Bayesian hierarchical modeling of size spectra

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

[1]University of South Dakota, Department of Biology, Vermillion, SD 57069

[2]Colorado Mesa University, Environmental Science and Technology, Grand Junction, CO 81501

[3]Great Lakes Research Center, Michigan Technological University, Houghton, MI 49931

[4]Louisiana Universities Marine Consortium, Chauvin, LA 70344

Corresponding Author: Jeff.Wesner@usd.edu

Running Head: Bayesian Size Spectra

Data Archiving Statement

All data, R code, and Stan code are available at https://github.com/jswesner/stan_isd (to be permanently archived at Zenodo upon acceptance). Body size data from the International Benthic Trawl Surveys was retrieved from https://github.com/andrew-edwards/sizeSpectra/tree/master/data.

Abstract

1. A fundamental pattern in ecology is that smaller organisms are more abundant than larger organisms. This pattern is known as the individual size distribution (ISD), which is the frequency of all individual sizes in an ecosystem, regardless of taxon.
2. The ISD is described by power law distribution with the form $f(x) = Cx^{\lambda}$, and a major goal of size spectra analyses is to estimate the ISD parameter λ. However, while numerous methods have been developed to do this, they have focused almost exclusively on estimating λ from single samples.
3. Here, we develop an extension of the truncated Pareto distribution within the probabilistic modeling language Stan. We use it to estimate multiple ISD parameters simultaneously with a hierarchical modeling approach.
4. The most important result is the ability to examine hypotheses related to size spectra, including the assessment of fixed and random effects, within a single Bayesian generalized (non)-linear mixed model. While the example here uses size spectra, the technique can also be generalized to any data that follows a power law distribution.

Keywords: *Bayesian, body size spectra, hierarchical, Pareto, power law, Stan*

Introduction

In any ecosystem, large individuals are typically more rare than small individuals. This fundamental feature of ecosystems leads to a remarkably common pattern in which relative abundance declines with individual body size, generating the individual size distribution (ISD), also called the community size spectrum (Sprules *et al.* 1983; White *et al.* 2008). Understanding how body sizes are distributed has been a focus in ecology for over a century (Peters & Wassenberg 1983), in part because they represent an ataxic approach that reflects fundamental measures of ecosystem structure and function, such as trophic transfer efficiency (Kerr & Dickie 2001; White *et al.* 2007; Perkins *et al.* 2019). Individual size distributions are also predicted as a result of physiological limits associated with body size, thereby emerging from predictions of metabolic theory and energetic equivalence (Brown *et al.* 2004).

More formally, the ISD is a frequency distribution that can be approximated by a bounded power law with a single free parameter λ, corresponding to the following probability density function (Edwards *et al.* 2020):

$$f(x) = Cx^{\lambda}, x_{min} \leq x \geq x_{max} \tag{1}$$

where x is the body size (e.g., mass or volume) of an individual regardless of taxon, x_{min} is the smallest individual attainable and x_{max} is the largest possible individual (White *et al.* 2008). C is a constant equal to:

$$C = \begin{cases} \frac{\lambda+1}{x_{max}^{\lambda+1} - x_{min}^{\lambda+1}}, \lambda \neq -1 \\ \frac{1}{log x_{max} - log x_{min}}, \lambda = -1 \end{cases} \tag{2}$$

This model is also known as the bounded power law or truncated Pareto distribution. The terms “bounded” or “truncated” refer to the limits of x_{min} and x_{max}, which represent the minimum and maximum attainable body size values (White *et al.* 2008). In practice, values of x_{min} and x_{max}

often come from the minimum and maximum body sizes in a data set or are estimated statistically (White *et al.* 2008; Edwards *et al.* 2017).

A compelling feature of size spectra is that λ may vary little across ecosystems as a result of physiological constraints that lead to size-abundance patterns more broadly. Metabolic scaling theory predicts $\lambda + 1 = \frac{log10\alpha}{log10\beta} - 3/4$, where α is trophic transfer efficiency in the food web and β is the mean predator-prey mass ratio (Reuman *et al.* 2008). The value of $-3/4$ is the predicted scaling exponent of log abundance and log mass (Damuth 1981; Peters & Wassenberg 1983). It is the reciprocal of scaling coefficient of metabolic rate and mass (0.75) (Brown *et al.* 2004) and as a result, values of $\lambda + 1$ have been used to estimate metabolic scaling across ecosystems (Reuman *et al.* 2008; Perkins *et al.* 2018, 2019). Values of $\sim$-2 represent a reasonable first guess of expected ISD exponents, with values of ranging from -1.2 to -2 appearing in the literature (Andersen & Beyer 2006; Blanchard *et al.* 2009; Pomeranz *et al.* 2022).

Whether λ represents a fixed or variable value is debated, but it varies among samples and ecosystems (Blanchard *et al.* 2009; Perkins *et al.* 2018; Pomeranz *et al.* 2022). It is often described by its "steepness", with more negative values (i.e., "steeper") indicating lower abundance of large relative to small individuals, and vice versa. These patterns of size frequency are an emergent property of demographic processes (e.g., age-dependent mortality), ecological interactions (e.g., size-structured predation, trophic transfer efficiency), and physiological constraints (e.g., size-dependent metabolic rates) (Muller-Landau *et al.* 2006; Andersen & Beyer 2006; White *et al.* 2008). As a result, variation in λ across ecosystems or across time can indicate fundamental shifts in community structure or ecosystem functioning. For example, overfishing in marine communities has been detected using size spectra in which λ was steeper than expected, indicating fewer large fish than expected (Jennings & Blanchard 2004). Shifts in λ have also been used to document responses to acid mine drainage in streams (Pomeranz *et al.* 2019), land use (Martínez *et al.* 2016), resource subsidies (Perkins *et al.* 2018), and temperature (O'Gorman *et al.* 2017; Pomeranz *et al.* 2022).

Given the ecological information it conveys, the data required to estimate size spectra are deceptively simple; only a single column of data are needed, in which each data point is a single measure of the body size of an individual. As long as the body sizes are collected systematically and without bias towards certain taxa or phenotypes, there is no need to know any more ecological information about the data points (e.g., taxon, trophic position, age, abundance). However, despite the simple data requirement, the statistical models used to estimate λ are diverse. Edwards *et al.* (2017) documented eight analytical methods. Six involved binning, in which the body sizes are grouped into size bins (e.g., 2-49 mg, 50-150 mg, etc.) and then counted, generating values for abundance within each size bin. Binning and log-transformation allows λ to be estimated using simple linear regression. Unfortunately, the binning process also removes most of the variation in the data, collapsing information about 1000's of individuals into just 6 or so bins. Doing so can lead to the wrong values of λ, sometimes drastically so (Goldstein *et al.* 2004; White *et al.* 2008; Pomeranz *et al.* 2023).

An improved alternative to binning and linear regression is to fit the body size data to a power law probability distribution (White *et al.* 2008; Edwards *et al.* 2017, 2020). This method uses all raw data observations directly to estimate λ, typically using the maximum likelihood estimation method (Edwards *et al.* 2017). In addition to estimating size spectra of single samples, ecologists have used this method to examine how λ varies across environmental gradients (Perkins *et al.* 2019; Pomeranz *et al.* 2022). However, these analyses typically proceed in two steps. First, λ is estimated individually from each collection (e.g., each site or year, etc.). Second, the estimates are used as response variables in a linear model to examine how they relate to corresponding predictor variables (Edwards *et al.* 2020). A downside to this approach is that it treats body sizes (and subsequent λ's) as independent samples, even if they come from the same site or time. It also removes information on sample size (number of individuals) used to derive λ. As a result, the approach not only separates the data generation model from the predictor variables, but is also unable to take advantage of partial pooling during model fitting.

Here, we develop a Bayesian model that uses the truncated Pareto distribution to estimate λ in response to both fixed and random predictor variables. The model extends the maximum likelihood approach developed by Edwards *et al.* (2020) and allows for a flexible hierarchical structure, including partial pooling, within the modeling language Stan (Stan Development Team 2022).

Methods

Translating to Stan

We first translated the probability density function described by Edwards *et al.* (2020) into Stan by converting it to the log probability density function (lpdf). Stan is a probabilistic modeling language that is capable of fitting complex models, including those with custom lpdf's. The resulting lpdf is given as

$$lpdf = \begin{cases} \text{log}\frac{\lambda+1}{x_{max}^{\lambda+1} - x_{min}^{\lambda+1}} + \lambda \text{log} x, \lambda \neq -1 \\ \text{-log}(\text{log} x_{max} - \text{log} x_{min}) - \text{log} x, \lambda = -1 \end{cases} \tag{3}$$

with all variables as described above. We call this the $paretocustom$ distribution, which we can now use to estimate λ of a given data set. For example, an intercept-only model would look like this:

$$x_i \sim paretocustom(\lambda, x_{min}, x_{max})$$

$$\lambda = \alpha$$

$$\alpha \sim Normal(\mu, \sigma) \tag{4}$$

where x_i is the ith individual body size, λ is the size spectrum parameter (also referred to as the exponent), x_{min} and x_{max} are as defined above, and α is the intercept with a prior probability distribution. In this case, we specified a Normal prior since λ is continuous and can be positive or negative, but this can be changed as needed.

The simple model above can be expanded to a generalized linear mixed model by including fixed predictors ($\beta\mathbf{X}$) and/or varying intercepts ($\alpha_{[x]}$):

$$
\begin{gathered}
x_{ij} \sim paretocustom(\lambda_j, x_{min,j}, x_{max,j}) \\
\lambda = \alpha + \beta\mathbf{X} + \alpha_{[j]} + \alpha_{[x]} \\
\alpha \sim Normal(\mu_\alpha, \sigma_\alpha) \\
\beta \sim Normal(\mu_\beta, \sigma_\beta) \\
\alpha_{[j]} \sim Normal(0, \sigma_{[j]}) \\
\sigma_{[j]} \sim Exponential(\phi) \\
\alpha_{[x]} \sim Normal(0, \sigma_{[x]}) \\
\sigma_{[x]} \sim Exponential(\phi)
\end{gathered} \tag{5}
$$

with one or more β regression parameters, represented by the vector β, for one or more fixed predictors $\mathbf{X}$, and one or more varying intercepts α_x. We specify α_j separately because it is needed to account for the non-independence of body sizes. In other words, each body size x_i is clustered within each sample j and so they are not independent and identically distributed. The addition of a varying intercept for each sample accounts for this non-independence. Prior distributions are given as $Normal$ for the parameters and varying intercept and $Exponential$ for $\sigma[x]$, but these can also be changed as needed.

The model above assumes that each body size x represents a single individual such that the data set might have many repeats for individuals of the same size (e.g., x = {0.2, 0.2, 0.2, 0.4, 0.4, 0.5, 9.8}). However, when individual body sizes are repeated in a data set, they are often accompanied by a count or density, such that the data set above might instead consist of two columns with x = {0.2, 0.4, 0.5, 9.8} and $counts$ = {3, 2, 1, 1}. To analyze this more compact data set, Edwards *et al.* (2020) developed a modification of the log probability density function to include $counts$:

$$lpdf = \begin{cases} counts(\log\frac{\lambda+1}{x_{max}^{\lambda+1} - x_{min}^{\lambda+1}} + \lambda\log x), \lambda \neq -1 \\ counts(-\log(\log x_{max} - \log x_{min}) - \log x, \lambda = -1 \end{cases} . \tag{6}$$

We refer to this as $paretocounts$, such that the model can be fit by using

$$x_i \sim paretocounts(\lambda, x_{min}, x_{max}, counts)$$

$$\lambda = \text{[linear or non-linear model] and}$$

$$\text{[priors].} \tag{7}$$

Aside from adding $counts$, the model is the same as presented above. These models ($paretocustom$ and $paretocounts$) allow us to test how the size distribution parameter, λ, varies in response to continuous or categorical predictors and to include hierarchical structure as needed.

Testing the models

The $paretocustom$ and $paretocounts$ lpdfs give the same results, differing only in how the data are aggregated. For simplicity, we demonstrate model performance here for the $paretocounts$ distribution, since the empirical data we used (see *Case Study* below) contains counts of individual body sizes. First, we tested for parameter recovery using data simulated from a bounded power law with known values of λ. Second, we fit the model to fisheries trawl data presented in Edwards *et al.* (2020) to estimate the hypothesis that λ declines over time.

Parameter recovery from simulated data

To ensure that the models could recover known parameter values, we simulated ten data sets from a bounded power law using the inverse cumulative density function:

$$x_i = (\mathrm{u}_i x_{max}^{(\lambda+1)} + (1 - \mathrm{u}_i) x_{min}^{(\lambda+1)})^{\frac{1}{(\lambda+1)}} \quad (8)$$

where x_i is the individual body size from the ith simulation, u_i is a unique draw from a $Uniform(0,1)$ distribution, and all other variables are the same as defined above. We set x_{min} = 1, x_{max} = 1000, and simulated i = 1000 values from each of 10 λ's ranging from -2.2 to -1.2. To generate $counts$, we rounded each simulated value to the nearest 0.001 and then tallied them.

We estimated the ten λ values in two ways. First, we fit a separate intercept-only model to each of the ten data sets. Second, we fit a varying intercept model (Gelman et al. 2014). The structure of this model is $\lambda = \alpha + \alpha_{[j]}$ where each group j represents an offset from the mean value of lambda.

Finally, we simulated data for a regression model with a single continuous predictor and a varying intercept: $\lambda = \alpha + \beta x + \alpha_{[j]}$, where α = -1.5, β = -0.1, and σ_j = 0.3. The predictor variable x was a continuous predictor. Using these parameters, we simulated 18 λ's, with each λ coming from one of three x-values (-2, 0, 2), nested within 3 groups with each replicated twice. From each λ, we simulated 1000 individuals using the procedure above, with x_{min} = 1 and x_{max} = 1000. Using those 18,000 simulated body sizes (1000 sizes simulated from 18 λ's), we fit a $paretocounts$ regression model 40 times to measure variation in parameter recovery among model runs.

Sample Size

We examined sensitivity to sample size (number of individual body sizes) across three λ values (-2, -1.6, -1.2). For each λ, we varied the number of simulated individuals from 2 to 2048, representing a 2^n sequence with n ranging from 1 to 11. Each of the 11 densities was replicated 10 times resulting in 110 datasets of individual body sizes. We fit each data set using separate intercept-only $paretocounts$ models and then plotted the resulting λ values as a function of sample size.

Case Studies

To examine model performance on empirical data, we re-ran a previously published analysis from Edwards et al. (2020). In Edwards' study, size spectra parameters were first estimated separately for each sample using maximum likelihood. Then the modeled parameters were used as response variables in linear regression models. The goal was to test for linear changes in size spectra over three decades using bi-yearly size data of marine fishes collected from the International Benthic Trawl Survey (IBTS). The data set and original model results are available in the `sizeSpectra` package (Edwards *et al.* 2017). We tested the same hypothesis as Edwards *et al.* (2020), but instead of using a two-step process we fit a single model using the $paretocounts$ lpdf.

Model Fitting

We fit each of the above models in `rstan` (Stan Development Team 2022) using 2 chains each with 1000 iterations. All models converged with R_{hat}'s <1.01. If a known parameter value fell inside the 95% Credible Intervals (CrI), we considered parameter recovery successful. For the replicated regression model, we also tallied the number of times that the known value fell outside of the 95% CrI. Assessments of prior influence and model checking are available in Appendix S1.

Data Availability Statement

All data, R code, and Stan code are available at https://github.com/jswesner/stan_isd (to be permanently archived on acceptance).

Results

Parameter Recovery

For models fit to simulated individual data sets, all 95% credible intervals included the true value of λ and posterior medians were no more than 0.05 units away from the true value (Table 1). Similarly, when the same data set was fit using a varying intercepts model, the posterior median intercept α and group standard deviation σ_j were nearly identical to the true values (Table 1). Using the varying intercept model to estimate group specific means yielded similar results as using separate models per group (Figure 1a), demonstrating that a single model can be used to estimate multiple size spectra. The change from "shallow" to "steep" size spectra is also evident in plots of the proportion of values $\geq x$ (i.e., $f(x)$ from Eq. 1) (Figure 1b-d).

We also recovered regression parameters (α, β) along with the group-level standard deviation (σ_j; Figure 2). Thirty-seven of the 40 models converged. Of those 37 models the true value fell outside of the 95% CrI once for α and σ_j and three times for β (Figure 2). Averaging the deviations (posterior median minus the true value) among the replicates indicated no bias in the modeled estimates (mean bias $\pm$ sd: α = -0.01 $\pm$ 0.05, β = 0001 $\pm$ 0.004, σ_j = 0.02 $\pm$ 0.05).

Sample Size

Variation in modeled estimates was high for samples containing less than 100 individuals (Figure 3). For example, when the true λ value was -2, samples with just 8 individuals yielded estimates ranging from -2.7 to -1.7. By contrast, all samples with more than 300 individuals captured the true λ with less than 0.1 unit of error (Figure 3).

Case Study

Using IBTS data (Edwards *et al.* 2017) with a Bayesian hierarchical regression, we found a negative trend over time. The ISD parameter of IBTS trawl data declined by ~0.001 units per year, but with a 95% CrI ranging from -0.005 to 0.002. These values were nearly identical to those reported by Edwards et al. (2020) using a two-step approach (Table 2). An advantage of fitting the model in a single Bayesian hierarchical framework is that estimates for individual groups are pulled toward the mean via partial pooling. This is apparent in comparing the unpooled MLE estimates (Figure 4a) to the partially pooled Bayesian estimates in each year (Figure 4b).

Discussion

The most important result of this work is the ability to analyze ISD parameters using fixed and random predictors in a hierarchical model. Our approach allows ecologists to test hypotheses about size spectra while avoiding the pitfalls of binning, which loses information and can lead to biased estimates of λ (White *et al.* 2008). Maximum likelihood solves this problem by directly estimating the ISD, but testing hypotheses with maximum likelihood is often done with a two-step process in which λ is estimated individually for each sample and the results are then used as response variables in linear or non-linear models (Edwards *et al.* 2020). Our approach merges these steps, allowing for the incorporation of prior probabilities and hierarchical structure.

The ability to incorporate prior information using Bayesian updating has two practical advantages over the two-step process described above. First, adding informative prior distributions can improve model fit by limiting the MCMC sampler to reasonable sampling space. In other words it would not be sensible to estimate the probability that λ is -1,234 or -9. Without informative priors, those values (and more extreme values) are considered equally likely and hence waste much of the algorithm's sampling effort on unlikely values (e.g., (Wesner & Pomeranz 2021)).

Second, and most importantly, ecologists have much prior information on the values that λ can take.

For example, global analysis of phytoplankton reveals values of -1.75, consistent with predictions based on sub-linear scaling of metabolic rate with mass of -3/4 (Perkins *et al.* 2019). Alternatively, Sheldon's conjecture suggests that λ is -2.05 (Andersen et al. 2006), a value reflecting isometric scaling of metabolic rate and mass, with support in pelagic marine food webs (Andersen & Beyer 2006). However, benthic marine systems typically have shallower exponents (e.g., $\sim$ -1.4; Blanchard *et al.* (2009)), similar to those in some freshwater stream ecosystems (~ -1.25 (Pomeranz *et al.* 2022). While the causes of these deviations from theoretical predictions are debated, it is clear that values of λ are restricted to a relatively narrow range between about -2.05 and -1.2. But this restriction is not known to the truncated Pareto, which has no natural lower or upper bounds on λ (White *et al.* 2008). As a result, a prior that places most of its probability mass on these values (e.g., $Normal(-1.75, 0.2)$ seems appropriate. Such a continuous prior does not prevent findings of larger or smaller λ, but instead places properly weighted skepticism on such values.

Similar to priors, partial pooling from varying intercepts provides additional benefits, allowing for the incorporation of hierarchical structure and pulling λ estimates towards the global mean (Gelman 2005; Qian *et al.* 2010). In the examples shown here, the amount of pooling is relatively small because the sample sizes are large (>1000 individuals). However, the primary benefit of pooling (both from varying effects and skeptical priors) is in prediction (Gelman 2005; Hobbs & Hooten 2015). This becomes especially important when models are used to forecast future ecosystem conditions. Forecasts are becoming more common in ecology (Dietze *et al.* 2018) and are likely to be easier to test with modern long-term data sets like NEON (National Ecological Observatory Network) in which body size samples will be collected at the continental scale over at least the next 20 years (Kuhlman *et al.* 2016). In addition, because the effects of priors and pooling increase with smaller samples sizes, varying intercepts are likely to be particularly helpful for small samples. In other words, priors and partial pooling contain built-in skepticism of extreme values, ensuring the maxim that "extraordinary claims require extraordinary evidence".

One major drawback to the Bayesian modeling framework here is time. Bayesian models of even minimal complexity must be estimated with Markov Chain Monte Carlo techniques. In this study, we used the No U-Turn sampling (NUTS) algorithm via `rstan` (Stan Development Team 2022). Stan can be substantially faster than other commonly used programs such as JAGS and WinBUGS, which rely on Gibbs sampling. For example, Stan is 10 to 1000 times more efficient than JAGS or WinBUGS, with the differences becoming greater as model complexity increases (Monnahan *et al.* 2017). In the current study, intercept-only models for individual samples with $\sim$ 300 to 1500 individuals could be fit quickly (<2 seconds total run time (warm-up + sampling on a Lenovo T490 with 16GB RAM)) with as little as 1000 iterations and two chains. However, the IBTS regression models took >2 hours to run with the same iterations and chains. These times include the fact that our models used several optimization techniques, such as informative priors, standardized predictors, and non-centered parameterization, each of which are known to improve convergence and reduce sampling time (McElreath 2016). But if Bayesian inference is desired, these run-times may be worth the wait. In addition, they are certain to become faster with the refinement of existing algorithms and the introduction of newer ones like Microcanonical HMC (Robnik *et al.* 2022).

Body size distributions in ecosystems have been studied for decades, yet comprehensive analytical approaches to testing these hypotheses are lacking. We present a single analytical approach that takes advantage of the underlying data structures of individual body sizes (Pareto distributions) while placing them in a generalized (non)-linear hierarchical modeling framework. We hope that ecologists will adopt and improve on the models here to critically examine hypotheses of size spectra or other power-law distributed data. Moreover, while the examples here are for ecological size spectra, the statistical approach is not limited to ecological data, but can be applied to analysis of power law distributions that are common in a wide variety of disciplines (Aban *et al.* 2006; Clauset *et al.* 2009).

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Nos. 2106067 to JSW and 2106068 to JRJ. We thank Professor Yuhlong Lio for statistical and mathemat-

ical advice and Edwards et al. (2017) and (2020) for placing their code and data in easily accessible repositories.

References

Aban, I.B., Meerschaert, M.M. & Panorska, A.K. (2006). Parameter estimation for the truncated pareto distribution. *Journal of the American Statistical Association*, **101**, 270–277.

Andersen, K.H. & Beyer, J.E. (2006). Asymptotic Size Determines Species Abundance in the Marine Size Spectrum. *The American Naturalist*, **168**, 54–61.

Blanchard, J.L., Jennings, S., Law, R., Castle, M.D., McCloghrie, P., Rochet, M.-J. & Benoît, E. (2009). How does abundance scale with body size in coupled size-structured food webs? *Journal of Animal Ecology*, **78**, 270–280.

Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M. & West, G.B. (2004). Toward a metabolic theory of ecology. *Ecology*, **85**, 17711789.

Clauset, A., Shalizi, C.R. & Newman, M.E. (2009). Power-law distributions in empirical data. *SIAM review*, **51**, 661703.

Damuth, J. (1981). Population density and body size in mammals. *Nature*, **290**, 699–700.

Dietze, M.C., Fox, A., Beck-Johnson, L.M., Betancourt, J.L., Hooten, M.B., Jarnevich, C.S., Keitt, T.H., Kenney, M.A., Laney, C.M., Larsen, L.G. & others. (2018). Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences*, **115**, 1424–1432.

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

Gelman, A. (2005). Analysis of variance—why it is more important than ever. *The Annals of Statistics*, **33**, 1–53.

Goldstein, M.L., Morris, S.A. & Yen, G.G. (2004). Problems with fitting to the power-law distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, **41**, 255–258.

Hobbs, N.T. & Hooten, M.B. (2015). *Bayesian models: A statistical primer for ecologists*. Princeton University Press.

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

Kerr, S.R. & Dickie, L.M. (2001). *The biomass spectrum: A predator-prey theory of aquatic production*. Columbia University Press.

Kuhlman, M.R., Loescher, H.W., Leonard, R., Farnsworth, R., Dawson, T.E. & Kelly, E.F. (2016). A New Engagement Model to Complete and Operate the National Ecological Observatory Network. *The Bulletin of the Ecological Society of America*, **97**, 283–287.

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. Retrieved from http://doi.wiley.com/10.1111/fwb.12680

McElreath, R. (2016). *Statistical rethinking: A bayesian course with examples in R and stan*, Secondn. CRC Press.

Monnahan, C.C., Thorson, J.T. & Branch, T.A. (2017). Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. *Methods in Ecology and Evolution*, **8**, 339–348.

Muller-Landau, H.C., Condit, R.S., Harms, K.E., Marks, C.O., Thomas, S.C., Bunyavejchewin, S., Chuyong, G., Co, L., Davies, S., Foster, R. & others. (2006). Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models. *Ecology letters*, **9**, 589–602.

O'Gorman, E.J., Zhao, L., Pichler, D.E., Adams, G., Friberg, N., Rall, B.C., Seeney, A., Zhang, H., Reuman, D.C. & Woodward, G. (2017). Unexpected changes in community size structure in a natural warming experiment. *Nature Climate Change*, **7**, 659–663.

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., Perna, A., Adrian, R., Cermeño, P., Gaedke, U., Huete-Ortega, M., White, E.P. & Yvon-Durocher, G. (2019). Energetic equivalence underpins the size structure of tree and phytoplankton communities. *Nature Communications*, **10**, 255.

Peters, R.H. & Wassenberg, K. (1983). The effect of body size on animal abundance. *Oecologia*, **60**, 89–96.

Pomeranz, J.P., Junker, J.R., Gjoni, V. & Wesner, J.S. (2023). Detecting differences in size spectra. *bioRxiv*, 2023–03.

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. (2019). Anthropogenic mining alters macroinvertebrate size spectra in streams. *Freshwater Biology*, **64**, 81–92.

Qian, S.S., Cuffney, T.F., Alameddine, I., McMahon, G. & Reckhow, K.H. (2010). On the application of multilevel modeling in environmental and ecological studies. *Ecology*, **91**, 355–361.

Reuman, D.C., Mulder, C., Raffaelli, D. & Cohen, J.E. (2008). Three allometric relations of population density to body mass: Theoretical integration and empirical tests in 149 food webs. *Ecology Letters*, **11**, 1216–1228.

Robnik, J., De Luca, G.B., Silverstein, E. & Seljak, U. (2022). Microcanonical Hamiltonian Monte Carlo. Retrieved from https://arxiv.org/abs/2212.08549

Sprules, W.G., Casselman, J.M. & Shuter, B.J. (1983). Size Distribution of Pelagic Particles in Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **40**, 1761–1769.

Stan Development Team. (2022). RStan: The R interface to Stan.

Wesner, J.S. & Pomeranz, J.P.F. (2021). Choosing priors in Bayesian ecological models by simulating from the prior predictive distribution. *Ecosphere*, **12**.

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.

Tables

Table 1: Table 1. Parameter recovery of the same data using two approaches. First, ten separate models individually recapture known lambda values. Second, the same ten data sets are estimated in a single hierarchical model. The true values are compared to the posterior median and 95% Credible Intervals.

Model	Parameter	True Value	q2.5	q50	q97.5
Separate Models	λ	-2.20	-2.23	-2.15	-2.08
Separate Models	λ	-2.09	-2.15	-2.09	-2.02
Separate Models	λ	-1.98	-2.08	-2.02	-1.96
Separate Models	λ	-1.87	-1.93	-1.87	-1.81
Separate Models	λ	-1.76	-1.75	-1.70	-1.65
Separate Models	λ	-1.64	-1.65	-1.60	-1.56
Separate Models	λ	-1.53	-1.54	-1.50	-1.46
Separate Models	λ	-1.42	-1.45	-1.42	-1.38
Separate Models	λ	-1.31	-1.34	-1.30	-1.26
Separate Models	λ	-1.20	-1.23	-1.20	-1.16
Single Model with Varying Intercepts	α	-1.70	-1.96	-1.71	-1.50
Single Model with Varying Intercepts	σ_[group]	0.34	0.23	0.36	0.58

Table 2: Table 2. Slope values from a regression testing the relationship between the ISD exponent and year for IBTS trawl data (Edwards et al. 2020). The values are derived using the Bayesian hierarchical model presented here or from the maximum likelihood approach described in Edwards et al. (2020).

Model	Mean	q2.5	q97.5
Bayesian - one step	-0.001	-0.005	0.002
MLE - two steps	-0.001	-0.005	0.003

Figure Captions

Figure 1. a) Modeled estimates (median +/- 95% Credible Intervals) of λ using either 10 separate models or a single model with ten varying intercepts. c-d) Fit of ISD relationships at 3 values of λ. Dots are raw data, lines are posterior medians, and shading is the 95% credible interval. Fits are from separate models.

Figure 2. Posterior distributions of n = 40 modeled estimates of alpha, beta, and sigma_j for a linear regression estimating the size spectrum exponent as a function of a continuous predictor. All data were simulated. Gray densities indicate that the 95% CrI contains the true value, while black densities indicate the true values fall outside of the CrI. The vertical lines indicate true values.

Figure 3. Estimates of λB across 11 different sample sizes (ranging from 2 to 2048 individuals) and three different true λ's (-2, -1.6, -1.2). Ten separate models were fit for each of the 11 sample sizes. The horizontal lines show the true value of λ.

Figure 4. Regression results from a) Edwards et al. (2020) using maximum likelihood and linear regression (two steps) and b) the Bayesian model with varying intercepts. In a) the points represent maximum likelihood estimates calculated separately for each year. In b) they represent hierarchical varying intercepts calculated from the model.