogeneity of variance and normal distribution of residuals. GAMs were computed using the mgcv R package (Wood, 2011). The best-fit models were then used to predict the dataset's baselines and associated errors. GAMs are reproducible but non-parametric models; that is, unlike linear models, they do not come with a set of fitted parameters to be used for rerunning simulations. For the use of the best models for further baseline estimations, we refer users to the associated repository (Perga, 2025).

2.4 | Relevance of simulated baselines for typical isotope-derived food-web metrics

Typical isotope trophic metrics, at the population- or community-levels, were computed for all fish isotope entries of the dataset ($\delta^{13}C_{fish}$; $\delta^{15}N_{fish}$), from sites where both baselines had been reported, that is, 35% of the fish entries and 44% of the sites. Baselines for the computations were the measured or predicted values at the primary consumers' level ($\delta^{13}C_{pelagic/benthic}$; $\delta^{15}N_{pelagic/benthic}$). Trophic discrimination factors were first set as fixed ($\Delta^{13}C=0.5‰$ and $\Delta^{15}N=3.4‰$ per trophic level), but in a second step, these were varied and randomly drawn from a normal distribution around the fixed values (SD of 1.3‰ and 1.0‰ for $\Delta^{13}C$ and $\Delta^{15}N$, respectively; Post, 2002).

Population-level isotope-based metrics were computed from the fish isotope data at the species level and for each site, and encompassed:

(i) $\alpha_{Benthic}$, that is, the consumers' reliance on the benthic trophic habitat (Vander Zanden et al., 1999)

$$\alpha_{Benthic} = \frac{\delta^{13}C_{fish} - \delta^{13}C_{benthic} - \Delta^{13}C}{\left(\delta^{13}C_{Pelagic} - \delta^{13}C_{benthic}\right)} \tag{2}$$

(ii) the species' trophic position (TP) computed from either the unweighted average ($TP_{unweighted}$) (Olsson et al., 2009) or weighted average (based on the per cent reliance on habitats) of baseline $\delta^{15}N$ values ($TP_{weighted}$, Vander Zanden et al., 1999):

$$TP_{unweighted} = 2 + \frac{\left[\delta^{15}N_{fish} - \text{mean}\left(\delta^{15}N_{baselines}\right)\right]}{\Delta^{15}N} \tag{3.1}$$

$$TP_{weighted} = 2 + \frac{\left[\delta^{15}N_{fish} - \delta^{15}N_{pelagic} - \alpha_{Benthic} \cdot \left(\delta^{15}N_{benthic} - \delta^{15}N_{pelagic}\right)\right]}{\Delta^{15}N} \tag{3.2}$$

Computations of $\alpha_{Benthic}$ and $TP_{weighted}$ require that the fish $\delta^{13}C$ (minus the allocated $\Delta^{13}C$) falls within the range of the measured $\delta^{13}C$ of the baselines. For such metrics, the comparison was therefore restricted to the fish entries that fulfilled that condition, that is, 43% of the dataset. Because it is independent from baseline $\delta^{13}C$ values, the unweighted TP could be computed for all dataset entries with corresponding baselines.

The error due to the use of the simulated baselines ($Error_{baseline}$) was computed as the difference of the population-specific metrics values computed from the observed $Metric_{Obs_{baseline}}$ or simulated $Metric_{Pred_{baseline}}$ baselines, using with a fixed isotope trophic discrimination factor; that is, $Metric_{Obs_{baseline}/Fixed_{frac}}$ and $Metric_{Pred_{baseline}/Fixed_{frac}}$, respectively:

$$Error_{baseline} = Metric_{Obs_{baseline}/Fixed_{frac}} - Metric_{Pred_{baseline}/Fixed_{frac}} \tag{4.1}$$

The error due to the uncertainty in the trophic discrimination factor ($Error_{fractionation}$) was computed as the difference of the population-specific metrics values computed from the fixed or variable trophic discrimination factors, using the observed baselines; that is, $\text{Metric}_{\text{Obs}_{\text{baseline}}/\text{Fixed}_{\text{frac}}}$ and $\text{Metric}_{\text{Obs}_{\text{baseline}}/\text{Var}_{\text{frac}}}$, respectively:

$$\text{Error}_{\text{fractionation}} = \text{Metric}_{\text{Obs}_{\text{baseline}}/\text{Fixed}_{\text{frac}}} - \text{Metric}_{\text{Obs}_{\text{baseline}}/\text{Var}_{\text{frac}}} \tag{4.2}$$

The total error ($\text{Error}_{\text{Total}}$) was computed as the difference of the population-specific metrics between the values computed from the fixed trophic discrimination factor and observed baselines; that is, $\text{Metric}_{\text{Obs}_{\text{baseline}}/\text{Fixed}_{\text{frac}}}$, and those computed from the variable trophic discrimination factor and simulated baselines; that is, $\text{Metric}_{\text{Pred}_{\text{baseline}}/\text{Var}_{\text{frac}}}$:

$$\text{Error}_{\text{Total}} = \text{Metric}_{\text{Obs}_{\text{baseline}}/\text{Fixed}_{\text{frac}}} - \text{Metric}_{\text{Pred}_{\text{baseline}}/\text{Var}_{\text{frac}}} \tag{4.3}$$

Community-wide metrics, computed per site and from both the observed and simulated baselines, included:

(i) Metrics related to the food chain, that is, food chain length (FCL, the maximum unweighted fish trophic position) and mean (unweighted) TP,
(ii) Metrics related to trophic similarity and redundancy between species, that is, mean distance to centroid (CD), mean nearest neighbour distance (NND) and standard deviation of the nearest neighbour distance (SDNND) (Layman et al., 2007);
(iii) Metrics quantifying the total amount of niche space occupied, estimated either from the Total Area of the convex hull (TA, Layman et al., 2007) or the Standard Ellipse Area (SEA and SEAc; when corrected for sample size; Jackson et al., 2011).

FCL and TPmean were directly computed from the population-level metrics while the Layman et al. (2007) and Jackson et al. (2011) metrics were computed from the SIBER R-Package (Jackson et al., 2011) after baseline standardization according to Olsson et al. (2009):

$$\delta^{13}\text{C}_{\text{cor}} = \frac{\delta^{13}\text{C}_{\text{fish}} - \text{mean}\left(\delta^{13}\text{C}_{\text{baselines}}\right)}{\text{abs}\left(\delta^{13}\text{C}_{\text{pelagic}} - \delta^{13}\text{C}_{\text{benthic}}\right)} \tag{5.1}$$

$$\delta^{15}\text{N}_{\text{cor}} = \delta^{15}\text{N}_{\text{fish}} - \text{mean}\left(\delta^{15}\text{N}_{\text{baselines}}\right) \tag{5.2}$$

A summary of variable names and descriptions can be found in Table SI-1.

The relevance and utility of the population and community-wide metrics for large-scale studies were tested first by measuring the linear correlation between the values of the metrics computed from the observed and simulated baselines. Additionally, we tested whether and which of the metrics computed from the estimated baselines were useful for macro-ecological studies, that is, whether they faithfully reflected the trends across sites. For population-wide metrics, we selected the nine most represented fish species in the dataset and compared their worldwide distribution of metrics as computed from the observed and estimated baselines, using Kolmogorov–Smirnov tests. To quantitatively test whether the use of simulated baselines preserved the overall spatial variability structure of the population or community metrics, we applied Kendall's coefficient of concordance, which measures the agreement of ranking of the metrics across sites.

3 | RESULTS

3.1 | Models for baseline predictions

Four models were selected based on their fit and capacity to fulfil GAM requirements (Table 2, Figure 2, Figure SI-2 for the diagnostic residuals plots of models). Many of the predictors showed significant concurvity with the geographical coordinates of the sites. Limiting concurvity in best-fit models therefore required an analytical decision whether to include either spatial coordinates or climatic variables as predictors in the models. For lentic systems, best-fit models necessarily included spatial coordinates, whereas models including climatic proxies produced lower goodness-of-fit (Table 2). For lotic systems, best-fit models were those excluding spatial coordinates (Table 2). For both lentic and lotic systems, fixed effects modelled higher δ^{13}C baselines for benthic than pelagic habitats, although the effect size was twice as large for lakes (Figure SI-3). TL fixed effects ended with δ^{15}N baselines greater for primary consumers than resources, with also a larger effect size in lakes. Anthropogenic impact, either as the human footprint or population in the catchment, was retained as a predictor for the N isotope baselines (Figure SI-3).

The best-fit model for lentic δ^{13}C also included latitude, longitude and lake area, with smooth effects leading to higher δ^{13}C baseline values for tropical and high latitude regions, and increased baseline δ^{13}C with lake size, although the effect varied between benthic and pelagic habitats (Figure SI-3). Additional predictors of the best-fit model for lentic δ^{15}N were elevation and human footprint. Smoothers modelled greater baseline δ^{15}N at mid-latitude, decreasing δ^{15}N with altitude and lake size, and a non-linear effect of human footprint (Figure SI-3).

In lotic systems, the Strahler order had strong predictive power for both C and N baselines and was a significant modulator of the smooth functions for both climatic and physiographic predictors (Table 2). Modelled climatic effects were non-linear for both δ^{13}C and δ^{15}N. The effect trend for soil organic carbon, retained as a predictor for streams and rivers δ^{13}C, implied decreasing δ^{13}C as soil OC increased, especially at lower Strahler order. Smoothers modelled greater baseline δ^{15}N at higher discharges and non-linear effects of human footprint as a function of the Strahler order (Figre SI-3).

Best-fit models explained 56%–65% of the total deviance of the baseline values. The quality of predictions did not decrease significantly between the training and the test datasets, confirming that models were stable and not overfitted (Table 2). The correlation between observations and modelled data was high (0.71–0.78, Figure 2). Overall, models fit better for δ^{15}N than for δ^{13}C, but model performance was similar for lotic and lentic systems. Root-mean-squared errors (RMSE) in the prediction of baselines were close to 4.0‰ for δ^{13}C and 2.5‰ for δ^{15}N (Table 2). Overall, the models for baseline δ^{13}C failed to reproduce the lowermost values for lentic ecosystems (< −35‰) and the uppermost values for lotic ecosystems (> −20‰).

TABLE 2 Equations for the best-fit models to predict δ^{13}C and δ^{15}N values of pelagic and benthic baselines in lentic and lotic ecosystems, estimated degrees of freedom (edf) and goodness-of-fit (Expl. dev.: explained deviance, *r*: correlation coefficient and RMSE: root mean squared error) on the training and test datasets.

					Benthic				Pelagic			
					Training dataset		Test dataset		Training dataset		Test dataset	
Ecosystem type	Response	Best models	edf	Expl. dev.	*r*	RMSE	*r*	RMSE	*r*	RMSE	*r*	RMSE
Lentic	δ^{13}C	*s(lat:HAB)+s(lon)+s(area:HAB)+HAB+TL*	36.1	56%	0.67	3.9 ‰	0.61	4.1 ‰	0.66	3.4 ‰	0.60	4.7 ‰
	δ^{15}N	*s(lat)+s(lon)+s(ele)+s(area)+s(hft)+HAB+TL*	34.9	60%	0.72	2.3 ‰	0.73	2.3 ‰	0.82	2.1 ‰	0.70	2.6 ‰
Lotic	δ^{13}C	*s(tmp:STRA)+s(pre:STRA)+s(soc:STRA)+HAB*	32.3	58%	0.75	2.8 ‰	0.56	2.8 ‰	0.73	3.1 ‰	0.58	3.2 ‰
	δ^{15}N	*s(dis:TL)+s(tmp:STRA)+s(pre)+s(pop:STRA)+TL*	47.8	65%	0.81	2.1 ‰	0.70	2.6 ‰	0.73	2.4 ‰	0.76	3.2 ‰

Note: *HAB* is a class variable referring to the benthic or pelagic nature of the baselines; *TL* is a class variable referring to the trophic level of the organism used as baselines (primary producer or consumer). Refer to Table 1 for other predictors' abbreviations.

3.2 | Relevance of simulated baselines for typical isotope-derived food-web metrics

To test the applicability of the modelled baselines, we used them to calculate a series of food web metrics at both population (α_{Benthic}, weighted and unweigthed TP) and community (FCL, TA, CD, NND, SDNND and SEA) levels and compared them to outputs from the measured baseline values. Population-level isotope metrics computed from the simulated baselines were all significantly correlated with those computed from observed baselines and aligned onto a 1:1 line with a normally distributed error, suggesting unbiased predictions (Figure 3). However, computing the metrics from the simulated baselines increased the spread in the values. Nevertheless, despite being correlated and unbiased, the benthic reliance computed from the simulated baselines spanned over a range five times wider than the range of values computed from the measured baselines (i.e. within 0–1) (Figure 3a). Because it is independent from the baseline δ^{13}C values, the unweighted metric for TP had a consistently better goodness-of-fit (Figure 3b) than the weighted metric (Figure 3c). Except for the benthic reliance (Figure 3d), errors in the TP estimates stemming from the use of simulated baselines were comparable to those due to the uncertainty in trophic discrimination (Figure 3e,f).

Community-level isotope-based metrics could be compared for only 25% of sites (the 221 sites where both baselines had been measured), and, due to a consistent lack of pelagic baselines in rivers, lotic sites were poorly represented (Figure 4). In all cases, metrics computed from simulated baselines were significantly correlated with those computed from observed baselines, and aligned onto a 1:1 line, suggesting an unbiased estimate. However, the strength of the correlation varied depending on the metric, with greatest correlations and lowest biases for the metrics related to community trophic diversity (CD, NND and SDNND, Figure 4d,e). Food chain length computed from the simulated baselines compared well with values computed from the observed baselines, but the correlation was much looser for the mean TP, despite both being independent from the C isotope baselines (Figure 4a,b). Finally, metrics computing the total isotopic niche (TA, SEA, SEAc) were the most sensitive to the source of the baseline estimates (simulated or observed) because, despite being strongly correlated and unbiased, they tended to result in extreme and unrealistic values in approximately 5% of sites (Figure 4f–h).

3.3 | Usefulness of the simulated baselines for macro-ecological approaches

Although there was a significant correlation of population- and community-wide metrics computed from the observed and

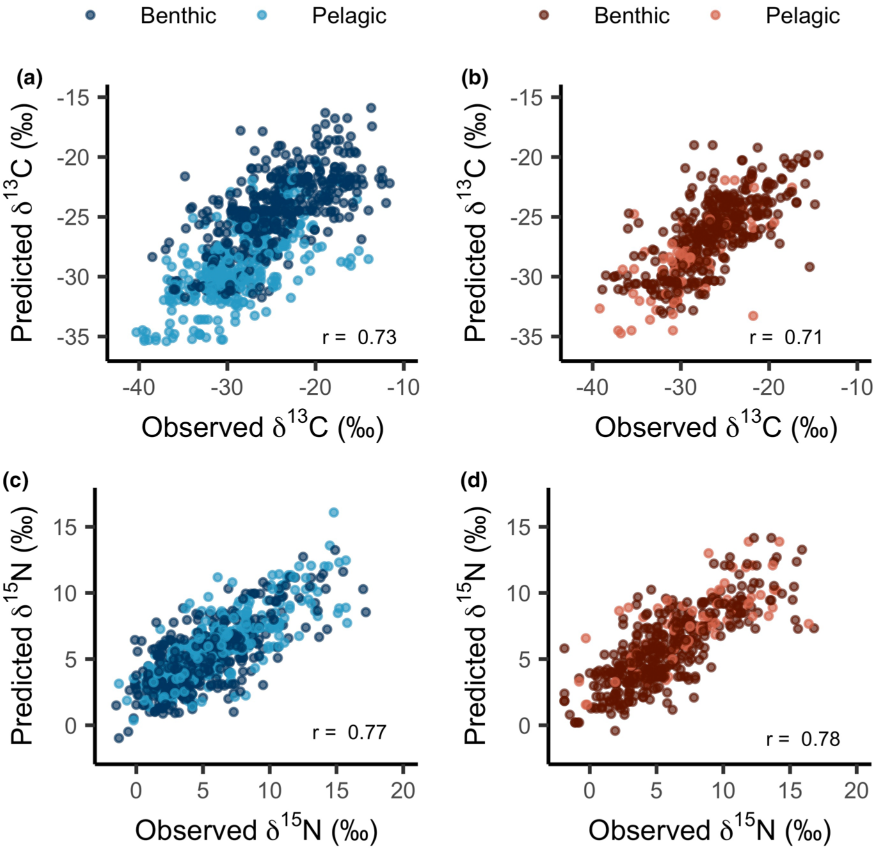

FIGURE 2 Goodness-of-fit for the selected generalized additive models predicting the $\delta^{13}C$ (a and b) and $\delta^{15}N$ (c and d) of baselines in lentic (blue) and lotic (red) freshwater ecosystems.

simulated baselines, the modelled metrics reproduced a variable range of the observed metrics variation, that is, from 9% in the worst case (α_{benthic}; Figure 3) up to 82% in the best case (CD; Figure 4). Therefore, we further assessed whether the error in the metrics computed from estimated baselines compromised the utility of the baseline estimator for species-specific, population- or community-wide macro-ecological studies (Figure 5).

For the nine most represented fish species in the dataset, there was no significant difference in the distribution of the unweighted TP between values computed from the observed or predicted baselines (Figure 5a and Table SI-4-1), and the rankings of values across sites were concordant between the observed and predicted baselines for all nine species (Figure 5b–d and Table SI-4-1). Therefore, for the unweighted TP, the error introduced by the use of estimated baselines did not modify the variability structure across sites (Figure 5b–d and Table SI-4-1). For α_{benthic} and the weighted TP, whose computation depends on the former, the distribution of the metric values could vary for some fish species (*C. carpio*, Table SI-4-1). The error introduced by the use of the simulated baselines led to the lack of concordance in the rankings of the metrics across sites for all species for α_{benthic}, and most species for the weighted TP (Table SI-4-1). In contrast to some of the population-wide metrics, ordered series were concordant between observed and simulated baselines for all community-wide metrics, confirming that the variability structure of community-wide metrics across sites was very robust to potential error introduced by the use of an estimated isotopic baseline (Figure 5e–g, Table SI-4-2).

4 | DISCUSSION

We report strong predictability of baseline C and N isotope values in lakes (lentic) and rivers (lotic) according to environmental attributes describing their waters and contributing watershed. We show that models are able to predict local baseline isotope values with an error within 2.5‰ to 4.0‰ for $\delta^{15}N$ and $\delta^{13}C$, respectively, that is, 11%–13% of the overall range of values observed worldwide (Figure 6). There is no previous attempt at creating baselines for inland waters to which performances could be compared, but models' performances were comparable to the performance of marine large-scale isoscape models (RMSE 2.5‰–3.0‰ for a 15‰ total range of $\delta^{13}C$, and 2.9‰–3.4‰ for a 15 ‰ total range of $\delta^{15}N$; e.g. St John Glew et al., 2021). The availability of these baselines helps address a persistent problem, opening up opportunities for global analyses, but the models are not suitable for local studies.

Although our models were developed to optimize prediction accuracy over interpretability, the most influential predictor variables were consistent with known drivers of spatial variation in C and N stable isotope composition of lower trophic levels (e.g. Casey & Post, 2011). The models integrate the overall role of climate variability either through coordinates (lentic sites) or climatic variables (lotic sites). The focal habitat (i.e. benthic or pelagic) was a significant factor for baselines in lentic ecosystems. The original dataset confirmed a significant ^{13}C-enrichment of benthic versus pelagic baseline organisms, consistent with the mean difference of 4‰–6‰ reported by Vander Zanden and Rasmussen (1999) and France (1995). Such a

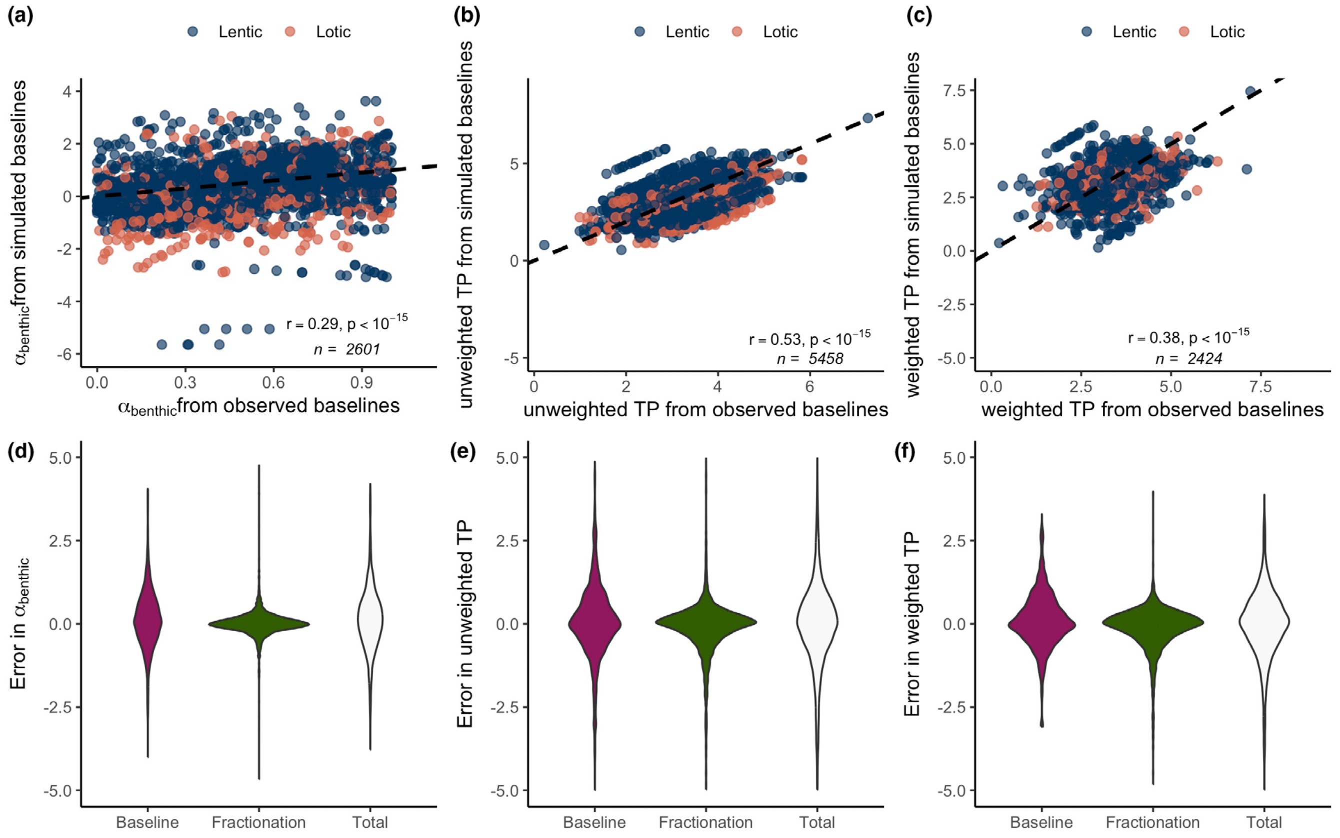

FIGURE 3 Comparison of population isotope metrics computed from measured or simulated baselines for both lentic (blue) and lotic (red) sites, including (a) benthic reliance, (b) unweighted trophic position (TP), and (c) weighted TP. Sample size (*n*), correlation coefficient (*r*) and *p*-value of the correlation test are indicated. The dotted line represents the 1:1 line. Also shown are error distributions for benthic reliance (d), and both unweighted (e) and weighted TP (f) estimates due to errors in the simulated baselines (purple), errors due to uncertainty in the trophic discrimination factor (green) and the sum of both (white).

difference is typically attributed to a less selective ^{13}C-fractionation during C-fixation in biofilms where the inorganic carbon supply is more limited by the diffusion at the boundary layer than for pelagic cells (France, 1995; Hecky & Hesslein, 1995).

For lotic sites, the model predicts fewer isotopic differences between benthic and pelagic endmembers, consistent with a released boundary layer effect at the biofilm's interface (Finlay, 2004). Predicted baseline δ^{13}C values in rivers and lakes also matched earlier observations and theory on the role of metabolism in influencing CO_2 concentrations and the isotope composition of dissolved inorganic carbon. For instance, Strahler order appeared in the lotic models, consistent with a change from respiration-dominated, low productivity headwaters to more productive and occasionally CO_2-limited, lower main-channel reaches (Finlay, 2001; Wollheim et al., 2022). The lotic δ^{13}C baselines were also predicted to decrease as the soil organic carbon content of the catchment increased, as expected for the transition from low to high CO_2 systems. Similarly, lake size was a significant predictor of lentic baselines, a regularly observed pattern attributed to the size dependence of lake metabolism (Perga & Gerdeaux, 2004; Post, 2002).

By contrast to δ^{13}C, the δ^{15}N of benthic baselines at lentic sites is predicted to be slightly lower than pelagic baselines, a regularly observed pattern (Vander Zanden & Rasmussen, 1999) attributed to a redox-dependent nitrogen source for primary producers (i.e. ammonia in oxygen-limited benthic environments versus nitrates in the oxic open water; Cadeau et al., 2021). The pelagic/benthic differences were less pronounced in lotic systems, consistent with typically more homogeneous N sources within a given site in flowing waters (Ribot et al., 2013). For nitrogen, anthropogenic influence, either as human population in the catchment or human development index, was significant in both the lentic and lotic models in line with expected higher δ^{15}N values associated with potentially greater agricultural runoff and sewage (Cabana & Rasmussen, 1996; Diebel & Zanden, 2009; Vander Zanden et al., 2005), a relationship that was, however, clearer for rivers of higher Strahler orders. Such a consistency between known drivers of isotope variability at a large spatial scale and the shape of the smoothers within the models provides confidence that our models are mechanistically sound.

Baseline predictions were more accurate for δ^{15}N than for δ^{13}C, reducing the accuracy of some of the trophic metrics strongly dependent on C isotope baselines when computed from simulated baselines. Both the lotic and lentic models failed to simulate the extreme baseline values for δ^{13}C. This limitation of the δ^{13}C baseline predictors may arise first from the lack of functionally relevant environmental variables included in the model. For instance, current

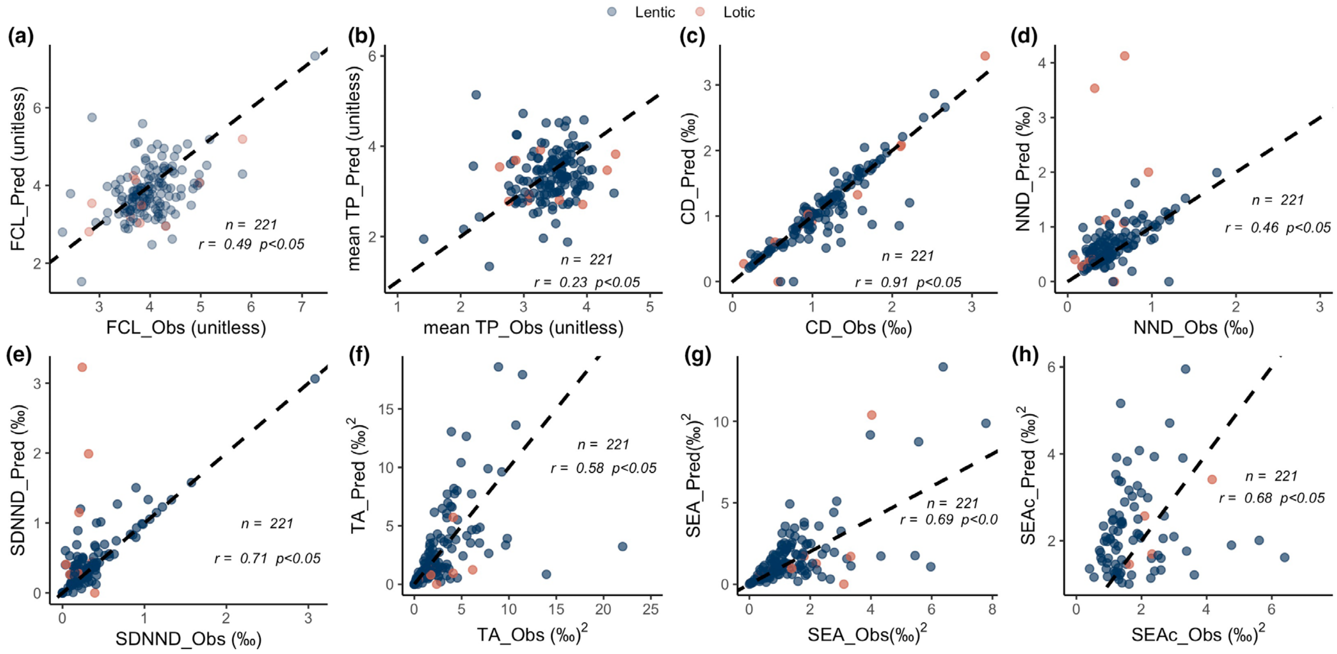

FIGURE 4 Comparison of community-wide isotope-metrics computed from measured (Obs) or simulated (Pred) baselines for both lentic (blue) and lotic (red) sites across metrics: (a) food chain length (FCL), (b) mean fish trophic position (TP), (c) centroid distance (CD), (d) mean nearest neighbour distance (NND), (e) standard deviation of the nearest neighbour distance (SDNND), (f) total Area of the convex hull (TA), (g) standard Ellipse Area (SEA) and (h) corrected standard ellipse Area (SEAc). The dotted line represents the 1:1 line. Sample size (*n*), correlation coefficient (*r*) and *p*-value of the correlation test are indicated.

global databases for water chemistry lack information on alkalinity (Filazzola et al., 2020; Naderian et al., 2024), although the water's alkalinity exerts a rather direct effect on the CO_2 concentration in the water and the $\delta^{13}C$ of the total dissolved inorganic carbon (Bade et al., 2004). A global database for inland waters' alkalinity does not yet exist, but as attempts are underway, we may hope for better accuracy of an updated $\delta^{13}C$ baseline model in the near future as this and other databases become available.

More generally, previous studies have shown the difficulties in obtaining good $\delta^{13}C$ estimates for baselines even from empirical data. Known temporal variation in short-lived organisms (Cabana & Rasmussen, 1996) leads to extreme values (Jardine et al., 2014) that do not represent the long-term integrated average baseline values that better reflect ecological reality for long-lived predators such as adult fishes in a given ecosystem (Post, 2002). Additionally, $\delta^{13}C$ in benthic organisms can vary considerably within a single site, for example, with increasing depth, especially when methane-derived carbon, produced in hypoxic patches and depths, can contribute to the benthic macroinvertebrate food web (Ayato et al., 2004; Jones et al., 2008). Not explicitly accounting for distinctively low ^{13}C-depleted methane-derived benthic endmembers may have limited the model's ability to reproduce the lowermost values of the lentic baseline $\delta^{13}C$ spectrum. Hence, the prediction model for carbon tends to underestimate the isotope difference between the pelagic and littoral habitats, as compared to the dataset of observed baselines, leading to a larger spread of benthic reliance when computed from the estimated baselines. However, benthic reliance is a metric that is rarely coherent even when computed from observed data, and indeed, herein, ~50% of the fish species $\delta^{13}C$ values fell outside of the site-specific range of the measured baselines. Some outliers in the predictions are from sites with the most extreme original baselines that the prediction model failed to reproduce, but they represented <5% of sites. At the population and community scale, the use of estimated baselines resulted in a low computation error for the trophic metrics that were the least dependent on $\delta^{13}C$ baselines, that is, trophic position and food chain length. Recomputed trophic metrics for trophic diversity and isotopic niches were sensitive to extremes, but once capped for unrealistic values, performed well when computing metrics from observed baselines. We however do not advise resorting to estimated baselines to compare benthic reliance or weighted trophic position across sites, as concordance varied between cases and species.

Besides, we recommend against using the predicted baseline values for single-system analysis or comparison at small spatial scales. At such limited scales, a high degree of accuracy is needed to ensure isotope metrics for the food web are calculated correctly, and the use of model-derived baseline values may misrepresent both short-term and long-term baselines at the site. In such cases, the key lies in matching the baseline value with the life span of the taxon of interest using locally measured baselines. For example, studies of short-lived nekton or zooplankton could be better served with a short-lived seston baseline (Heuvel et al., 2023). Such an approach would allow close tracking of the taxon of interest with the spatial and temporal variation in the baseline that can vary by season in a predictable manner (Woodland et al., 2012), or in systems with episodic nutrient inputs with variable $\delta^{15}N$ values (Jardine et al., 2014;

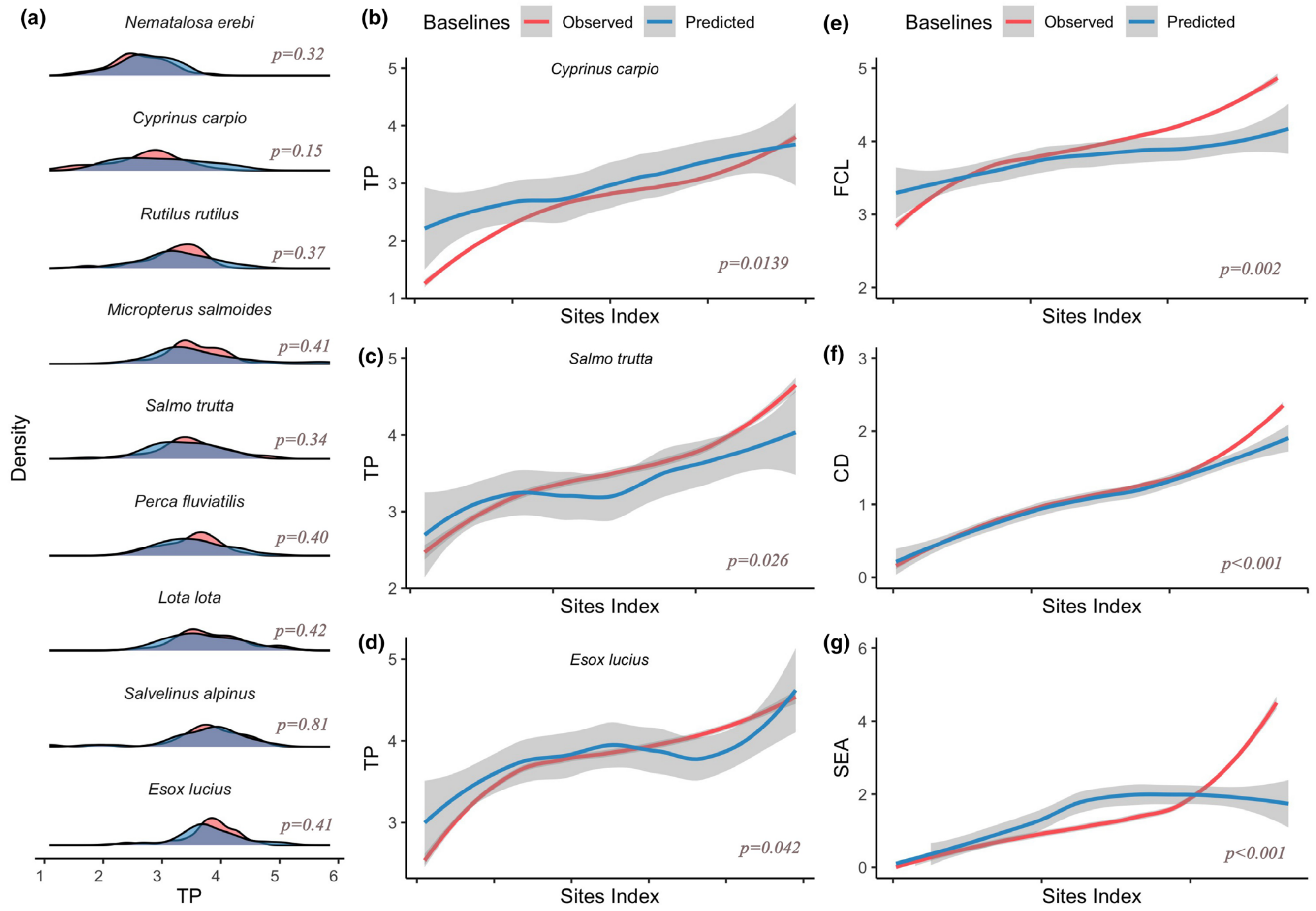

FIGURE 5 Tests of the usefulness of the isotope-metrics computed from simulated baselines at the species- (a), population- (b–d) and community-levels (e–g): (a) density distribution of trophic positions (TP) across sites for the nine most represented species in the database, computed from observed (blue) and simulated (red) baselines. *p*-values correspond to the result of the Kolmogorov–Smirnov tests for the compared distribution. (b–d) variability of trophic position (TP) for (b) *C. carpio*, (c) *S. trutta* and (d) *E. lucius* across sites, computed from observed (blue) and simulated (red) baselines. Sites are ordered according to increasing TP as computed from observed baselines and curves are local polynomial regression (LOESS) with standard error, (e–g) inter-sites variability in food chain length (e—FCL), centroid distance (f—CD) and standard ellipse area (g—SEA), computed from observed (blue) and simulated (red) baselines. Sites are ordered according to increasing FCL, CD and SEA as computed from observed baselines and curves are local polynomial regression (LOESS) with standard error. *p*-values correspond to the results of the Kendall's test of concordance. Results for all nine species and all trophic metrics are reported in Table SI-4.

Matthews & Mazumder, 2005). For local studies of long-lived taxa of interest, to avoid mismatches with the baseline, we recommend following standard practice by using long-lived primary consumers such as molluscs (Post, 2002) or employing repeated baseline sampling prior to collecting the taxon of interest (Jardine et al., 2014) and averaging or dynamically modelling baseline values (Woodland et al., 2012) rather than estimating them from our models.

With the above limitations in mind, when a high degree of precision is not required and data are noisy, our baseline predictions may hold considerable value. While original baseline measures were available for only 221 of the study sites in the ISOFRESH database, the modelled baseline values allow us to compute and estimate food web trophic metrics (unweighted TP and all community-wide metrics) at four times more sites. Our models were effective at simulating baselines over broad spatial scales. Therefore, the baseline estimator expands the geographical scale at which the variability of the (unweighted) trophic position of the species of interest can be investigated, or at which the food chain length, trophic similarity of fish species or occupied niche space can be quantified. Since we demonstrated that these metrics could be accurately recreated using the simulated baselines across a large range of ecosystems, we foresee that surrogate baselines may unleash the potential of stable isotope analyses to test food web theory at a macroecological scale.

AUTHOR CONTRIBUTIONS

Marie-Elodie Perga and Tim Jardine conceived the ideas and designed the methodology; Julien Cucherousset, Julian D. Olden and Chloé Vagnon collected the data; Marie-Elodie Perga analysed the data; Marie-Elodie Perga and Tim Jardine led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

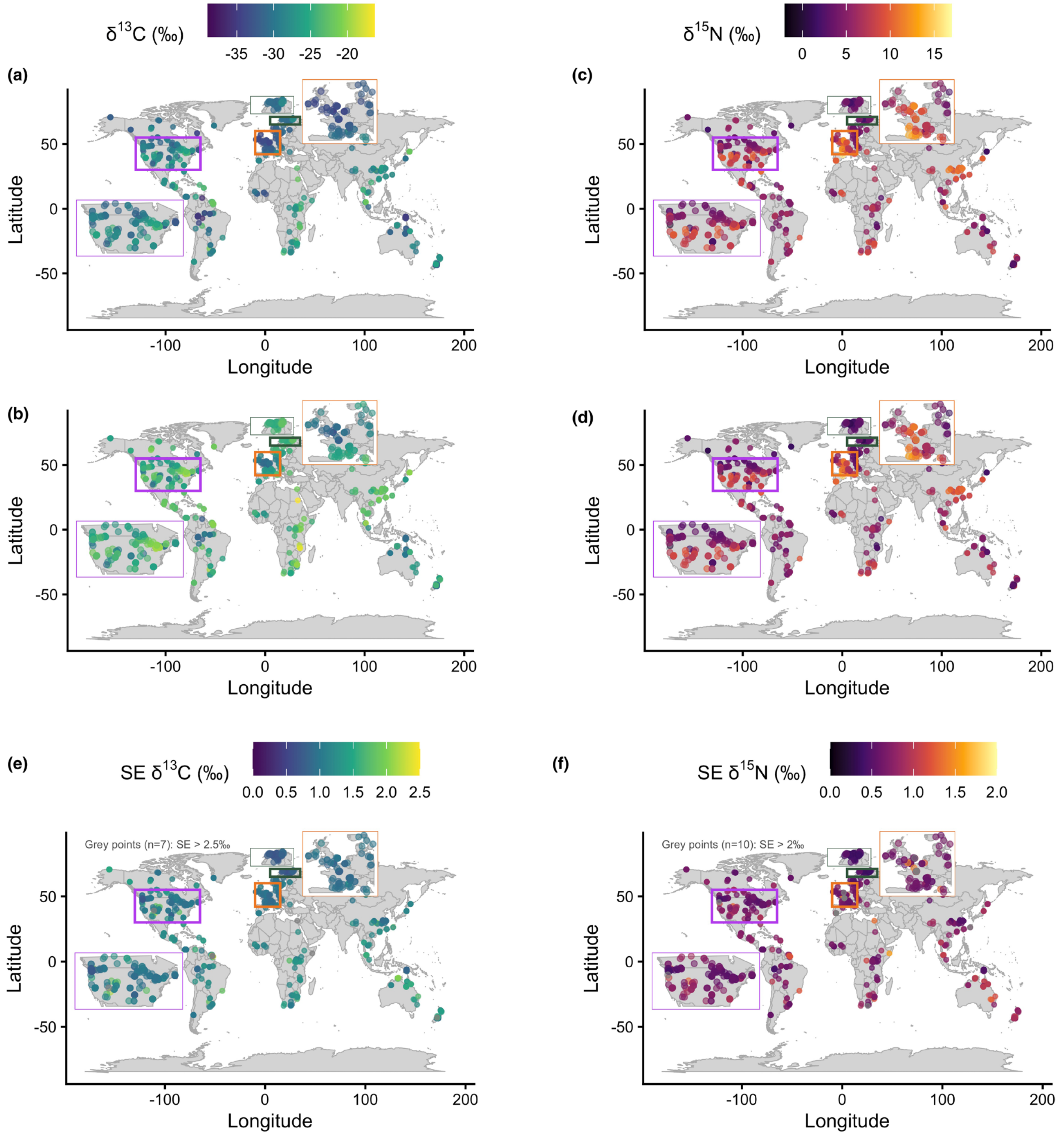

FIGURE 6 Simulated C and N stable isotope baselines for 472 lakes and 521 rivers worldwide. Simulated site-specific (a) δ^{13}C and (b) δ^{15}N values for pelagic baselines at the primary consumer level. Simulated site-specific (c) δ^{13}C and (d) δ^{15}N values for benthic baselines at the primary consumer level. Standard errors associated with site-specific (e) δ^{13}C and (f) δ^{15}N values for baseline estimates.

ACKNOWLEDGEMENTS

We thank the two anonymous reviewers and the Journal's editor for their comments on our manuscript. This research is a product of the Fooδ-webs research group funded by the Centre for the Synthesis and Analysis of Biodiversity (CESAB) of the Foundation for Research on Biodiversity (FRB). This work is part of the FUNELAKES project (grant no. IRP 00020) from INEE CNRS (Centre National de la Recherche Scientifique). M.-E.P. was funded by the SNF scientific exchange grant IZSEZ0_229270 and a fellowship from the Herbette Foundation for her sabbatical stay at The Center for Limnology at the University of Wisconsin–Madison. Open access publishing facilitated by Universite de Lausanne, as part of the Wiley - Universite de Lausanne agreement via the Consortium Of Swiss Academic Libraries.

CONFLICT OF INTEREST STATEMENT

Authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210x.70225.

DATA AVAILABILITY STATEMENT

All data and codes are available via GitHub https://doi.org/10.5281/zenodo.17718458 (Perga, 2025).

ORCID

Marie-Elodie Perga https://orcid.org/0000-0002-9003-0769
Chris Harrod https://orcid.org/0000-0002-5353-1556
Angus McIntosh https://orcid.org/0000-0003-2696-8813
Julian D. Olden https://orcid.org/0000-0003-2143-1187
Chloé Vagnon https://orcid.org/0000-0002-6201-7110

REFERENCES

Arnoldi, J. F., Bortoluzzi, J. R., Rowland, H., Harrod, C., Parnell, A. C., Payne, N., Donohue, I., & Jackson, A. L. (2024). How strongly does diet variation explain variation in isotope values of animal consumers? *PLoS One*, *19*, e0301900. https://doi.org/10.1371/journal.pone.0301900

Ayato, K., Kato, C., Iwata, T., Kishi, D., Murakami, M., Nakano, S., & Wada, E. (2004). Stream food web fueled by methane-derived carbon. *Aquatic Microbial Ecology*, *36*, 189–194.

Bade, D. L., Carpenter, S. R., Cole, J. J., Hanson, P. C., & Hesslein, R. H. (2004). Controls of δ13C-DIC in lakes: Geochemistry, lake metabolism, and morphometry. *Limnology and Oceanography*, *49*(4), 1160–1172. https://doi.org/10.4319/lo.2004.49.4.1160

Barrus, N. T., Maitland, B. M., & Rahel, F. J. (2024). Assessing a standardized method to identify optimal baselines for trophic position estimation in stable isotope studies of stream ecosystems. *Hydrobiologia*, *851*(19), 4673–4691. https://doi.org/10.1007/s10750-024-05618-y

Bodin, N., Pethybridge, H., Duffy, L. M., Lorrain, A., Allain, V., Logan, J. M., Ménard, F., Graham, B., Choy, C. A., Somes, C. J., Olson, R. J., & Young, J. W. (2021). Global data set for nitrogen and carbon stable isotopes of tunas. *Ecology*, *102*, e03265. https://doi.org/10.1002/ecy.3265

Bouletreau, 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 and Management of Aquatic Ecosystems*, *426*, 15. https://doi.org/10.1051/kmae/2025010

Bowen, G. J. (2010). Isoscapes: Spatial Pattern in Isotopic Biogeochemistry. *Annual Review of Earth and Planetary Sciences*, *38*, 161–187.

Cabana, G., & Rasmussen, J. B. (1996). Comparison of aquatic food chains using nitrogen isotopes. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 10844–10847. https://doi.org/10.1073/pnas.93.20.10844

Cadeau, P., Ader, M., Jézéquel, D., Chaduteau, C., Sarazin, G., Bernard, C., & Leboulanger, C. (2021). Nitrogen isotope discrepancy between primary producers and sediments in an anoxic and alkaline lake. *Frontiers in Earth Science*, *9*, 787386. https://doi.org/10.3389/feart.2021.787386

Casey, M. M., & Post, D. M. (2011). The problem of isotopic baseline: Reconstructing the diet and trophic position of fossil animals. *Earth-Science Reviews*, *106*, 131–148. https://doi.org/10.1016/j.earscirev.2011.02.001

De Niro, M. J., & Epstein, S. (1978). Influence of diet on distribution of carbon isotopes in animals. *Geochimica et Cosmochimica Acta*, *42*, 495–506. https://doi.org/10.1016/0016-7037(78)90199-0

De Niro, M. J., & Epstein, S. (1981). Influence of diet on the distribution of nitrogen isotopes in animals. *Geochimica et Cosmochimica Acta*, *45*, 341–351. https://doi.org/10.1016/0016-7037(81)90244-1

Diebel, M. W., & Zanden, M. J. V. (2009). Nitrogen stable isotopes in streams: Effects of agricultural sources and transformations. *Ecological Applications*, *19*, 1127–1134. https://doi.org/10.1890/08-0327.1

Döll, P., Kaspar, F., & Lehner, B. (2003). A global hydrological model for deriving water availability indicators: Model tuning and validation. *Journal of Hydrology*, *270*, 105–134. https://doi.org/10.1016/S0022-1694(02)00283-4

Espinasse, B., Sturbois, A., Basedow, S. L., Hélaouët, P., Johns, D. G., Newton, J., & Trueman, C. N. (2022). Temporal dynamics in zooplankton δ^{13}C and δ^{15}N isoscapes for the North Atlantic Ocean: Decadal cycles, seasonality, and implications for predator ecology. *Frontiers in Ecology and Evolution*, *10*, 986082. https://doi.org/10.3389/fevo.2022.986082

Filazzola, A., Mahdiyan, O., Shuvo, A., Ewins, C., Moslenko, L., Sadid, T., Blagrave, K., Imrit, M. A., Gray, D. K., Quinlan, R., O'Reilly, C. M., & Sharma, S. (2020). A database of chlorophyll and water chemistry in freshwater lakes. *Scientific Data*, *7*, 310.

Finlay, J. C. (2001). Stable-carbon-isotope ratios of river biota: Implications for energy flow in lotic food webs. *Ecology*, *82*, 1052–1064. https://doi.org/10.1890/0012-9658(2001)082[1052:SCIROR]2.0.CO;2

Finlay, J. C. (2004). Patterns and controls of lotic algal stable carbon isotope ratios. *Limnology and Oceanography*, *49*, 850–861. https://doi.org/10.4319/lo.2004.49.3.0850

France, R. L. (1995). Differentiation between littoral and pelagic food webs in lakes using stable carbon isotopes. *Limnology and Oceanography*, *40*, 1310–1313. https://doi.org/10.4319/lo.1995.40.7.1310

Goericke, R., & Fry, B. (1994). Variations of marine plankton δ^{13}C with latitude, temperature, and dissolved CO_2 in the world ocean. *Global Biogeochemical Cycles*, *8*, 85–90. https://doi.org/10.1029/93GB03272

Gu, B. (2009). Variations and controls of nitrogen stable isotopes in particulate organic matter of lakes. *Oecologia*, *160*, 421–431. https://doi.org/10.1007/s00442-009-1323-z

Gu, B., Schelske, C. L., & Waters, M. N. (2011). Patterns and controls of seasonal variability of carbon stable isotopes of particulate organic matter in lakes. *Oecologia*, *165*, 1083–1094. https://doi.org/10.1007/s00442-010-1888-6

Hastie, T., & Tibshirani, R. (1987). Generalized additive models: Some applications. *Journal of the American Statistical Association*, *82*, 371–386. https://doi.org/10.1080/01621459.1987.10478440

Hecky, R. E., & Hesslein, R. H. (1995). Contributions of benthic algae to lake food webs as revealed by stable isotope analysis. *Journal of the North American Benthological Society*, *14*, 631–653. https://doi.org/10.2307/1467546

Hengl, T., de Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B. M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J. G. B., Walsh, M. G., & Gonzalez, M. R. (2014). SoilGrids1km—Global soil information based on automated mapping. *PLoS One*, *9*, e105992. https://doi.org/10.1371/journal.pone.0105992

Heuvel, C. E., Zhao, Y., & Fisk, A. T. (2023). Influence of spatial and temporal variation on establishing stable isotope baselines of δN, δC, and δS in a large freshwater lake. *Freshwater Biology*, *68*, 806–820. https://doi.org/10.1111/fwb.14065

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978. https://doi.org/10.1002/joc.1276

Hixon, S. W., Fernandes, R., Andriamahaihavana, A., Baden, A. L., Blanco, M. B., Caulier, G., Dammhahn, M., Eeckhaut, I., Eppley, T. M., Frédérich, B., Ganzhorn, J. U., Garbaras, A., Gibson, D., Goodman, S. M., Irwin, M., Kelley, E. A., Michel, L. N., Lepoint, G., Loudon, J. E., … Crowley, B. E. (2024). Introducing IsoMad, a compilation of isotopic datasets for Madagascar. *Scientific Data*, *11*, 857. https://doi.org/10.1038/s41597-024-03705-2

Hoeinghaus, D. J., & Zeug, S. C. (2008). Can stable isotope ratios provide for community-wide measures of trophic structure? Comment. *Ecology*, *89*, 2353–2357. https://doi.org/10.1890/07-1143.1

Jackson, A. L., Inger, R., Parnell, A. C., & Bearhop, S. (2011). Comparing isotopic niche widths among and within communities: SIBER—Stable Isotope Bayesian Ellipses in R. *The Journal of Animal Ecology*, *80*, 595–602. https://doi.org/10.1111/j.1365-2656.2011.01806.x

Jardine, T. D., Hadwen, W. L., Hamilton, S. K., Hladyz, S., Mitrovic, S. M., Kidd, K. A., Tsoi, W. Y., Spears, M., Westhorpe, D. P., Fry, V. M., Sheldon, F., & Bunn, S. E. (2014). Understanding and overcoming baseline isotopic variability in running waters. *River Research and Applications*, *30*, 155–165. https://doi.org/10.1002/rra.2630

Jepsen, D. B., & Winemiller, K. O. (2002). Structure of tropical river food webs revealed by stable isotope ratios. *Oikos*, *96*(1), 46–55.

Jones, R. I., Carter, C. E., Kelly, A., Ward, S., Kelly, D. J., & Grey, J. (2008). Widespread contribution of methane-cycle bacteria to the diets of lake profundal chironomid larvae. *Ecology*, *89*, 857–864.

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*, 24. https://doi.org/10.1002/ecm.1451

Kling, G. W., Fry, B., & Obrien, W. J. (1992). Stable isotopes and planktonic trophic structure in arctic lakes. *Ecology*, *73*, 561–566. https://doi.org/10.2307/1940762

Kristensen, P. B., Riis, T., Dylmer, H. E., Kristensen, E. A., Meerhoff, M., Olesen, B., de Teixeira- Mello, F., Baattrup-Pedersen, A., Cavalli, G., & Jeppesen, E. (2016). Baseline identification in stable -isotope studies of temperate lotic systems and implications for calculated trophic positions. *Freshwater Science*, *35*, 909–921. https://doi.org/10.1086/687284

Laws, E. A., Bidigare, R. R., & Popp, B. N. (1997). Effect of growth rate and CO_2 concentration on carbon isotopic fractionation by the marine diatom *Phaeodactylum tricornutum*. *Limnology and Oceanography*, *42*, 1552–1560. https://doi.org/10.4319/lo.1997.42.7.1552

Layman, C. A., Araujo, M. S., Boucek, R., Hammerschlag-Peyer, C. M., Harrison, E., Jud, Z. R., Matich, P., Rosenblatt, A. E., Vaudo, J. J., Yeager, L. A., Post, D. M., & Bearhop, S. (2012). Applying stable isotopes to examine food-web structure: An overview of analytical tools. *Biological Reviews*, *87*, 545–562. https://doi.org/10.1111/j.1469-185X.2011.00208.x

Layman, C. A., Arrington, D. A., Montaña, C. G., & Post, D. M. (2007). Can stable isotope ratios provide for community-wide measures of trophic structure? *Ecology*, *88*, 42–48. https://doi.org/10.1890/0012-9658(2007)88[42:Csirpf]2.0.Co;2

Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, *27*, 2171–2186. https://doi.org/10.1002/hyp.9740

Lehner, B., Messager, M. L., Korver, M. C., & Linke, S. (2022). Global hydro-environmental lake characteristics at high spatial resolution. *Scientific Data*, *9*, 351. https://doi.org/10.1038/s41597-022-01425-z

Lejeune, B., Lepoint, G., & Denoël, M. (2024). Food web collapse and regime shift following goldfish introduction in permanent ponds. *Global Change Biology*, *30*, 14. https://doi.org/10.1111/gcb.17435

Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, F., & Thieme, M. (2019). Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. *Scientific Data*, *6*, 283. https://doi.org/10.1038/s41597-019-0300-6

Magozzi, S., Yool, A., Zanden, H. B. V., Wunder, M. B., & Trueman, C. N. (2017). Using ocean models to predict spatial and temporal variation in marine carbon isotopes. *Ecosphere*, *8*, 20. https://doi.org/10.1002/ecs2.1763

Majdi, N., Hette-Tronquart, N., Auclair, E., Bec, A., Chouvelon, T., Cognie, B., Danger, M., Decottignies, P., Dessier, A., Desvilettes, C., Dubois, S., Dupuy, C., Fritsch, C., Gaucherel, C., Hedde, M., Jabot, F., Lefebvre, S., Marzloff, M. P., Pey, B., … Perga, M.-E. (2018). There's no harm in having too much: A comprehensive toolbox of methods in trophic ecology. *Food Webs*, *17*, e00100. https://doi.org/10.1016/j.fooweb.2018.e00100

Matthews, B., & Mazumder, A. (2005). Consequences of large temporal variability of zooplankton δ^{15}N for modeling fish trophic position and variation. *Limnology and Oceanography*, *50*(5), 1404–1414.

Mullineaux, S. T., Kostka, B., Rock, L., Ogle, N., Marks, N. J., Doherty, R., Harrod, C., Montgomery, W. I., & Scantlebury, D. M. (2022). Soil δ^{13}C and δ^{15}N baselines clarify biogeographic heterogeneity in isotopic discrimination of European badgers (*Meles meles*). *Scientific Reports*, *12*, 200. https://doi.org/10.1038/s41598-021-04011-2

Naderian, D., Noori, R., Heggy, E., Bateni, S. M., Bhattarai, R., Nohegar, A., & Sharma, S. (2024). A water quality database for global lakes. *Resources, Conservation and Recycling*, *202*, 107401.

Olsson, K., Stenroth, P., NystrÖM, P. E. R., & GranÉLi, W. (2009). Invasions and niche width: Does niche width of an introduced crayfish differ from a native crayfish? *Freshwater Biology*, *54*, 1731–1740. https://doi.org/10.1111/j.1365-2427.2009.02221.x

Perga, M.-E. (2025). mepLAKES/FoodWeb_baselines: A global estimator of C and N isotope baselines for fresh waters (V1.0_isotope_baseline_estimator). *Zenodo*. https://doi.org/10.5281/zenodo.17718459

Perga, M.-E., & Gerdeaux, D. (2004). Changes in the δ^{13}C of pelagic food webs: The influence of lake area and trophic status on the isotopic signature of whitefish (*Coregonus lavaretus*). *Canadian Journal of Fisheries and Aquatic Sciences*, *61*, 1485–1492. https://doi.org/10.1139/f04-089

Post, D. M. (2002). Using stable isotopes to estimate trophic position: Models, methods, and assumptions. *Ecology*, *83*, 703–718. https://doi.org/10.2307/3071875

Post, D. M., Pace, M. L., & Hairston, N. G. (2000). Ecosystem size determines food-chain length in lakes. *Nature*, *405*, 1047–1049. https://doi.org/10.1038/35016565

Ramsay, T. O., Burnett, R. T., & Krewski, D. (2003). The effect of concurvity in generalized additive models linking mortality to ambient particulate matter. *Epidemiology*, *14*, 18–23.

Ribot, M., von Schiller, D., Peipoch, M., Sabater, F., Grimm, N. B., & Martí, E. (2013). Influence of nitrate and ammonium availability on uptake kinetics of stream biofilms. *Freshwater Science*, *32*, 1155–1167. https://doi.org/10.1899/12-209.1

St John Glew, K., Espinasse, B., Hunt, B. P. V., Pakhomov, E. A., Bury, S. J., Pinkerton, M., Nodder, S. D., Gutiérrez-Rodríguez, A., Safi, K., Brown, J. C. S., Graham, L., Dunbar, R. B., Mucciarone, D. A., Magozzi, S., Somes, C., & Trueman, C. N. (2021). Isoscape models of the southern ocean: Predicting spatial and temporal variability in carbon and nitrogen isotope compositions of particulate organic matter. *Global Biogeochemical Cycles*, *35*, e2020GB006901. https://doi.org/10.1029/2020GB006901

Stiling, R. R., Olden, J. D., Boulêtreau, S., Cucherousset, J., & Holtgrieve, G. W. (2023). Global investigation of lake habitat coupling by fishes. *Oecologia*, *202*, 12. https://doi.org/10.1007/s00442-023-05424-8

Trueman, C. N., MacKenzie, K. M., & St John Glew, K. (2017). Stable isotope-based location in a shelf sea setting: Accuracy and precision are comparable to light-based location methods. *Methods in Ecology and Evolution*, *8*, 232–240. https://doi.org/10.1111/2041-210X.12651

Vander Zanden, M. J., Casselman, J. M., & Rasmussen, J. B. (1999). Stable isotope evidence for the food web consequences of species invasions in lakes. *Nature*, *401*, 464–467. https://doi.org/10.1038/46762

Vander Zanden, M. J., & Fetzer, W. W. (2007). Global patterns of aquatic food chain length. *Oikos*, *116*, 1378–1388. https://doi.org/10.1111/j.2007.0030-1299.16036.x

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*, 1395–1404.

Vander Zanden, M. J., Vadeboncoeur, Y., Diebel, M. W., & Jeppesen, E. (2005). Primary consumer stable nitrogen isotopes as indicators of nutrient source. *Environmental Science & Technology*, *39*, 7509–7515. https://doi.org/10.1021/es050606t

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., Laurance, W. F., Wood, P., Fekete, B. M., Levy, M. A., & Watson, J. E. M. (2016). Global terrestrial Human Footprint maps for 1993 and 2009. *Scientific Data*, *3*, 160067. https://doi.org/10.1038/sdata.2016.67

Wainright, C. A., Muhlfeld, C. C., Elser, J. J., Bourret, S. L., & Devlin, S. P. (2021). Species invasion progressively disrupts the trophic structure of native food webs. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(5), e2102179118. https://doi.org/10.1073/pnas.2102179118

Wollheim, W. M., Harms, T. K., Robison, A. L., Koenig, L. E., Helton, A. M., Song, C., Bowden, W. B., & Finlay, J. C. (2022). Superlinear scaling of riverine biogeochemical function with watershed size. *Nature Communications*, *13*, 1230. https://doi.org/10.1038/s41467-022-28630-z

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *73*, 3–36.

Woodland, R. J., Rodríguez, M. A., Magnan, P., Glémet, H., & Cabana, G. (2012). Incorporating temporally dynamic baselines in isotopic mixing models. *Ecology*, *93*(1), 131–144.

Yoshikawa, C., Shigemitsu, M., Yamamoto, A., Oka, A., & Ohkouchi, N. (2024). A nitrogen isoscape of phytoplankton in the western North Pacific created with a marine nitrogen isotope model. *Frontiers in Marine Science*, *11*, 1294608. https://doi.org/10.3389/fmars.2024.1294608

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Table S1. Summary of models' abbreviations and variable names.

Figure S2. Diagnostic residuals plots for the fitted and selected GAM models.

Figure S3. Smooth functions of the selected and fitted GAM for baselines prediction.

Table S4. Relevance of simulated baselines for macro-ecological studies.

Table S4-1. Compared distribution and concordance for population-level isotope metrics computed from observed and predicted baselines.

Table S4-2. Compared concordance for community-level isotope metrics computed from observed and predicted baselines.

How to cite this article: Perga, M.-E., Bouletreau, S., Cucherousset, J., Harrod, C., McIntosh, A., Olden, J. D., Vagnon, C., & Jardine, T. (2026). A global estimator of C and N isotope baselines for fresh waters. *Methods in Ecology and Evolution*, *17*, 992–1006. https://doi.org/10.1111/2041-210x.70225