Received: 17 August 2023 | Accepted: 21 November 2023

DOI: 10.1111/1365-2656.14044

Journal of Animal Ecology | BRITISH ECOLOGICAL SOCIETY

RESEARCH METHODS GUIDE

Maximum likelihood outperforms binning methods for detecting differences in abundance size spectra across environmental gradients

Justin Pomeranz[1] | James R. Junker[2,3] | Vojsava Gjoni[4] | Jeff S. Wesner[4]

[1]Colorado Mesa University, Grand Junction, Colorado, USA

[2]Great Lakes Research Center, Michigan Technological University, Houghton, Michigan, USA

[3]Louisiana Universities Marine Consortium, Chauvin, Louisiana, USA

[4]Department of Biology, University of South Dakota, Vermillion, South Dakota, USA

Correspondence
Justin Pomeranz
Email: jfpomeranz@gmail.com

Handling Editor: Julien Cucherousset

Abstract

1. Individual body size distributions (ISD) within communities are remarkably consistent across habitats and spatiotemporal scales and can be represented by size spectra, which are described by a power law. The focus of size spectra analysis is to estimate the exponent (λ) of the power law. A common application of size spectra studies is to detect anthropogenic pressures.
2. Many methods have been proposed for estimating λ most of which involve binning the data, counting the abundance within bins, and then fitting an ordinary least squares regression in log–log space. However, recent work has shown that binning procedures return biased estimates of λ compared to procedures that directly estimate λ using maximum likelihood estimation (MLE). While it is clear that MLE produces less biased estimates of site-specific λ's, it is less clear how this bias affects the ability to test for changes in λ across space and time, a common question in the ecological literature.
3. Here, we used simulation to compare the ability of two normalised binning methods (equal logarithmic and $\log_2$ bins) and MLE to (1) recapture known values of λ, and (2) recapture parameters in a linear regression measuring the change in λ across a hypothetical environmental gradient. We also compared the methods using two previously published body size datasets across a natural temperature gradient and an anthropogenic pollution gradient.
4. Maximum likelihood methods always performed better than common binning methods, which demonstrated consistent bias depending on the simulated values of λ. This bias carried over to the regressions, which were more accurate when λ was estimated using MLE compared to the binning procedures. Additionally, the variance in estimates using MLE methods is markedly reduced when compared to binning methods.
5. The error induced by binning methods can be of similar magnitudes as the variation previously published in experimental and observational studies, bringing into question the effect sizes of previously published results. However, while the methods produced different regression slope estimates, they were in qualitative agreement on the sign of those slopes (i.e. all negative or all positive). Our results provide

further support for the direct estimation of λ and its relative variation across environmental gradients using MLE over the more common methods of binning.

KEYWORDS
body size-abundance relationships, community biomass distributions, individual size distributions, logarithmic binning methods, maximum likelihood, size spectra, statistical methods

1 | INTRODUCTION

Body size distributions are a fundamental characteristic of communities. In general, abundance declines with increasing body size, and this is thought to be a consequence of simple size-dependent metabolic constraints on organisms' energy use predicted by the metabolic theory of ecology (Brown et al., 2004; Nee et al., 1991). The remarkable consistency of these relationships across spatiotemporal scales and ecosystems has led them to be recommended as a "universal" indicator of ecological status (Petchey & Belgrano, 2010). Variation in size-abundance relationships have been documented through space (Pomeranz et al., 2022), time (Evans et al., 2022; McGarvey & Kirk, 2018) and in response to human activities (Jennings & Blanchard, 2004; Martínez et al., 2016; Pomeranz et al., 2018). Likewise, variation in size-abundance relationships have been used to explain fundamental differences in how communities are organised. For example, external resource subsidies "bend the rules" and allow higher abundances of large body sizes than would be expected based on metabolic theory (Perkins et al., 2018, 2021). However, recent research has shown that these results may be an artefact of how the data were treated. Edwards et al. (2020) analysed a time series of marine fisheries data and found that the parameter explaining the relationship was either invariant, or that it changed consistently through time depending on the methodology used.

Individual size distributions (ISD sensu White et al., 2007), also referred to as abundance size spectra, are one of the size-abundance relationships commonly used. Generally, there is a negative relationship between individual body size (m, measured in mass) on the x-axis and abundance (N) on the y-axis. Theoretical and empirical data support this relationship being described as a simple power law with exponent λ in the form of

$$N \sim m^{\lambda} \tag{1}$$

(Andersen & Beyer, 2006; Sheldon & Kerr, 1972). ISDs represent frequency distributions of body sizes within a community. Specifically, let m be a random variable of body sizes described by the probability density function:

$$f(m) = Cm^{\lambda}, m_{\min} \le m \le m_{\max} \tag{2}$$

where

$$C = \begin{cases} \dfrac{\lambda+1}{m_{\max}{}^{\lambda+1} - m_{\min}{}^{\lambda+1}}, & \lambda \neq -1 \\ \dfrac{1}{\log m_{\max} - \log m_{\min}}, & \lambda = -1 \end{cases} \tag{3}$$

and where m is body mass in milligrams, λ is the exponent describing the power law distribution bounded by the minimum ($m_{\min}$) and maximum ($m_{\max}$) body sizes in the data (Edwards et al., 2017).

The primary goal of ISD analyses is to estimate λ, and ecologists have devised multiple methods for doing so. Commonly, N_i is the count of body sizes in bins i, where bin i has midpoint $m_{\text{mid},i}$, and λ is estimated as the slope from ordinary least squares (OLS) regressions in log–log space (commonly $\log_{10}$) as:

$$\log_{10} N_i = \beta_0 + \lambda \log_{10} m_{\text{mid},i} + \varepsilon \tag{4}$$

where N_i is the count in bin i, $m_{\text{mid},i}$ is the mid-point of bin i, λ is the parameter describing the decline in abundance (estimate of the power law exponent), β_0 is the intercept, and ε is the error term.

Myriad binning methods have been proposed, including different bin widths on linear and logarithmic scales. Likewise, some methods rely on the absolute counts in the bins (referred to here as "abundance spectra") and others use normalization techniques (referred to here as "normalized abundance spectra") such as dividing the count by the bin width. To further complicate matters, the total biomass in a body size bin can be summed to estimate the biomass spectrum or the normalised (when the total biomass is divided by the bin width) biomass spectrum. The common feature that ties binning methods together is data reduction, in which all the variation in individual sizes within bins is removed by assigning each individual to a single body size (like the midpoint of a bin between 10 and 100 mg). As an alternative to binning methods, λ can be estimated directly on un-binned data using maximum likelihood estimation (MLE; see Sprules & Barth, 2016 for a review on size spectra methods). The major advantage of using MLE is that it does not require any binning and hence no abundance estimates. Instead, it uses only the individual body sizes provided in the data, consistent with theoretical expectations of ISD (Edwards et al., 2017).

Previous work has shown that the estimates of λ differ between MLE and size-binned OLS techniques (Edwards et al., 2017, 2020; White et al., 2008). Size-binned OLS methods are particularly sensitive to decisions made in the binning process including the number, width, and beginning and ending "edges of the bins" (Edwards et al., 2020; White et al., 2008). Simulation studies have shown that MLE offers consistently more accurate estimates of λ (Edwards et al., 2017, 2020; White et al., 2008), and reanalysis of empirical data also indicates that the conclusions are dependent on the method used (White et al., 2008). For example, White et al. (2008) reanalyzed the data of Enquist and Niklas (2001) and Meehan (2006) using MLE methods. The original publications supported theoretically expected quantities, whereas the less biased MLE methods produced estimates which deviated from those expected by theory. Likewise, Edwards et al. (2020) analysed a time-series of individual body sizes of demersal fish from the North Sea using different methods. Depending on the method

used, they either found a decrease (steepening) of λ or invariance of λ through time (fig. 1 in Edwards et al., 2020). Although not explicitly discussed in either White et al. (2008) or Edwards et al. (2020), it is not known what the "true" relationship of the empirical dataset was. For instance, the differences detected between the methods in the marine fish community through time could have drastic implications for what future management recommendations may be made. Steeper size-abundance relationships imply that the abundance of large fish is declining (or the abundance of small fish is increasing). If this were the conclusion reached, it seems reasonable that managers might recommend reductions in fish catch quotas to allow the community to recover. Alternatively, if the size-abundance relationship is invariant through time, it would seem reasonable to not make any changes to fisheries decisions based solely on the size-abundance relationship. Furthermore, it may imply that the biological communities organise themselves to have a consistent size-abundance relationship, even when human impacts (i.e. fishing) is present. This is particularly intriguing given that seminal analyses of marine fish communities have had profound impacts on the establishment, development, application and interpretation of size-abundance relationships in ecology (Jennings & Blanchard, 2004).

While there is a growing consensus that MLE methods offer more reliable estimates of λ than binning methods, it remains unclear if these biases are consistent and systematic or stochastic, and whether the relative change in ISD parameters is consistent across space and time. In other words, if the data within a study are all treated the same, does a relative change of size-binned OLS slope of 0.1 coincide with a relative change of MLE λ estimates of 0.1?

We had three primary objectives in this study: (1) to compare how well different methods estimate site-specific λ's, (2) recapture parameters in a linear regression measuring the change in λ across a hypothetical environmental gradient and (3) to see if the conclusion reached on empirical datasets were dependent on the different methods used. Objective 1 extends work by Edwards et al. (2017, 2022) and compares MLE methods to two common logarithmic binning methods for constructing normalised abundance spectra. Objective 2 is a novel simulation exercise to make recommendations for detecting differences in ISD relationships in future studies. We find that MLE provides more accurate estimates of site-specific λ values, as well as recapturing relative changes in λ values across a hypothetical gradient. We recommend that future work uses MLE methods to fit size-abundance relationships.

2 | METHODS

2.1 | Data simulation

To investigate the performance of commonly used methods, we simulate body size observations from a bounded power law distribution using the `rPLB()` function in the *sizeSpectra* package (Edwards, 2020) for the R statistical language (version 4.0.3, R Core Team, 2020), as described in (Edwards et al., 2017). Given known values of λ, m_{min} and m_{max}, the `rPLB()` function generates random body sizes (m) as described in Edwards et al. (2017). For all simulations, we set $m_{min} = 0.0026$ and $m_{max} = 1.2 * 10^3$. These values were based on empirical body sizes of stream benthic communities reported in (Pomeranz et al., 2022). In a review of size spectrum methods, Sprules and Barth (2016) indicate that the results of analyses may depend on the range of body sizes present in the data (i.e. partial community, such as zooplankton or fish, compared with a community including body sizes from zooplankton to fish). Our results were not dependent on the range of body sizes (Supporting Information). No ethical approval was required for this study.

2.2 | Experiment 1: Site-specific λ estimates

Using the procedure above, we independently sampled $n = 999$ body sizes from nine different λ's: (−0.50, −0.75, −1.00, −1.25, −1.50, −1.75, −2.00, −2.25, −2.50). The values of λ describe how quickly the abundance of large body sizes decline within a community. For example, a value of −0.5 means there would be a relatively high number of large body sizes (shallow decline) whereas a value of −2.5 means there would be relatively very few large body sizes (steep decline). For each value of λ, we repeated the process 1000 times (reps), resulting in 9000 (9 λ's * 1000 reps) estimates.

2.3 | Estimation of ISD parameter λ

After simulating data, we used three different methods (described below) to estimate the value of λ (maximum likelihood, equal logarithmic bins normalised [ELBn] and $\log_2$ bins normalised [L2n]) and plotted the distribution of estimated values obtained for each method against the known value of λ.

2.4 | Maximum likelihood estimation

The MLE is a method for estimating parameters of an assumed probability function directly by maximizing a likelihood function. MLE directly estimates λ by finding the value of λ which maximises the likelihood function based on the specific data analysed (see Edwards et al., 2017). Edwards et al. (2020) provided an R package called `sizeSpectra` which has MLE methods and tutorials on how to apply them to datasets. Here, we modified MLE functions from the `sizeSpectra` package for the R language to estimate λ (Edwards, 2020). Throughout the manuscript, these estimates are referred to as MLE.

2.5 | Equal logarithmic bins normalised: ELBn

For the first binning method, we created six equal logarithmic bins covering the range of body sizes. This method (without normalization)

has been used extensively in previous studies (Table 1). It is important to note that the number of bins is set a priori, and the widths of the bins is determined based on the range of the body sizes present in the data. For example, with six equal logarithmic bins, if the data ranged from 1 to 100 g the first bin would be from 1.0 to 2.3 g, and the final bin would be from 64.5 to 100 g, whereas if the data ranged from 1 to 1000 g the first bin would be from 1.0 to 3.2 g and the final bin would be from 341.5 to 1096 g (Figure S1 in supplemental information). Here, we set the body size range to be the same for all simulations, so the widths of the bins do not vary. The count in each bin was normalised by dividing by the bin width to account for the unequal bin sizes. Normalization corrects for the distortion caused by logarithmic bins and generally improves the linear fits of OLS (Sprules & Barth, 2016; White et al., 2008). Although previous publications using this method generally do not use normalization, the process of normalizing shifts the OLS estimate of λ by −1. In other words, an un-normalised OLS estimate of −0.75 would result in an OLS estimate of −1.75 when normalizing the data (Edwards et al., 2017; Pomeranz et al., 2022; Sprules & Barth, 2016). Throughout the manuscript, the normalised equal logarithmic binning method will be referred to as ELBn which was not tested in Edwards et al. (2017).

2.6 | Log_2 bins normalised: L2n

The second binning method was similar to ELBn but bins of equal with on a log2 scale were used, where the width of each bin is twice that of the previous one. When working with empirical data with different size ranges, this can alter the number of bins per site which is known to alter parameter estimates (Sprules & Barth, 2016; White et al., 2008). However, since the data here were simulated from a known size range, the number of bins for each site is identical. Essentially, the ELBn method sets the number of bins and the width varies based on the data, whereas the L2n method sets the bin widths and the number of bins varies based on the data. The count in each bin is normalised in the same way as described above for the ELBn approach. Log_2 bins have been used extensively in the literature to construct biomass and abundance spectrum, both normalised and un-normalised (Table 1). The L2n method is like the LBNbiom method in Edwards et al., 2017 except in the present study the count in each bin is used as opposed to the sum of the total biomass in each bin.

As mentioned above, normalization consistently shifts the OLS estimate of $\lambda - 1$ when compared with OLS estimates from un-normalised counts. It is also worth noting the relationship between abundance (the focus of the present study) and biomass spectrum (commonly used in studies of marine systems). OLS estimates of the size spectra exponent when using un-normalised data is actually estimating $\lambda + 1$ (hence, it is necessary to subtract 1 from the OLS estimate to calculate λ; Edwards et al., 2017; Sprules & Barth, 2016; White et al., 2008), whereas OLS using normalized abundance spectrum is estimating λ. Likewise, OLS estimates of the exponent when using biomass are actually estimating $\lambda + 2$, and when using normalised biomass are estimating $\lambda + 1$. Although we do not test the biomass spectrum here directly, our conclusions apply to studies of biomass spectra after accounting for the shift in estimates from abundance to biomass relationships and accounting for normalization (if applicable). After processing the simulated data through the ELBn and L2n binning procedures, λ was estimated using simple OLS regression (Equation 4), which is directly comparable with the MLE estimates.

TABLE 1 Selected citations demonstrating the use of the two binning methods (or variations) assessed here. ELB refers to equal logarithmic bins, and the number of bins used in the study is indicated in parentheses. The studies are organised by whether they used abundance (N, count of individuals in a bin) or biomass (B, sum of individuals in a bin), and whether the results were normalised (sum or count in a bin divided by bin width) or presented un-normalised (raw count or sum in a bin). See the main text for a discussion on converting estimated exponent values from abundance to biomass, and from un-normalised to normalised.

Authors (Year)	Abundance (N) or biomass(B)	Bin size	Normalised?
Maxwell and Jennings (2006)	B	log_2	No
Jennings and Blanchard (2004)	B	log_2	No
Jennings et al. (2002)	B	log_2	No
Gaedke et al. (2004)	B[a]	log_2	Yes
Mehner et al. (2018)	B[a]	log_2	Yes
Mazurkiewicz et al. (2020)	B[a]	log_2	Yes
McGarvey and Kirk (2018)	N[b]	log_2	Yes
Fraley et al. (2018)	N	log_2	No
Pomeranz et al. (2019a)	N	log_2	Yes
Chang et al. (2014)	N	log_2	Yes
Martínez et al. (2016)	N	ELB (bin $N = 6$)	No
Yvon-Durocher et al. (2011)	N	ELB (bin $N = 10$)	No
Perkins et al. (2018)	N	ELB (bin $N = 6$)	No
Dossena et al. (2012)	N	ELB (bin $N = 6$)	No
Perkins et al. (2021)	N	ELB (bin $N = 6$)	No

[a]These references use the LBNbiom method as described in Edwards et al., 2017.

[b]McGarvey and Kirk present the results as $D \sim M$, where D is the number of individuals per m^{-2}.

2.7 | Experiment 2: Variation in λ across a hypothetical environmental gradient

A common application of size spectra analyses is to test for changes in λ across some sort of gradient (i.e. anthropogenic pollution, Pomeranz et al., 2019a; resource subsides, Perkins et al., 2018; time Edwards et al., 2020; environmental temperature, Pomeranz et al., 2022). However, it is unknown how biases in site-specific estimates of λ "scale-up" and potentially alter the conclusions reached for changes in λ across gradients. The focus of experiment 2 was to investigate the ability of the three methods to recapture known changes in λ across a hypothetical environmental gradient (Figure 1). To do this, we set λ to vary at a known rate (β) across a hypothetical gradient (x) and then sampled body sizes from bounded power laws described by λ. We then used the three methods (ELBn, L2n, and MLE) to estimate site-specific λ at each point across the gradient, and then performed OLS regressions to estimate β. We conducted this process 1000 times (replicates) to get a distribution of estimated β's and compared this distribution to the known value of β (see Figure 1).

Experiment 2 had two parts: (2.1) Assessing how the bias of site-specific λ estimates influences our estimates of β (the change in λ across a hypothetical environment) depending on where λ falls in parameter space (λ from −2.5 to −0.5); (2.2) How does the magnitude of β (0, 0.25, 0.5) across the gradient influence our estimates of β?

2.8 | Experiment 2.1: λ "scenarios"

To test how variation in site-specific λ "scale-up" across gradients, we performed OLS regression analysis based on simulations according to the following equation:

$$\lambda_{jkl} = \beta_{0,kl} + \beta_{\text{env},kl} x_j + \varepsilon_{kl} \quad (5)$$

where λ_{jkl} is the estimated parameter from site j, replicate k and method l, x_j is the environmental value at site j, $\beta_{0,kl}$ and $\beta_{\text{env},kl}$ are the regression intercept and slope coefficients, for replicate k, and method l, respectively, and ε_{kl} is the error term. The distribution of $\beta_{\text{env},kl}$ estimates across simulation were then plotted against the known value of β_{env}.

For each replicate ($k = 1, \ldots, 1000$), we had five sites (j) uniformly spaced across a hypothetical environment (x) with values of x between −1 and 1 (i.e. values of x_j were −1.0, −0.5, 0, 0.5, 1.0). This can be compared to standardizing values of an environmental value

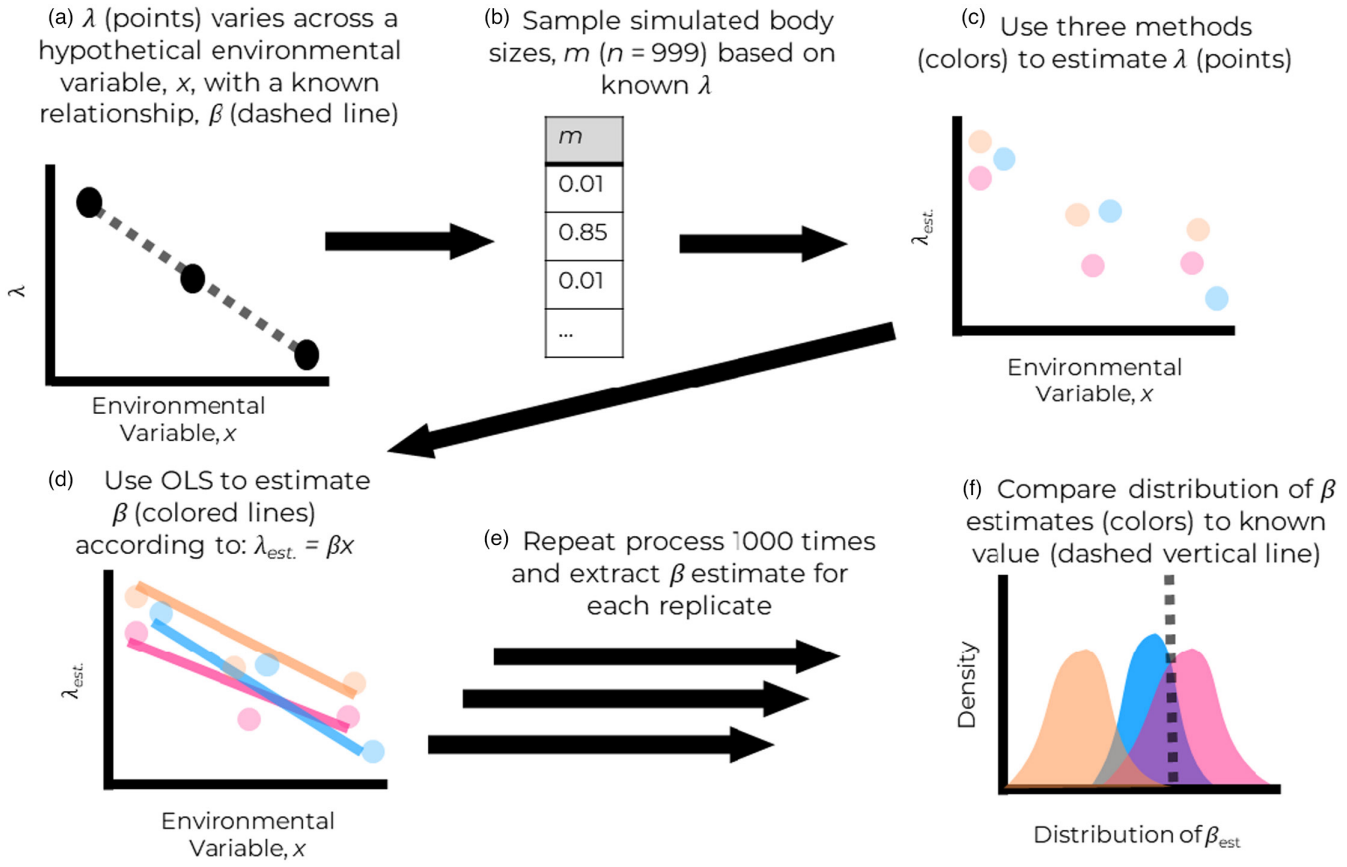

FIGURE 1 Conceptual figure of the simulation procedure used in experiment 2. (a) we set λ (points) to vary at a known relationship (β, dashed line). (b) Using known values of λ, we sampled 999 body sizes for each site. (c) We estimated site-specific λ using the three methods (ELbn, L2n, MLE, see main text) and plotted them across the hypothetical environmental gradient (points, coloured by method). (d) We then estimated β using OLS regressions for each method separately (coloured lines). (e) We repeated a–d 1000 times to get a distribution of β estimates. (f) We compared the distribution of β estimates (coloured density plots) with the known values of β (dashed vertical line).

TABLE 2 Variables for experiment 2.1 The environmental value, *x* are the values for the hypothetical environmental gradient. The subsequent columns represent the corresponding λ values across three different "scenarios". For each scenario $\beta_{env}=0.5$.

Environmental value, *x*	Steep λ	Medium λ	Shallow λ
−1	−2.5	−2	−1.5
−0.5	−2.25	−1.75	−1.25
0	−2	−1.5	−1
0.5	−1.75	−1.25	−0.75
1	−1.5	−1	−0.5

(i.e. temperature) across a gradient with *z*-scores. Each site (x_j) had a corresponding value of λ based on the λ scenario (steep, medium and shallow, Table 2). In the "steep" scenario λ ranged from −2.5 to −1.5, compared to −2.0 to −1 in the "medium" scenario and from −1.5 to −0.5 in the "shallow" scenario. For all scenarios in experiment 2.1, the relationship of λ across the environment, *x*, was set as $\beta=0.5$.

For each site *j*, and replicate *k* we independently sampled 999 body-size observations from a bounded power law distribution described by λ and with m_{min} and m_{max} set as in experiment 1. Within each simulation, we estimated the value of λ for each data set using the three methods (*l*, = L2n, ELBn, and MLE) as described above. We then fit an OLS regression separately for each replicate *k* using each method according to equation 5:

For experiment 2.1, we performed a total of 15,000 simulations (5 sites * 3 scenarios * 1000 replicates). The main results presented here were not dependent on the range of x-values or the number of sites (Supporting Information).

2.9 | Experiment 2.2: Varying the effect size of the known relationship

In the previous process, the β_{env} had an effect size (i.e. slope) of −0.5, and the intercept was varied to be equal to −2, −1.5, and −1 (i.e. shifting the window in parameter space). We wanted to test the robustness of our results by varying the effect size. We repeated the process with λ centred at −1.5, but varied the values of λ across the hypothetical gradient to have a relationship of $\beta_{env}=-0.25$ or 0 (Table 3).

TABLE 3 Variables for experiment 2.2. The environmental value, x_j are the *x*-values for the hypothetical environmental gradient. The subsequent columns represent the corresponding λ values across three different values of β_{env}.

Environmental value, x_j	$\beta_{env}=0.5$	$\beta_{env}=0.25$	$\beta_{env}=0$
−1	−1.5	−1.25	−1.5
−0.5	−1.25	−1.375	−1.5
0	−1	−1.5	−1.5
0.5	−0.75	−1.625	−1.5
1	−0.5	−1.75	−1.5

2.10 | Empirical data

We re-analysed two data sets of benthic macroinvertebrate communities from stream habitats across two different gradients. In the first, quantitative macroinvertebrate samples were collected from streams across an acid mine drainage (AMD) stress gradient. Details of the sample collection and processing can be found in (Pomeranz et al., 2019a). Briefly, all individuals from each sample were identified to the lowest practical taxonomic unit and body lengths were measured using Adobe Acrobat 9 Pro (San Jose, California, USA) photos taken with a Leica DFC295 digital camera mounted to a Leica model M125 microscope. Body mass was estimated using taxon-specific published length-weight regressions.

The second dataset was from the wadeable stream sites of National Ecological Observatory Network (NEON) (2022). NEON stream sites are located across a wide temperature gradient in the United States, from Puerto Rico to Alaska. Quantitative macroinvertebrate samples were collected using the most appropriate method based on the local habitat. All individuals were identified and had their body lengths measured, and body mass was estimated using taxon-specific published length-weight regressions. This data has been analysed for ISD relationships previously using methods described in Pomeranz et al. (2022). Detailed methods of the sample collection and initial data QA/QC processing can be found in the macroinvertebrate data product information documents found on the NEON website https://data.neonscience.org/data-products/DP1.20120.001.

Estimates of the slope coefficient (β_{AMD}, and β_{NEON}, respectively) ±1 SD, were compared across methods. This allowed us to determine whether the main results published previously differed depending on the method used.

2.11 | Performance metrics

We compared performance of each procedure (L2n, ELBn, MLE) by first plotting the distribution of site-specific λ estimates from experiment 1, and the distribution of β_{env} estimates from experiment 2 from the three methods against the known values. For each procedure we estimated the width of the 95% CI's to compare uncertainty, as well as calculated bias for the procedures overall as the median absolute difference (averaged across all simulations) between the known values and the modelled estimates. Finally, we recorded the proportion of model estimated 95% CI's which contained the known value.

3 | RESULTS

3.1 | Experiment 1: λ estimates

There was considerable variation in the λ estimate across methods (Figure 2). The distribution of estimates from the MLE method was always symmetrical and centred at the known value

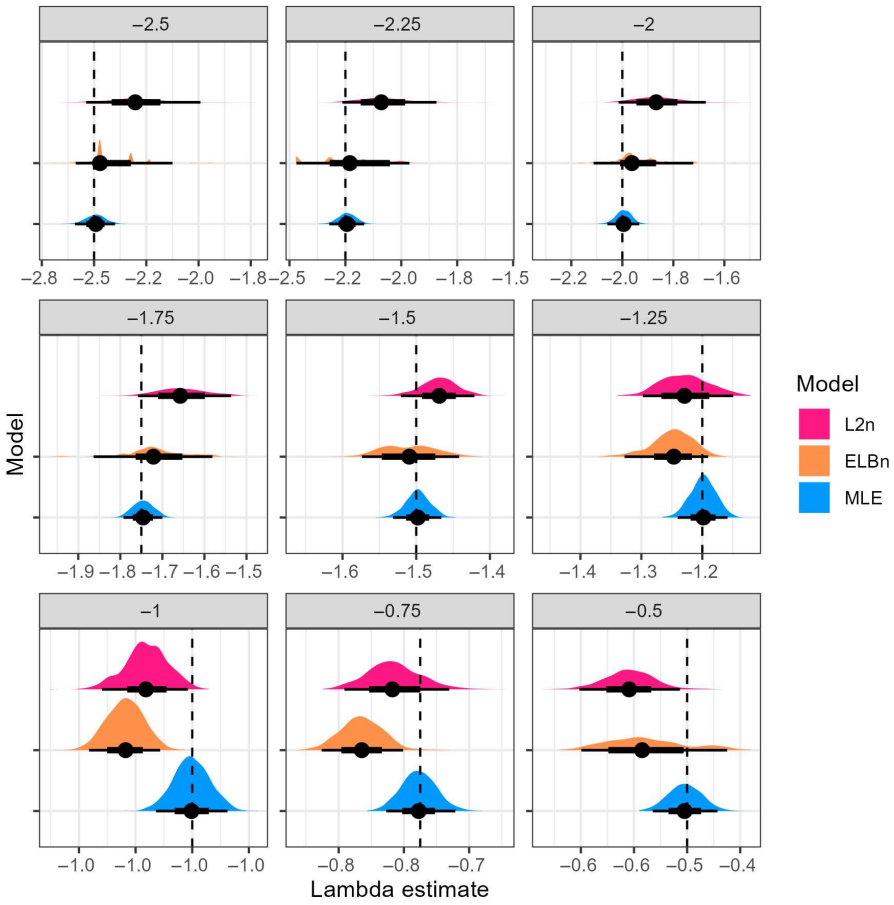

FIGURE 2 Distribution of Lambda estimates by method (colour) from random samples of body sizes from bounded power law distributions with varying exponents (−2.5 to −0.5). The figure is facetted by the known lambda parameter (facet title) and is also shown as the dashed line in each facet. Note that the *x*-axis varies in each facet.

of λ (Figure 2). The distribution of estimates from the binning methods were generally wider and occasionally asymmetrical (i.e. long right-tails for L2n and ELBn when $\lambda = -2.25$, Figure S3), or bimodal (i.e. ELBn method for $\lambda = -1.5$). We also compared the proportion of 95% CI's which contained the known value of λ for each method. The proportion of CI's produced by MLE estimates which had the known value of λ was exactly 95%, as expected. However, only 93% of the CI's from the ELBn method contained the known value of λ. The L2n method performed the worse, with only 70% of the CI's containing the known value. On average, the CI's for the λ estimates produced by the L2n and ELBn methods were ~two times wider than those produced by MLE (Table 4), indicating greater consistency of estimates from MLE. Similarly, estimates of λ deviated from the true value by an average of 0.035 or 0.045 absolute units for the L2n and ELBn methods, up to four times higher than the deviation (0.012) observed for the MLE (Table 4).

Interestingly, the two binning methods systematically overestimated λ when the simulated relationships were steeper (i.e. distributions of estimates for the binning methods are to the right when $\lambda = \sim -2.5$ to -1.5, Figure 2) and slightly underestimated λ when the simulated relationships were shallower (distributions of estimates to the left when $\lambda > \sim -1.25$). This finding was more pronounced in the L2n method compared with the ELBn method.

3.2 | Relationship across hypothetical environmental gradients

3.2.1 | Experiment 2.1: λ "scenarios"

We wanted to assess the ability of the three methods in recapturing parameters describing a known change in λ across gradients (i.e. a known relationship of $\beta_{env} = -0.5$). However, the binning methods provide inaccurate estimates of site-specific λ, and these inaccuracies were not equivalent across the range of λ values tested. The magnitude of the deviations of λ estimates for the binning methods increased with more negative (i.e. steeper) values of λ. Because of the different performance of the two binning methods at steep and shallow values of λ, we performed simulations for three separate scenarios across the λ parameter space. The three scenarios were steep (λ varies from −2.5, −1.5), medium (λ varies from −2, −1) and shallow (λ varies from −1.5, −0.5).

The MLE method (Figure 3, blue) recaptured the known slope value in each of the scenarios, with a median absolute difference of ~0.008 units between the modelled and known values (Table 4). By contrast, the binning methods systematically overestimated the known slope (Figure 3), with median absolute differences three to four times greater than the MLE. Similarly, uncertainty in the slope estimates derived from binning methods was twice that of the uncertainty in the MLE method (Table 4).

TABLE 4 Summary of three methods in recapturing the known values of site-specific λ values or the regression slopes (β_{env}) simulated in this study. Performance is determined by comparing the uncertainty (range of 95% CI's) and the absolute distance of the model estimates from the known values (median and SD of the difference). We also report the proportion of 95% CI's which contained the target value. Values are summarised across all $n = 9000$ or 6000 simulated data sets. See figures for more specific comparisons.

Target	Method	*n*	Median range of 95% CI	Median absolute deviation	SD absolute deviation	Proportion of 95% CI containing
λ	MLE	9000	0.0659	0.0119	0.0194	95%
λ	L2n	9000	0.1315	0.0450	0.0959	70%
λ	ELBn	8587	0.1699	0.0351	0.0677	93%
β_{env}	MLE	6000	0.0377	0.0084	0.0094	82%
β_{env}	L2n	6000	0.0805	0.0507	0.0424	38%
β_{env}	ELBn	6000	0.0905	0.0299	0.0326	67%

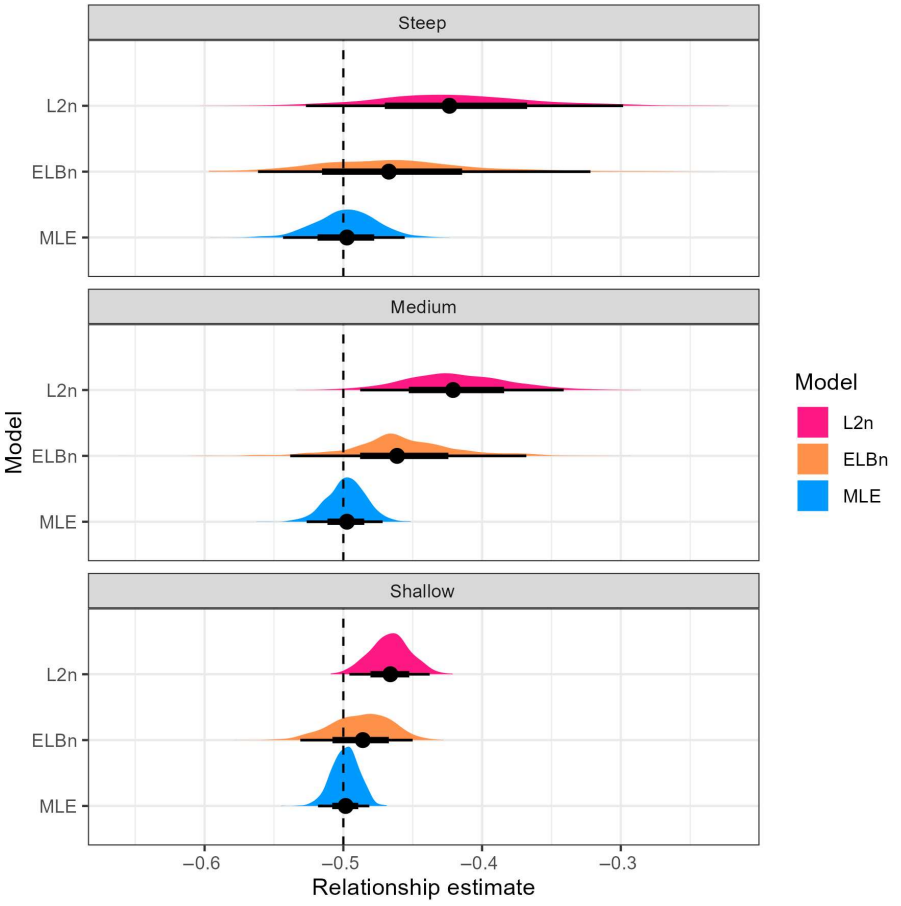

FIGURE 3 Distribution of relationship estimates (β_{env}) in three different "scenarios" of lambda values; steep: $\lambda = -2.5$ to -1.5; medium: $\lambda = -2.0$ to -1.0; steep: $\lambda = -1.5$ to -0.5. The dashed vertical line is the known relationship value of −0.5.

3.3 | Experiment 2.2: Varying the effect size of the known relationship

All methods recaptured the correct sign of the slopes, yielding qualitative consistency (Figure 4). However, the binning methods systematically underestimated the true value of the slope by ~0.05 units (Figure 5). Likewise, uncertainty in the slope estimates was always greater in the binning methods, with the width of the distributions increasing with stronger relationships across a hypothetical gradient. By comparison, the MLE showed no evidence of bias and was always centred at the known value with relatively narrow variation.

3.4 | Empirical data

The empirical dataset contained λ estimates that spanned the range observed across diverse ecosystems (−2.28 to 0.02, depending on the data set and method). Both empirical data sets yielded similar patterns to those observed in the simulated data. There was qualitative agreement in that the direction of the coefficients (i.e. β_{AMD}, β_{NEON}, coefficients) were the same among methods, with positive slopes across the pollution gradient (Figure 6a) and negative slopes across the temperature gradient. However, the magnitude of change differed between methods. The AMD slopes ranged from ~0.062 with MLE to ~0.078 with L2n (Figure 6b), while temperature slopes

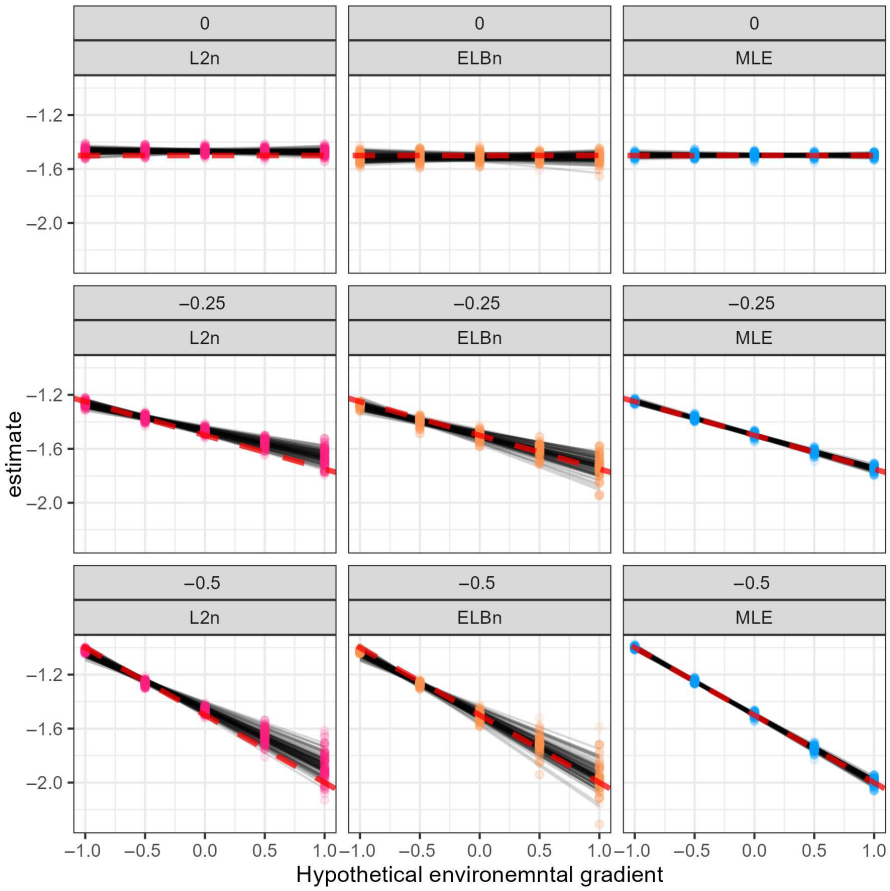

FIGURE 4 Individual regressions by method (columns) for three different known relationship values (rows from bottom to top, red dashed line). There were a total of 1000 replicates simulated for each combination of method and known relationship, but only 500 are plotted here to illustrate the variability in the regression lines.

ranged from −0.0058 with MLE to −0.0019 with ELBn (Figure 6d). As with simulated data, slope uncertainty (±1 SD) was larger in the binning methods, particularly the ELBn method (Figure 5b). Likewise, the ISD relationship parameters consistently decrease (become steeper) with increasing temperature across the NEON sites (Figure 6c).

4 | DISCUSSION

The relationship between body size and abundance has been extensively studied in a wide range of taxa inhabiting both terrestrial and aquatic ecosystems (reviewed by Brown, 1995; White et al., 2007). Empirical data shows generally consistent patterns and can be explained by the metabolic theory of ecology (Brown et al., 2004). Measuring parameters describing the decline in abundance with increasing body size in communities is being done with increasing frequency across ecology. Previous work has investigated the accuracy and inherent biases associated with different estimation methods (Edwards et al., 2017, 2020; White et al., 2007). However, the extent to which these inaccuracies and biases compound across environmental gradients remains uncertain, making it difficult to detect variation in size-abundance relationships across environmental gradients with confidence. The most important outcome of our work is that binning methods not only generate biased λ values for individual datasets but that bias carries over to affect the parameters of subsequent regressions that use those λ's as response variables. This makes it challenging to understand how λ varies in response to environmental gradients if binning is used to estimate ISD exponents.

Binning methods are easy to use and interpret, which most likely accounts for their wide use in ecological studies (Collyer et al., 2023; Martínez et al., 2016; Perkins et al., 2018). However, aggregating individuals into logarithmic bins removes a large amount of information within the data by collapsing body size variation into a single value within each bin. This is particularly true when using logarithmic bins. For example, all individuals placed into a bin that ranges from 2 to 4 g of mass are all treated as having a mass of 3 g, the midpoint of that bin. Likewise, a single abundance value is taken for each bin, despite that fact that there is almost certainly variation in the abundance of individuals that weigh ~2, ~3 or ~4 g. Moreover, the number of logarithmic-sized bins that can be produced by any dataset is limited. For the ELBn method, the number of bins is set a priori. However, a higher number of bins increases the chances of having empty bins and can lead to poorer OLS fits. This is why the number of bins using this method is often $n = 6$ data points, as this is a reasonable number of data points for a regression, but minimses the chances of having empty bins (Dossena et al., 2012; Martínez et al., 2016; Perkins et al., 2018). If the range of body sizes is sufficient, using $\log_2$ bins can increase the number of bins available, but this is sensitive to the underlying data. Finally, linear bins could be used (i.e. 1–2, 2–3, 3–4 g, etc.) to increase the number of data points. However, White

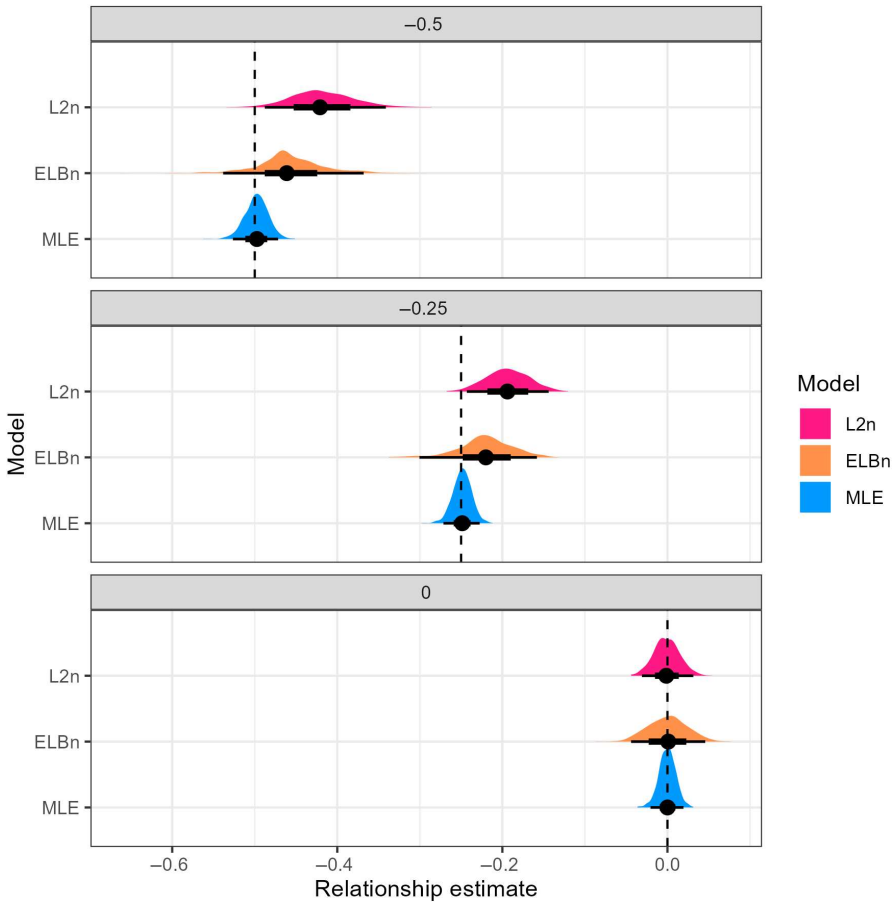

FIGURE 5 Distribution of relationship estimates (β_{env}) when estimating from different known relationships.

et al. (2008) showed that linear bins perform poorly under nearly any circumstance, and this is generally not recommended.

Regardless of the method used, binning data treats all individuals within a bin as identical, even though there are likely 100's or 1000's of individual body sizes available in each bin. By contrast, the MLE uses all the individual body size data to directly estimate λ, meaning that it not only produces more accurate estimates, but does so with less uncertainty than binning, even when the underlying data sets are identical. Likewise, MLE produces less variable and more accurate estimates of changes in λ across gradients, making it the preferred method for assessing change in λ across spatiotemporal and environmental gradients. Even when the underlying data (i.e. vectors of individual body sizes) are not available, there is a method for estimating the exponent of size-abundance relationships using maximum likelihood, the `MLEbin()` method from the *sizeSpectra* package (Edwards, 2020), which solves the same issues we discuss here. In other words, any future analyses of size spectra could use MLE estimates of λ, even if the data are only available in binned form.

At first glance, the variation in λ produced by different methods may seem trivial. For example, when the true λ was −2, the three methods gave values of −2 (MLE), − 1.96 (ELBn), and −1.87 (L2n). By themselves, all the methods appear reasonably close to the true value. However, the small differences imply very different food web structures because λ represents an emergent property governed by three ecological parameters: trophic transfer efficiency (t), predator–prey mass ratio (r), and the reciprocal of the metabolism-mass scaling exponent (b), such that

$$\lambda = \frac{\log_{10} t}{\log_{10} r} + b - 1 \quad (6)$$

Typical starting values assume that $b = 3/4$, $t \sim 0.1$, and $r \sim 10^4$ (Brown et al., 2004), and these result in a λ value of −2. To get a λ value of −1.96 (the ELBn estimate) requires a change in at least one of the three parameters. For example, keeping all else the same, we can only get $\lambda = -1.96$ if $t = 0.14$. In other words, a 2% change in λ (−1.96 vs. −2) reflects a 140% change in t (0.14 vs. 0.1). More strikingly, to get $\lambda = -1.87$ implies a change in trophic transfer efficiency of 340% (0.34 vs. 0.1, or compensating changes in r, b or both). These two examples become even more concerning in the context of environmental gradients. As shown in Figure 5, the L2n method erroneously estimates a regression slope between λ and the environmental variable of ~−0.19, when the true slope is −0.25. In other words, the two slopes diverge by 0.06 for every unit increase in the environmental variable. If we assume that the average λ bias between L2n and MLE is −0.13, then just two units of increase in the predictor variable would nearly double the size of the difference ($0.13 + 2^{*}0.06 = 0.25$), predicting drastically different food web structures despite the same underlying data (individual body sizes). The fact that small changes in λ imply large changes in food web structure emphasizes the importance of estimating λ properly, particularly when

Journal of Animal Ecology BRITISH ECOLOGICAL SOCIETY

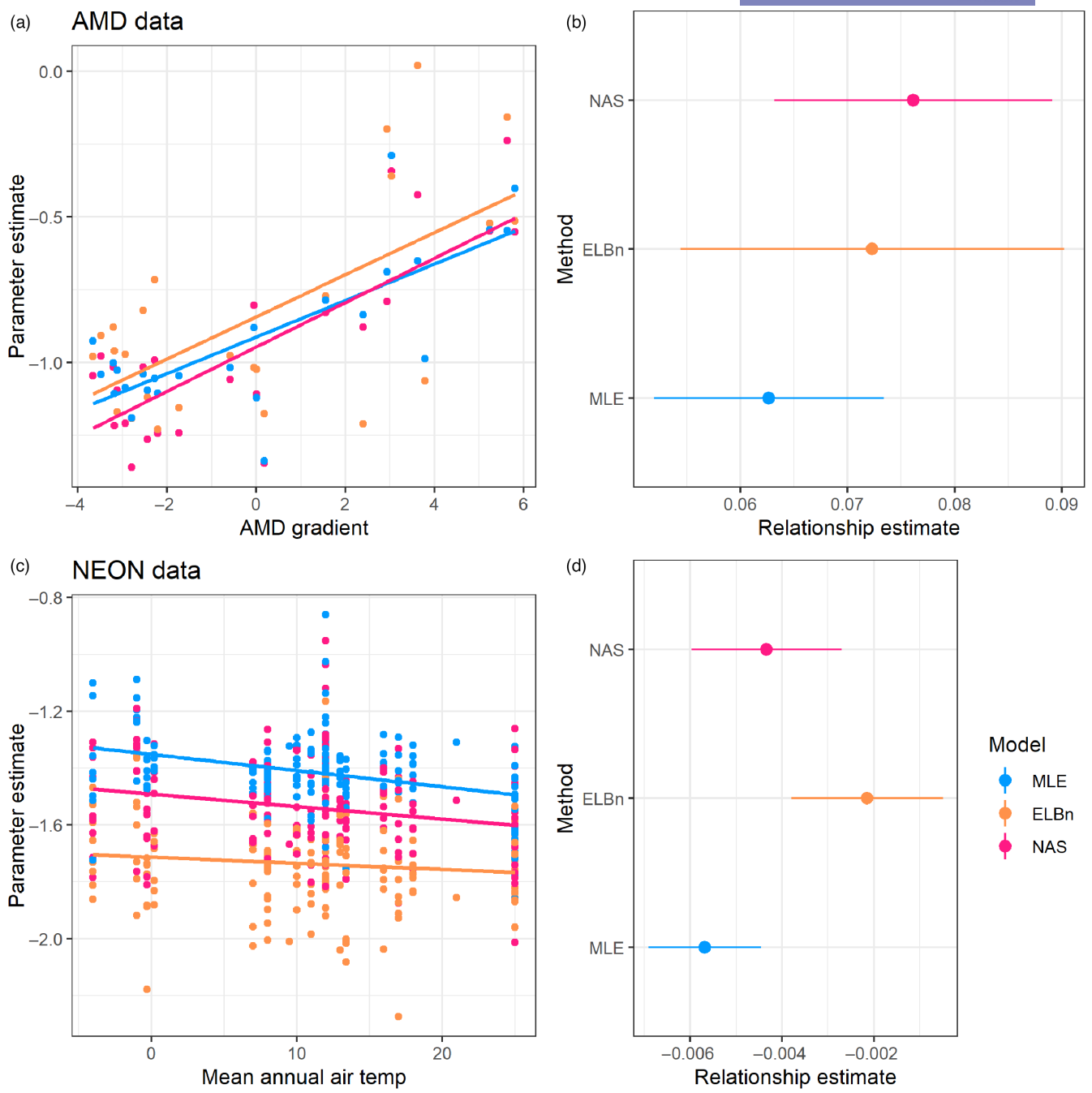

FIGURE 6 Estimates of change in exponent for size-abundance relationships across gradients from empirical data estimates. Panels (a) and (c) show the individual λ estimates for each site and the line shows the estimated relationship based on method (colour) for the natural pollution and temperature gradients, respectively. Panels (b) and (d) show the mean estimated relationship coefficient (β, point) $\pm$ 1 standard deviation (error bars) from the OLS model for both empirical data sets. All the methods estimate the same sign of the relationship, but the estimates from the binning methods are generally larger than the MLE estimates.

the changes are used to assess environmental impacts such as temperature (Pomeranz et al., 2022) or overfishing (Jennings & Blanchard, 2004).

Although there were differences in the value of the empirical relationship parameters, they were in a consistent direction and of a similar magnitude. This suggests that previously reported changes in size-abundance relationships across environmental gradients and in experimental manipulations are plausible. However, the biases and inconsistencies in the estimates of both λ and environmental response parameters presented here suggest that it may be difficult if not impossible to directly compare the relative changes across different published studies which use different methods.

The publication of individual body size data with future studies of size-abundance relationships would greatly aid in our ability to generalize changes to this fundamental aspect of community

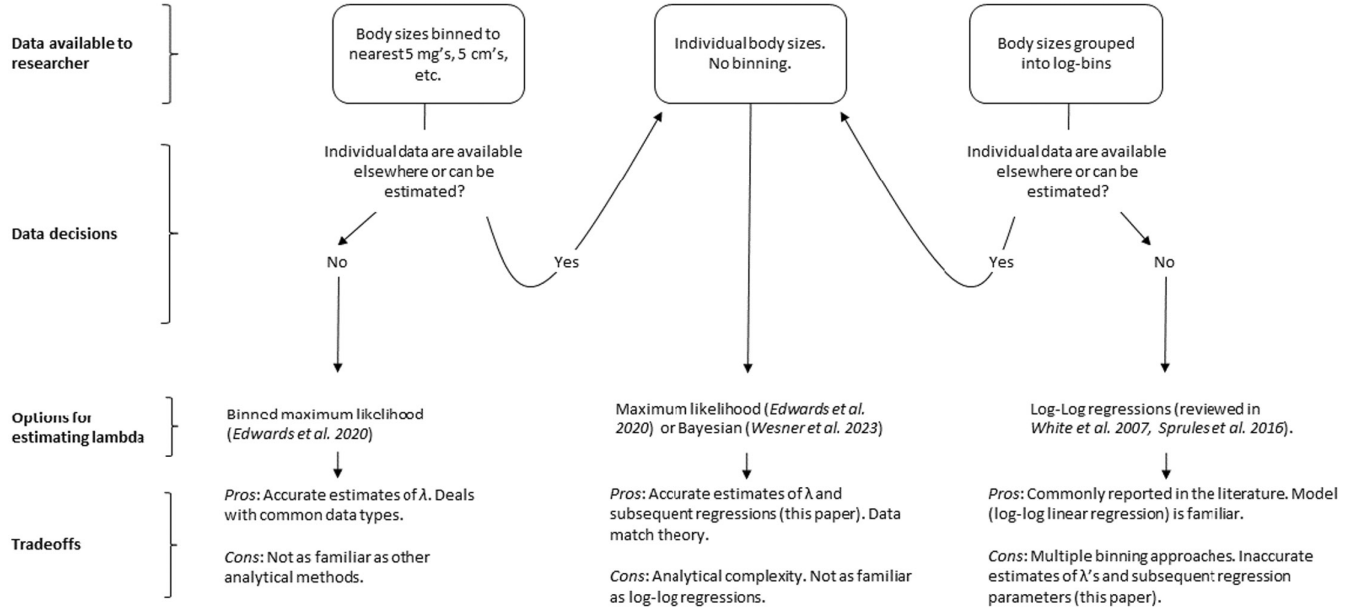

FIGURE 7 Decision tree to aid researchers in choosing an analytical framework for estimating individual size distributions. Major pros and cons of each framework are included.

organization across spatiotemporal scales and in response to environmental conditions.

5 | CONCLUSIONS

The MLE method outperformed binning methods under nearly any measure. With the publication of the `sizeSpectra` package (Edwards et al., 2020), and modifications available in other publicly available repositories (i.e. Pomeranz et al., 2022 GitHub repository: https://github.com/Jpomz/Pomeranz-Junker-Wesner), producing MLE estimates of size spectra parameters is a relatively easy task. Therefore, we recommend using it in all future studies of size-abundance relationships rather than binning. There are some circumstances where data are only available in binned formats, and individual sizes are not recorded (for example many fishery data sets group body sizes into size class bins and count the occurrence). In these situations, we recommend using the `MLEbin()` function from the `sizeSpectra` package (Edwards et al., 2020), which specifically accounts for the uncertainty of placing individuals into a size class bin. One issue with MLE analysis of ISD relationships is that it is inherently a two step process, where site-specific λ's are estimated, and then a separate analysis needs to be performed on the λ estimates across a predictor variable. However, computational techniques could be developed which will allow for the simultaneous estimation of λ and the effects of predictor variables across gradients (for example, see Wesner et al., 2023 for a hierarchical Bayesian modelling framework). Likewise, there may be other situations that we have not covered directly in this study and we provide a decision tree (Figure 7) to help guide future analyses of size-abundance relationships.

We reiterate the recommendations of White et al. (2007), Sprules and Barth (2016) and Edwards et al. (2017) to estimate ISD's using MLE methods due to their superior performance in nearly every context. Size spectra are an emergent property and depend on a number of internal processes. Even slight deviations in estimates of λ could have profound implications for interpretations of food web structure and patterns of community biomass distributions. Using MLE methods to estimate ISD parameters of communities will improve our understanding of these processes as well as aid in ecologists' ability to describe and predict the structure and organization of natural food webs and communities, as well as those which are affected by human activities. Furthermore, we strongly encourage authors to publish individual size data whenever possible. This will allow for the consistent re-analysis of existing data sets as methods develop and improve and will aid in the ability to synthesize results between research groups and across scales.

AUTHOR CONTRIBUTIONS

Justin Pomeranz, Vojsava Gjoni, Jeff S. Wesner and James R. Junker conceived the ideas and wrote the introduction. Justin Pomeranz designed the methodology, led the writing of the simulation and analysis code and made the figures. Jeff S. Wesner calculated the bias table. Justin Pomeranz led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

ACKNOWLEDGEMENTS

We thank CJP for help and support while writing this manuscript. We would also like to thank the three anonymous reviewers whose comments greatly improved this manuscript.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to share.

DATA AVAILABILITY STATEMENT

The empirical data used are publicly available (Pomeranz et al., 2019b; data dryad https://datadryad.org/stash/dataset/doi:10.5061/dryad.v6g985s, and National Ecological Observatory Network (NEON), 2022). All simulations and analyses were performed in the R statistical language (version 4.0.3, R Core Team, 2020). Scripts to reproduce the full simulation and analysis can be found at: https://github.com/Jpomz/detecting-spectra-differences.

ORCID

Justin Pomeranz https://orcid.org/0000-0002-3882-7666
James R. Junker https://orcid.org/0000-0001-9713-2330
Vojsava Gjoni https://orcid.org/0000-0003-1740-6093
Jeff S. Wesner https://orcid.org/0000-0001-6058-7972

REFERENCES

Andersen, K. H., & Beyer, J. E. (2006). Asymptotic size determines species abundance in the marine size Spectrum. *The American Naturalist*, *168*, 54–61.

Brown, J. H. (1995). *Macroecology*. University of Chicago Press.

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., & West, G. B. (2004). Toward a metabolic theory of ecology. *Ecology*, *85*, 1771–1789.

Chang, C.-W., Miki, T., Shiah, F.-K., Kao, S., Wu, J.-T., Sastri, A. R., & Hsieh, C. (2014). Linking secondary structure of individual size distribution with nonlinear size–trophic level relationship in food webs. *Ecology*, *95*, 897–909.

Collyer, G., Perkins, D. M., Petsch, D. K., Siqueira, T., & Saito, V. (2023). Land-use intensification systematically alters the size structure of aquatic communities in the Neotropics. *Global Change Biology*, *29*, 4094–4106. https://doi.org/10.1111/gcb.16720

Dossena, M., Yvon-Durocher, G., Grey, J., Montoya, J. M., Perkins, D. M., Trimmer, M., & Woodward, G. (2012). Warming alters community size structure and ecosystem functioning. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 3011–3019.

Edwards, A. M. (2020). *sizeSpectra: R package for fitting size spectra to ecological data (including binned data)*. https://github.com/andrew-edwards/sizeSpectra

Edwards, A. M., Robinson, J., Blanchard, J., Baum, J., & Plank, M. (2020). Accounting for the bin structure of data removes bias when fitting size spectra. *Marine Ecology Progress Series*, *636*, 19–33.

Edwards, A. M., Robinson, J., Plank, M., Baum, J., & Blanchard, J. (2017). Testing and recommending methods for fitting size spectra to data. *Methods in Ecology and Evolution*, *8*, 57–67.

Enquist, B. J., & Niklas, K. J. (2001). Invariant scaling relations across tree-dominated communities. *Nature*, *410*, 655–660.

Evans, T. M., Feiner, Z. S., Rudstam, L. G., Mason, D. M., Watkins, J. M., Reavie, E. D., Scofield, A. E., Burlakova, L. E., Karatayev, A. Y., & Sprules, W. G. (2022). Size spectra analysis of a decade of Laurentian Great Lakes data. *Canadian Journal of Fisheries and Aquatic Sciences*, *79*, 183–194.

Fraley, K. M., Warburton, H. J., Jellyman, P. G., Kelly, D., & McIntosh, A. R. (2018). Responsiveness of fish mass–abundance relationships and trophic metrics to flood disturbance, stream size, land cover and predator taxa presence in headwater streams. *Ecology of Freshwater Fish*, *27*, 999–1014.

Gaedke, U., Seifried, A., & Adrian, R. (2004). Biomass size spectra and plankton diversity in a shallow eutrophic Lake. *International Review of Hydrobiology*, *89*, 1–20.

Jennings, S., & Blanchard, J. L. (2004). Fish abundance with no fishing: Predictions based on macroecological theory. *Journal of Animal Ecology*, *73*, 632–642.

Jennings, S., Warr, K. J., & Mackinson, S. (2002). Use of size-based production and stable isotope analyses to predict trophic transfer efficiencies and predator–prey body mass ratios in food webs. *Marine Ecology Progress Series*, *240*, 11–20.

Martínez, A., Larrañaga, A., Miguélez, A., Yvon-Durocher, G., & Pozo, J. (2016). Land use change affects macroinvertebrate community size spectrum in streams: The case of *Pinus radiata* plantations. *Freshwater Biology*, *61*, 69–79.

Maxwell, T. A. D., & Jennings, S. (2006). Predicting abundance–body size relationships in functional and taxonomic subsets of food webs. *Oecologia*, *150*, 282–290.

Mazurkiewicz, M., Górska, B., Renaud, P. E., & Włodarska-Kowalczuk, M. (2020). Latitudinal consistency of biomass size spectra—Benthic resilience despite environmental, taxonomic and functional trait variability. *Scientific Reports*, *10*, 4164.

McGarvey, D. J., & Kirk, A. J. (2018). Seasonal comparison of community-level size-spectra in southern coalfield streams of West Virginia (USA). *Hydrobiologia*, *809*, 65–77.

Meehan, T. D. (2006). Energy use and animal abundance in litter and soil communities. *Ecology*, *87*, 1650–1658.

Mehner, T., Lischke, B., Scharnweber, K., Attermeyer, K., Brothers, S., Gaedke, U., Hilt, S., & Brucet, S. (2018). Empirical correspondence between trophic transfer efficiency in freshwater food webs and the slope of their size spectra. *Ecology*, *99*, 1463–1472.

National Ecological Observatory Network (NEON). (2022). *Macroinvertebrate collection (DP1.20120.001)*. National Ecological Observatory Network (NEON).

Nee, S., Read, A. F., Greenwood, J. J. D., & Harvey, P. H. (1991). The relationship between abundance and body size in British birds. *Nature*, *351*, 312–313.

Perkins, D. M., Durance, I., Edwards, F. K., Grey, J., Hildrew, A. G., Jackson, M., Jones, J. I., Lauridsen, R. B., Layer-Dobra, K., Thompson, M. S. A., & Woodward, G. (2018). Bending the rules: Exploitation of allochthonous resources by a top-predator modifies size-abundance scaling in stream food webs. *Ecology Letters*, *21*, 1771–1780.

Perkins, D. M., Durance, I., Jackson, M., Jones, J. I., Lauridsen, R. B., Layer-Dobra, K., Reiss, J., Thompson, M. S. A., & Woodward, G. (2021). Systematic variation in food web body-size structure linked to external subsidies. *Biology Letters*, *17*, 20200798.

Petchey, O. L., & Belgrano, A. (2010). Body-size distributions and size-spectra: Universal indicators of ecological status? *Biology Letters*, *6*, 434–437.

Pomeranz, J. P. F., Junker, J. R., & Wesner, J. S. (2022). Individual size distributions across North American streams vary with local temperature. *Global Change Biology*, *28*, 848–858.

Pomeranz, J. P. F., Warburton, H. J., & Harding, J. S. (2019a). Anthropogenic mining alters macroinvertebrate size spectra in streams. *Freshwater Biology*, *64*, 81–92.

Pomeranz, J. P. F., Warburton, H. J., & Harding, J. S. (2019b). Data from: Anthropogenic mining alters macroinvertebrate size spectra in streams. *Dryad* https://doi.org/10.5061/dryad.v6g985s

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing https://www.R-project.org/

Sheldon, R. W., & Kerr, S. R. (1972). The population density of monsters in loch ness. *Limnology and Oceanography*, *17*, 796–798.

Sprules, W. G., & Barth, L. E. (2016). Surfing the biomass size spectrum: Some remarks on history, theory, and application. *Canadian Journal of Fisheries and Aquatic Sciences*, *73*, 477–495.

Wesner, J. S., Pomeranz, J. P. F., Junker, J. R., Gjoni, V., & Lio, Y. (2023, February 15). Bayesian hierarchical modeling of size spectra. *bioRxiv*.

White, E. P., Enquist, B. J., & Green, J. L. (2008). On estimating the exponent of power-law frequency distributions. *Ecology*, *89*, 905–912.
White, E. P., Ernest, S. K. M., Kerkhoff, A. J., & Enquist, B. J. (2007). Relationships between body size and abundance in ecology. *Trends in Ecology & Evolution*, *22*, 323–330.
Yvon-Durocher, G., Montoya, J. M., Trimmer, M., & Woodward, G. (2011). Warming alters the size spectrum and shifts the distribution of biomass in freshwater ecosystems. *Global Change Biology*, *17*, 1681–1694.

SUPPORTING INFORMATION

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

Figure S1. The edges of the first 4 bins are shown when the body sizes range from 0 to 100 (bottom line) and 0–1000 (top line).

Figure S2. The edges of all bins for the L2n method are shown when body sizes range from 0 to 100 (bottom line) and 0–1000 (top line).

Figure S3. The variation in the distributions of λ estimates for "steep" values in Figure 2 in the main text makes it difficult to compare them with the "shallower" distirbutions.

Figure S4. Absolute deviation in estimate from known value of λ.

Figure S5. We varied the number of replicates ($n = 10$, 50, 100, 200, 250, 500, 750, 1000) and plotted the distribution of β estimates across the three λ scenarios.

Figure S6. CI's and λ estimates. CI's are colored if they contain the true value (blue) or if they do not (red). Only the first 500 replicates are displayed for visualization purposes.

Figure S7. CI's and β estimates when varying the known value of β.

Figure S8. CI's and β estimates when varying the scenario of λ parameter space.

Table S1. Deviation of estimates across methods for different sample sizes of body size values.

Table S2. Deviation of estimates across methods when the values of the hypothetical gradient (large *x*) and the range of body sizes (small *m*) are changed compared with the results presented in the manuscript (main).

Table S3. Deviation of estimates across methods when the number of sites was changed.

How to cite this article: Pomeranz, J., Junker, J. R., Gjoni, V., & Wesner, J. S. (2024). Maximum likelihood outperforms binning methods for detecting differences in abundance size spectra across environmental gradients. *Journal of Animal Ecology*, *93*, 267–280. https://doi.org/10.1111/1365-2656.14044