Testing and recommending methods for fitting size spectra to data

Andrew M. Edwards[1,2,5], James P. W. Robinson[2], Michael J. Plank[3,4], Julia K. Baum[2] and Julia L. Blanchard[5]

[1]*Pacific Biological Station, Fisheries and Oceans Canada, 3190 Hammond Bay Road, Nanaimo, British Columbia, V9T 6N7, Canada;* [2]*Department of Biology, University of Victoria, PO Box 1700 STN CSC, Victoria, British Columbia, V8W 2Y2, Canada;* [3]*School of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand;* [4]*Te Pūnaha Matatini, a New Zealand Centre of Research Excellence;* [5]*Institute for Marine and Antarctic Studies, University of Tasmania, Private Bag 129, Hobart, TAS 7001, Australia.*

[5]Email: Andrew.Edwards@dfo-mpo.gc.ca

Correspondence author: Andrew M. Edwards, Pacific Biological Station, Fisheries and Oceans Canada, 3190 Hammond Bay Road, Nanaimo, British Columbia, V9T 6N7, Canada. Tel: +1 250 756 7146. Fax: +1 250 756 7053. Email: Andrew.Edwards@dfo-mpo.gc.ca

Running title: Size spectra methods

Word count (including references, tables and figure legends): 7072

May 24, 2016

Summary

1. The size spectrum of an ecological community characterises how a property, such as abundance or biomass, varies with body size. Size spectra are often used as ecosystem indicators of marine systems. They have been fitted to data from various sources, including groundfish trawl surveys, visual surveys of fish in kelp forests and coral reefs, sediment samples of benthic invertebrates and satellite remote-sensing of chlorophyll.

2. Over the past decade several methods have been used to fit size spectra to data. We document eight such methods, demonstrating their commonalities and differences. Seven methods use linear regression (of which six require binning of data), while the eighth uses maximum likelihood estimation. We test the accuracy of the methods on simulated data.

3. We find that four of the methods can sometimes give reasonably accurate estimates of the exponent of the individual size distribution (which is related to the slope of the size spectrum). However, sensitivity analyses find that maximum likelihood estimation is the only method that is consistently accurate, and the only one that yields reliable confidence intervals for the exponent.

4. We therefore recommend the use of maximum likelihood estimation when fitting size spectra. To facilitate this we provide documented R code for fitting and plotting results. This should provide consistency in future studies and improve the quality of any resulting advice to ecosystem managers. In particular, the calculation of reliable confidence intervals will allow proper consideration of uncertainty when making management decisions.

Key-words: individual size distribution, ecosystem indicators, ecosystem approach to fisheries, biomass size spectrum, abundance size spectrum, bounded power-law distribution, truncated Pareto distribution.

Introduction

For aquatic ecosystems, size-based indicators are tools for understanding food-web structure and enabling cost-effective monitoring (Shin *et al.*, 2005). One indicator, the size spectrum (Sheldon and Parsons, 1967; Sheldon *et al.*, 1972), has been adopted by several fields in ecology as a method of quantifying the distribution of body size, or other biological or ecological traits, across a community. Size spectra are commonly used to examine fishing impacts at the community or ecosystem level (Rice and Gislason, 1996; Bianchi *et al.*, 2000; Shin *et al.*, 2005; Law *et al.*, 2012; Jacobsen *et al.*, 2014; Thorpe *et al.*, 2015) and have been more broadly used in analyses of macroecological patterns (Jennings *et al.*, 2008; Reuman *et al.*, 2008) and dynamical food-web models (Blanchard *et al.*, 2009; Hartvig *et al.*, 2011). Despite widespread use of the size spectrum, its success as a general tool in marine and terrestrial ecology has been hampered by confusion surrounding definitions of size spectra (White *et al.*, 2007), and by methodological inconsistencies in how size spectra are fitted to data (Vidondo *et al.*, 1997).

For a fish community, Rice and Gislason (1996) define size spectra as generally being 'the variation in a community property across the size range of fish in the community'. This allows for different types of spectra, such as the traditional biomass size spectrum (Boudreau and Dickie, 1992), the abundance size spectrum (Rice and Gislason, 1996), and the diversity size spectrum (Reuman *et al.*, 2014).

White *et al.* (2007) give a more specific definition of a size spectrum as the relationship between the number of individuals in a body-size class and the average size of that body-size class. Typically, the pattern is linear on logarithmic axes and is quantified by the slope, which ideally should be uniquely defined. However, if the same data set (e.g. individual body masses of fish in a community) is given to two researchers, under current practices it is not clear that they would obtain the same value for slope of the size spectrum. This is because there are usually choices to be made in determining the slope: (i) how to define

the size classes to bin the data, and (ii) how to plot the binned data.

White *et al.* (2007) point out that the size spectrum is, more generally, a frequency distribution or probability density of body sizes of individuals in a community, and recommend the term 'individual size distribution' (ISD). We adopt this approach because it moves away from the need to define somewhat arbitrary body-size classes. By thinking of body-size data as individual measurements drawn from a probability distribution, we can fit the distribution using likelihood methods (that do not require binning), to give a uniquely defined parameter that is analogous to the size-spectrum slope.

To determine such a parameter requires specifying a probability distribution for the ISD. Size spectra typically exhibit power-law relationships (Platt and Denman, 1978; Boudreau and Dickie, 1992; White *et al.*, 2007; Reuman *et al.*, 2008). For example, in community size-spectrum models 'the number of individuals in each size group is often found to exhibit a power-law relationship with size' (Andersen and Beyer, 2006), and in empirical studies, fitting of straight lines on logarithmic axes implicitly implies the fitting of a power-law relationship (Newman, 2005). Therefore a power-law distribution (or Pareto distribution or Zipf's law, Newman 2005) is the distribution to be specified; Vidondo *et al.* (1997) recommended thinking about size spectra in such a context. Specifically, we specify a bounded (truncated), rather than the usual unbounded, power-law distribution (see Methods).

Here, we describe and test eight different methods that have been used to fit size spectra. Six of these methods require binning the data in some way, plotting the binned data and fitting a linear regression. The seventh involves no binning and fits a linear regression to all data points, while the eighth involves maximising the likelihood of a distribution. Using simulated data, we test the accuracy of each method in determining point estimates and confidence intervals for the exponent of the ISD.

Our results first demonstrate that estimated slopes are not comparable between regression-based methods because the different methods are not estimating the same parameter, even

though this may have been assumed or implied in the past. However, for most methods the estimated slopes can be adjusted to provide comparable estimates of the exponent of the ISD. Some methods perform much better than others, but sensitivity analyses show that maximum likelihood estimation is the only method that is consistently accurate, and the only one that yields reliable confidence intervals. We also extend it to deal with data that are only available in binned form.

Therefore, we recommend maximum likelihood estimation, in contrast to previous advice (Vidondo *et al.*, 1997). Since this method is computationally more complicated than the regression-type approaches, in the Supporting Information we provide fully documented and functionalised R code (R Core Team, 2015) intended to be used by other researchers to reproduce our results and to apply methods to their own data.

Materials and Methods

Individual size distribution

Let the random variable X represent the body mass of an individual fish (or other organism). Considering X to come from a bounded power-law (PLB) distribution, the probability density function for X is

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

where

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

x represents possible values of X, log is the natural logarithm, b is an exponent, and $x_{\min}$ and $x_{\max}$ are the minimum and maximum possible values of body mass (with $0 < x_{\min} < x_{\max}$).

The normalisation constant C is calculated by solving $\int_{x_{\min}}^{x_{\max}} f(x)\mathrm{d}x = 1$. Assuming that the body mass of each individual fish is independently distributed according to (1) means that (1) is the ISD. Because of the normalisation constant, it describes the shape of the size spectrum independently of the total abundance of fish. The ISD is characterised by the exponent b that needs to be estimated from data. This exponent is expected to be negative, and can be related to the slope of the size spectrum, though exactly how depends on the method used to estimate the slope (see Results). A steepening slope (e.g. due to selective fishing of larger fish) corresponds to a more negative b.

We use a bounded rather than unbounded ($x_{\max} \to \infty$) distribution for several reasons. By definition the unbounded distribution assumes that individuals can, and occasionally will, attain extremely large body masses, even though such body masses are unrealistic. In related tests of the distribution of the mean body masses of species, the bounded power law was overwhelmingly more supported than the unbounded power law (Reuman *et al.*, 2008) – real biological data inherently have an upper bound. Also, ecological surveys are often designed to sample a specific range of body sizes, leading to size spectra being fit across a finite range (e.g. Dulvy *et al.* 2004; Trebilco *et al.* 2015), so a bounded distribution is being implicitly assumed (even though for most methods the distinction cannot be made). Finally, Graham *et al.* (2005), for example, calculated size-spectra slopes that estimated b to be between -0.24 and -0.20. Such values of $b > -1$ are only possible for bounded, and not for unbounded (e.g. Edwards 2008), power-law distributions.

For a community of n individuals, the abundance density function, $N(x)$, is

$$N(x) = nf(x) = nCx^b, \qquad x_{\min} \leq x \leq x_{\max}. \tag{3}$$

This leads to the biomass density function, $B(x)$, that describes how biomass is distributed with respect to body mass:

$$B(x) = xN(x) = nCx^{b+1}, \qquad x_{\min} \leq x \leq x_{\max}. \tag{4}$$

This is the equation for the biomass size spectrum (Boudreau and Dickie, 1992), and allows calculation of the total biomass of all individuals with body mass $\leq x$ (see Appendix); see also Vidondo *et al.* (1997).

Some studies (e.g. Dulvy *et al.*, 2004; Daan *et al.*, 2005; Boldt *et al.*, 2012) use length to represent size, and calculate the slope of the length size spectra. Thus, body mass x in (1) would be replaced by length l, but our results regarding the calculation of slopes and the exponent b still hold. There is no direct length-based equivalent to the biomass size spectrum (4); calculating (4) would require first converting lengths to body masses, via species-specific allometric relationships (e.g. Shin *et al.* 2005; Trebilco *et al.* 2015).

Simulated data

We simulate a data set that consists of individual body masses of $n = 1,000$ fish. Define x_i to be the body mass (g) of fish i, where $i = 1, 2, 3, ..., n$. The 1,000 simulated values are independently drawn from the PLB distribution (1) using the inverse method (see Appendix), with $x_{\min} = 1$, $x_{\max} = 1,000$, and the exponent $b = -2$. The exponent $b = -2$ comes from the Sheldon *et al.* (1972) conjecture (Andersen and Beyer, 2006), and theoretical and empirical estimates are often close to this value (e.g. Platt and Denman, 1978; Boudreau and Dickie, 1992; Gaedke, 1992; San Martin *et al.*, 2006). Other values of $x_{\max}, b$ and n are tested later.

We use seven methods that have previously been used to estimate the slope of a size spectrum, and one that estimates the exponent b directly. We test each method on the simulated data set to obtain an estimated slope. Motivated by other ecological contexts, similar approaches were taken by White *et al.* (2008) and Edwards (2008) to test methods used to fit unbounded power-law distributions ($x_{\max} \to \infty$ in (1)), though only three of the eight size-spectra methods tested here were investigated, and neither study investigated confidence intervals, as we do here.

We then estimate b for 10,000 simulated data sets to determine the accuracy of each method and the reliability of confidence intervals. Our overall aim is to investigate whether the different methods, which sometimes differ by seemingly minor details, give consistent results. We acknowledge that authors themselves may be aware of any differences, but this is not necessarily apparent from published studies. For clarity we describe each method in the Results section in conjunction with the figure that arises from applying it to simulated data.

Results

For each method in turn (summarised in Table 1), we prescribe a name, describe the method, plot the results and give the estimated slope for the simulated data set of 1,000 values. The slope is what is usually reported, but we explain how it can be an estimate of b, $b + 1$ or $b + 2$, depending upon the method used. Thus, slopes cannot be interpreted as comparable if derived from different methods. Figure 1 is a standard histogram of the simulated data set; the y-axis has a break because so many of the counts end up in the first bin (size interval), since the data are power-law distributed.

Llin (log-linear) method

The Llin (log-linear) method involves binning the data into bins of constant width, plotting log(count of the number of individuals within a size interval) against the midpoint of the size interval, and then using linear regression to estimate the spectrum slope. Essentially the histogram in Figure 1 gets replotted as Figure 2(a) with the counts plotted on a logarithmic y-axis and the midpoints of each bin on the x-axis. Such a method was used by Daan *et al.* (2005). Note that they (and Dulvy *et al.* 2004, Boldt *et al.* 2012 and Trebilco *et al.* 2015) subtracted the midpoint of the full range of data, $(x_{\max} - x_{\min})/2$, off the midpoints of all

size intervals, in order to centre the size classes around zero. But such a constant shift does not affect the calculated value of the slope, and so for simplicity we omit it in this manuscript.

Applying the Llin method to our simulated data set estimates a slope of -0.0156. We used eight bins but two are empty (Figure 1) and so do not appear on the logarithmic scale of Figure 2(a). The use of log-linear axes suggests an exponential distribution, and so the slope cannot be related to b.

LT (log-transform) method

The LT (log-transform) method involves binning the data into bins of constant width, plotting log(count within a size interval) against log(midpoint of the size interval), then using linear regression to estimate the spectrum slope. Thus, the only difference to the Llin method is the logging of the values on the x-axis. Such a method was used on length-based data from groundfish trawl surveys by Rice and Gislason (1996) for the North Sea and Boldt *et al.* (2012) for the Eastern Bering Sea. Figure 2(b) shows the result of applying the LT method to the simulated data set, using the same eight bins (and thus counts), as in Figures 1 and 2(a). The LT method estimates a slope of -2.64, which is an estimate of b because of the logarithmic axes (White *et al.*, 2008).

LTplus1 (log-transform plus 1) method

The LTplus1 (log-transform plus 1) method is similar to the LT method, except that the count in each bin is increased by one. Dulvy *et al.* (2004) and Graham *et al.* (2005) used it to examine effects of fishing intensity on coral reef fish communities in Fiji. Their choice of $\log_{10}$ axes, rather than log axes as for the LT method, does not affect the slope (this is true for all regression-based methods – see Appendix). Consequently, $\log_{10}$(count +1 within a size interval) is plotted against against $\log_{10}$(midpoint of the size interval), and a linear

regression is fitted. Adding one to the count avoids bins with zero counts not appearing in the plots and not contributing to the regression calculation, as occurred in Figures 2(a) and (b) for the Llin and LT methods. For the LTplus1 method, Figure 2(c) has eight points (one for each bin), and the slope of the regression is -2.33, which is an estimate of b. Adding one to the counts has estimated b closer to the true value of $b = -2$, compared to the LT method's estimate of -2.64.

LBmiz ($\log_{10}$ binning plotted on log axes used in `mizer`) method

The LBmiz method involves binning the data using bins that have equal width on a $\log_{10}$ scale (e.g bin breaks of 1, 10, 100, 1000), but with the largest bin set to the same arithmetic width as the penultimate bin. It then involves plotting and fitting the regression of log(count within a size interval) against log(lower bound of the size interval). It was used in the R package `mizer` (Scott *et al.*, 2014), which simulates the potential consequences of various fishing patterns using an approach based on the McKendrick-von Foerster equation, and calculates resulting size spectra. The user specifies the number of bins, and the lower bounds of the lowest and highest bins. For our simulated data we know the minimum and maximum values of the data and can derive the bin breaks (see Appendix). Our estimated slope is -1.11. For logarithmically spaced bin breaks, as used here except for the largest bin, the slope estimates $b + 1$ (Appendix A of White *et al.* 2008), such that this method estimates $b = -2.11$. Repeating the LBmiz method using the midpoint of bins (as per the other binning methods), rather than the minimum, estimates $b = -2.13$, suggesting that the LBmiz method's use of minima is not important.

LBbiom ($\log_2$ binning with biomass in each bin plotted on $\log_{10}$ axes) method

The LBbiom method involves binning the individual fish into size intervals that have equal width on a $\log_2$ scale, and then plotting and fitting the regression of $\log_{10}$(biomass within a size interval) against $\log_{10}$(midpoint of the size interval), as used by Maxwell and Jennings (2006) for data on benthic invertebrates in the North Sea, and Jennings *et al.* (2007) for theoretical work and analyses of fish data from bottom trawl surveys. Trebilco *et al.* (2015) used it (with $\log_2$-$\log_2$ axes) to examine the role of habitat complexity on the size structure of the rockfish-dominated fish community in kelp forests off Haida Gwaii, Canada. So in contrast to the above methods based on number of fish in each bin, this method uses the total biomass in each bin and is effectively fitting the biomass spectrum rather than the ISD, though these are related via (3) and (4). Maxwell and Jennings (2006) and Jennings *et al.* (2007) used bin breaks at integer powers of two that spanned their data, and so we set bin breaks at $1, 2, 4, 8, ...$. Vidondo *et al.* (1997) described how early instruments measured numbers of particles within $\log_2$ size classes, and such binning was adopted by later scientists (even when sizes could be individually measured). We obtain an estimated *biomass* size spectrum slope of -0.0937. The biomass size spectrum (4) has exponent $b+1$ and, since logarithmically spaced bins mean the slope is the exponent plus one (White *et al.*, 2008), the slope is estimating $b + 2$, giving $b = -2.09$.

LBNbiom ($\log_2$ binning with normalised biomass in bins plotted on $\log_{10}$ axes) method

The LBNbiom (log-binning with normalisation using biomass) method is the LBbiom method but with the biomass in each bin normalised by dividing it by the bin width, i.e. plotting and fitting the regression of $\log_{10}$(biomass within a size interval divided by the

width of that size interval) against $\log_{10}$(midpoint of the size interval). Blanchard *et al.* (2005) used it to analyse groundfish survey data from the Celtic Sea, and Roy *et al.* (2011) used it (with log-log axes) to investigate temporal changes in the slope of the normalised phytoplankton biomass size spectrum for a location in the North Atlantic Ocean. Platt and Denman (1977, 1978) introduced the idea of dividing the total biomass in a size class by the width of that size class. For our simulated data set, using the same bin breaks as for the LBbiom method, the estimated biomass size spectrum slope is -1.09. This correctly estimates the biomass size spectrum (4) exponent $b+1$ because of the normalised counts (White *et al.*, 2008), giving $b = -2.09$.

LCD (log cumulative distribution) method

The LCD (log of the cumulative distribution) method requires no binning because it plots all data points. Body masses are ranked from largest (rank 1) to smallest (rank n), and $\log(\text{rank}(x)/n)$ against $\log\ x$ is plotted, with one point for each body mass x. A regression is fitted to estimate the slope. Note that $\text{rank}(x)/n$ is the fraction of values $\geq x$, which is estimating $\text{P}(X \geq x) = 1 - F(x)$, where $F(x)$ is the probability *distribution* function (Grimmett and Stirzaker, 1990) or cumulative distribution function (Bolker, 2008), and the resulting slope is approximately $b+1$ (see Appendix). Vidondo *et al.* (1997) recommended this method (for the unbounded power law), and it was recently used by Rogers *et al.* (2014) to investigate vulnerability of coral reef fisheries in The Bahamas. Figure 2(g) demonstrates this method for our data set, yielding an estimated slope of -1.04, giving $b = -2.04$.

MLE (maximum likelihood estimate) method

The MLE (maximum likelihood estimate) method directly estimates the parameter b using a standard statistical likelihood approach (e.g. Hilborn and Mangel 1997; Bolker 2008). It finds the value of b that maximises the likelihood function for the given data set. In the

context of *unbounded* power-law distributions it has been tested (e.g. Newman 2005; White *et al.* 2008; Edwards 2008), and used (together with other methods) by Arim *et al.* (2011) on body-size data from ponds in Uruguay. The MLE for b requires numerical maximisation of the log-likelihood function (see Appendix). The MLEs for $x_{\min}$ and $x_{\max}$ are the minimum and maximum observed values, respectively (Edwards *et al.*, 2012). For our data set the MLE for b is -2.03.

The MLE method does not require any plotting to estimate b. To visualise the resulting fit, in Figure 2(h) we show a rank/frequency plot which gives, on logarithmic axes, the rank of x (the number of values $\geq x$) against the value of x (e.g. Edwards 2008). We label axes using actual values (rather than log values) for easier interpretation of the results; the points in Figures 2(g) and (h) are essentially the same with the axes defined differently. The fitted PLB model (red curve) is calculated across the range of x values as $(1 - F(x))n$ using the MLE value for b, and characterises the *abundance size spectrum* based on (3); see Appendix. It is not linear because we have used the MLE method to explicitly fit a bounded power-law distribution; the fit from the LCD method in Figure 2(g) is linear because that method implicitly assumes an unbounded power-law distribution.

Summary of methods applied to the simulated data set

Overall, the *slopes* differ considerably between methods, from -2.64 to -0.02. But the slopes cannot be directly compared because they are estimating different quantities. Translating the slopes into estimates of b means that five of the methods estimate b in the range $(-2.11, -2.03)$, just below the true value of $b = -2$.

While some of the above differences in what each method calculates will have been appreciated by some authors, it is not always clear that subtle methodological differences are important. For example, Daan *et al.* (2005) initially talk about the 'slope of the log-linear size spectrum of the total fish community' (i.e. the Llin method), and then mention

Rice and Gislason (1996) as having showed that the spectrum slope for a North Sea fish community had steepened over time. However, Rice and Gislason (1996) used the LT method. Thus, spectrum slopes were being defined using different methods and so cannot be considered comparable.

Repeated simulations – accuracy of the methods

The above results depend on the single simulated data set of $n = 1,000$ random numbers drawn from the PLB distribution (1). To build a more detailed picture of the accuracy of each method, we now repeat the above calculations on 10,000 independent simulated samples (a number recommended by Crawley 2002), each containing 1,000 values drawn randomly from the PLB distribution (still with $b = -2$, $x_{\min} = 1$ and $x_{\max} = 1,000$). So for each method we obtain 10,000 estimates of b (or slope for the Llin method). For the MLE method, $x_{\min}$ and $x_{\max}$ are explicitly estimated as the minimum and maximum data values, respectively, for each of the 10,000 samples.

The resulting estimates of b are shown in the blue histograms in Figure 3, with summary statistics in Table 2. The Llin method gives a narrow range of slopes that are just below zero, which is intuitive when looking at the scales of the axes in Figure 2(a). The distribution of estimates of b for the LT and LTplus1 methods are fairly wide and highly biased (Figures 3(b) and (c)), with 99% and 82%, respectively, of the estimates being below the true value of $b = -2$ (Table 2).

For the remaining five methods the means and medians of the estimates are all within 0.01 of the true value of b (Table 2), with LBmiz having 47% of the estimates below the true value, which is the closest any of the methods get to the desired value of 50% (equally likely to be above or below the true value). The LBmiz, LBiom and LBNbiom methods show similar distributions, with the LCD and then MLE methods having progressively narrower distributions. Thus, overall, the final five methods appear to be fairly accurate, with MLE

showing the least variation.

The shaded gold histograms in Figure 3 show the same analyses but with $x_{\max} = 10,000$ (rather than $x_{\max} = 1,000$). Such a 10,000-fold range of body sizes can be observed for coral-reef fish (JPWR, personal observation). The results for the MLE method remain essentially unchanged from the $x_{\max} = 1,000$ results, while the accuracy of some of the other methods is diminished. For example, for the LBNbiom method the distribution of estimated b values shifts to the right in Figure 3(f), such that only 20% (rather than 45%) of the estimated values fall below the known value of -2. See Appendix for full details.

Confidence intervals

The previous results estimate b using the different methods. Bolker (2008) states that such types of best-fit estimates require some measurement of uncertainty to be meaningful. However, uncertainty of slopes has only been occasionally calculated in previous studies (e.g. Rice and Gislason 1996, Graham *et al.* 2005), a situation that is 'particularly unsettling' (Rice, 2000). Therefore we now construct confidence intervals of b for each method, and test how well they perform.

For the regression-based methods, a confidence interval for the slope can be calculated (e.g. Crawley 2002) using the R command `confint`. The confidence interval for b can then be obtained by subtracting one or two as appropriate for each method (see Table 2). For the MLE method, a 95% confidence interval for b can be calculated using the profile likelihood-ratio test (Hilborn and Mangel, 1997).

By definition, 95% of the 95% confidence intervals should contain the true value of the estimated quantity. To see if this holds, for each method we compute a confidence interval for b for each of the 10,000 simulated data sets (with $x_{\max} = 1,000$), and see what percentage of a method's intervals contain the true value of $b = -2$. This percentage is the 'observed coverage' and should ideally equal the 'nominal coverage' of 95% (Bolker, 2008).

Figure 4 shows the resulting confidence intervals for subsamples of the 10,000 simulated data sets; we use subsamples for clarity (see Appendix). For each method the true value of b is shown as a vertical red line, and each confidence interval is coloured grey if it encompasses the true value and blue if it does not. Thus, we would expect 95% of the intervals to be grey and 5% to be blue. The resulting percentage (the observed coverage) based on all 10,000 confidence intervals is indicated for each method.

Figure 4(a) shows that the confidence intervals of the slope for the Llin method never include the true value of b. The confidence intervals of b for the LT and LTplus1 methods are so wide that they essentially always include the true value (Figures 4(b) and (c)); such intervals are therefore not of practical use. For the LBmiz, LBbiom and LBNbiom methods, the confidence intervals include the true value of b only 90% of the time (Figure 4), thereby overstating their reliability. For the LCD method, only 6% of the confidence intervals include the true value of b because the intervals are very narrow (Figure 4(g)). Intuitively, such narrow intervals can be inferred from Figure 2(g) – the regression line is being fitted to all $n = 1,000$ points, and there is clearly not a large possible range in the slope (compared to, say, Figure 2(e)). Thus, the very narrow confidence intervals from the LCD method give a misleading impression of accuracy.

For the MLE method, 95% of the confidence intervals include the true value of b (Figure 4(h)). The intervals are of a relatively consistent width, which is an intuitively desirable property that is lacking for the other methods.

With $x_{\max} = 10,000$, the observed coverage declines from 90% to 84% (LBmiz method) and 74% (LBbiom and LBNbiom methods), and remains at 6% for the LCD method and at the desired 95% for the MLE method (see Appendix). Thus, overall we find the MLE method to be the only one that produces reliable estimators of the uncertainty of b.

Sensitivity analyses – robustness of the MLE method

In the Appendix we modify the MLE method to fix $x_{\max}$ across the 10,000 data sets rather than estimating it individually for each data set, which gives only minor numerical differences in results. We also repeat our main simulations with $b = -2.5$, $b = -1.5$ and $b = -0.5$ instead of $b = -2$, and with a ten-fold increase in sample size to $n = 10,000$. The conclusions for most methods are sensitive to the value of b or n (e.g. the LBNbiom method performs worse with $b = -2.5$). However, only the conclusions for the MLE method are robust – estimates of b are accurate and confidence intervals are reliable (observed coverage of 94% or 95%), unlike for other methods. We also find our results and conclusions are not dependent on the seed used for the random-number generator.

MLEbin method for binned data

Sometimes data (or model output, Thorpe *et al.* 2015) are only available in binned form. We extend the MLE method for such data sets to give the MLEbin method (see Appendix). We test it using the same 10,000 simulated data sets as earlier, but first binning each data set (using bin breaks at 1, 2, 4, 8, ...) and then applying the method to the counts in each bin. The MLEbin method appears as accurate as the MLE method (Table 2 and Figure 5). Sensitivity analyses (e.g. regarding binning) will be conducted in future work. Researchers can adapt our code for their particular data sets, and also investigate different binning protocols for data that require binning when being collected.

Discussion

We have expanded upon White *et al.*'s (2007) recommendation to think of size spectra in terms of ISDs, because it places such work in the context of probability densities. Our results show that the slopes of size spectra arising from commonly-used methods cannot be

interpreted as equivalent since they do not all directly estimate the exponent b of the ISD, and that the methods estimate b with different levels of accuracy. We recommend the MLE method for estimating b and its confidence intervals, since only its performance was robust under sensitivity analyses. This is in contrast to Vidondo *et al.*'s (1997) recommendation to use the LCD method over the MLE method (based on unpublished simulations for unbounded power laws).

The MLE method avoids binning and regression. Binning in general can be problematic (for example, if a data set has no body masses < 10 g but the lowest bin is defined as 8-16 g), and the choice of bin widths can affect the estimated slope (Vidondo *et al.*, 1997). Regression-based methods are problematic because the intercept and the slope implicitly determine $x_{\min}$, which can erroneously be greater than some data values (James *et al.*, 2011). They also assume that the sampling errors in the logarithmic counts for each bin have the same variance, which may not be justified. Although regression can be understood in a likelihood context, this is different to explicitly using a likelihood-based method (Edwards *et al.*, 2012).

However, researchers are used to seeing biomass size spectra in the form of log-log plots of the normalised biomass in logarithmic bins, as in Figure 2(f). Thus, we recommend presenting results as the two plots in Figure 6 – a biomass size spectrum and an abundance size spectrum, with the MLE estimate for b (and bounds of the 95% confidence interval) used in (4) for biomass and (3) for abundance.

Rice (2000) called for an objective way to determine if differences among values of a community metric are meaningful. The calculation of reliable confidence intervals for b will allow this. Furthermore, quantifying the uncertainty in b should improve the quality of advice to fisheries or ecosystem managers, because without uncertainty numerical results can give a misleading impression of accuracy. Uncertainty can be accounted for when investigating changes in b (e.g. using weighted linear regression) that could represent steepening

of the size spectra in response to fishing.

We can only partially determine the consequences of our results for previous conclusions. For example, Dulvy *et al.* (2004) found a significant relationship between size spectrum slopes and fishing intensity across 13 fishing grounds. The slopes were all between -0.04 and -0.01, derived using the LTplus1 method. However, Figure 3(c) suggests that such a small change in size spectrum slope could be an artefact of the LTplus1 method. In general, previously calculated slopes must be interpreted with respect to the method used.

We have used a bounded power-law distribution for the ISD since power laws are commonly-used models for size spectra (Platt and Denman, 1978; Boudreau and Dickie, 1992; Andersen and Beyer, 2006). However, we echo Vidondo *et al.*'s (1997) warning that there will be datasets for which power-law distributions are not appropriate. Dynamic models of size spectra in marine communities predict non-power-law size distributions at the level of individual species (Hartvig *et al.*, 2011; Jacobsen *et al.*, 2014; Law *et al.*, 2014), although the aggregate community ISD may be closer to a power law (Andersen and Beyer, 2006). We have compared different methods for estimating the exponent b *on the common assumption that the ISD is a power law.* In applications, the validity of this assumption could be investigated using goodness-of-fit tests and Akaike Information Criteria (e.g. Edwards 2008).

The likelihood approach could be further developed to incorporate measurement errors that researchers are aware of in their data sets. Our current results (and R code) have application in ecology beyond size spectra, since power-law distributions arise in several areas (White *et al.*, 2008).

Our take-home messages are: (i) size spectra should be formally expressed in terms of individual size distributions, (ii) the MLE method should be used to estimate the ISD exponent b and its confidence intervals, (iii) there is no need to bin data, but if data are only available in binned form then the MLEbin method can be used and tested. We hope

that these will be adopted and applied in size spectra research. To facilitate this, we have formalised the mathematics used to analyse size spectra, tested the methods, and provided usable R code for researchers.

Acknowledgements

For useful discussions we thank Finlay Scott, Jennifer Boldt, Carrie Holt, Jean-Baptiste Lecomte, Brooke Davis and Rowan Haigh. We thank three anonymous reviewers for their insightful comments that have improved this work. This project arose from a sabbatical visit by MJP to AME. JLB was supported by the UK's Natural Environment Research Council and Department for Environment, Food and Rural Affairs (grant number NE/L003279/1, Marine Ecosystems Research Programme).

Data Accessibility

For complete reproducibility of our work we provide all R code (R version 3.1.0) as online Supporting Information. We have functionalised and documented the code with the aim of it being used by researchers for their own data (e.g. to produce an equivalent Figure 6). Furthermore, to share future enhancements the code is available at `https://github.com/andrew-edwards`. The Supporting Information also includes an Appendix of extended methods and results.

References

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

Arim, M., Berazategui, M., Barreneche, J. M., Ziegler, L., Zarucki, M., and Abades, S. R. (2011). Determinants of density-body size scaling within food webs and tools for their detection. *Advances in Ecological Research*, **45**, 1–39.

Bianchi, G., Gislason, H., Graham, K., Hill, L., Jin, X., Koranteng, K., Manickchand-Heileman, S., Payá, I., Sainsbury, K., Sanchez, F., and Zwanenburg, K. (2000). Impact of fishing on size composition and diversity of demersal fish communities. *ICES Journal of Marine Science*, **57**, 558–571.

Blanchard, J. L., Dulvy, N. K., Jennings, S., Ellis, J. R., Pinnegar, J. K., Tidd, A., and Kell, L. T. (2005). Do climate and fishing influence size-based indicators of Celtic Sea fish community structure? *ICES Journal of Marine Science*, **62**, 405–411.

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

Boldt, J. L., Bartkiw, S. C., Livingston, P. A., Hoff, G. R., and Walters, G. E. (2012). Investigation of fishing and climate effects on the community size spectra of eastern Bering Sea fish. *Transactions of the American Fisheries Society*, **141**, 327–342.

Bolker, B. (2008). *Ecological Models and Data in R*. Princeton University Press, Princeton and Oxford.

Boudreau, P. R. and Dickie, L. M. (1992). Biomass spectra of aquatic ecosystems in relation to fisheries yield. *Canadian Journal of Fisheries and Aquatic Sciences*, **49**, 1528–1538.

Crawley, M. J. (2002). *Statistical Computing: An Introduction to Data Analysis using S-Plus*. John Wiley & Sons Ltd., Chichester, UK.

493 Daan, N., Gislason, H., Pope, J. G., and Rice, J. C. (2005). Changes in the North Sea fish
494 community: evidence of indirect effects of fishing? *ICES Journal of Marine Science*, **62**,
495 177–188.

496 Dulvy, N. K., Polunin, N. V. C., Mill, A. C., and Graham, N. A. J. (2004). Size structural
497 change in lightly exploited coral reef fish communities: evidence for weak indirect effects.
498 *Canadian Journal of Fisheries and Aquatic Sciences*, **61**, 466–475.

499 Edwards, A. M. (2008). Using likelihood to test for Lévy flight search patterns and for
500 general power-law distributions in nature. *Journal of Animal Ecology*, **77**, 1212–1222.

501 Edwards, A. M. (2011). Overturning conclusions of Lévy flight movement patterns by
502 fishing boats and foraging animals. *Ecology*, **92**, 1247–1257.

503 Edwards, A. M., Phillips, R. A., Watkins, N. W., Freeman, M. P., Murphy, E. J., Afanasyev,
504 V., Buldyrev, S. V., da Luz, M. G. E., Raposo, E. P., Stanley, H. E., and Viswanathan,
505 G. M. (2007). Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees
506 and deer. *Nature*, **449**, 1044–1048.

507 Edwards, A. M., Freeman, M. P., Breed, G. A., and Jonsen, I. D. (2012). Incorrect likelihood
508 methods were used to infer scaling laws of marine predator search behaviour. *PLoS ONE*,
509 **7**, e45174.

510 Gaedke, U. (1992). The size distribution of plankton biomass in a large lake and its seasonal
511 variability. *Limnology and Oceanography*, **37**, 1202–1220.

512 Graham, N. A. J., Dulvy, N. K., Jennings, S., and Polunin, N. V. C. (2005). Size-spectra
513 as indicators of the effects of fishing on coral reef fish assemblages. *Coral Reefs*, **24**,
514 118–124.

Grimmett, G. R. and Stirzaker, D. R. (1990). *Probability and Random Processes*. Oxford University Press, Oxford.

Hartvig, M., Andersen, K. H., and Beyer, J. E. (2011). Food web framework for size-structured populations. *Journal of Theoretical Biology*, **272**, 113–122.

Hilborn, R. and Mangel, M. (1997). *The Ecological Detective: Confronting Models with Data*. Vol. 28, Monographs in Population Biology, Princeton University Press, New Jersey.

Jacobsen, N. S., Gislason, H., and Andersen, K. H. (2014). The consequences of balanced harvesting of fish communities. *Proceedings of the Royal Society, Series B*, **281**, 20132701.

James, A., Plank, M. J., and Edwards, A. M. (2011). Assessing Lévy walks as models of animal foraging. *Journal of the Royal Society Interface*, **8**, 1233–1247.

Jennings, S., de Oliveira, J. A. A., and Warr, K. J. (2007). Measurement of body size and abundance in tests of macroecological and food web theory. *Journal of Animal Ecology*, **76**, 72–82.

Jennings, S., Mélin, F., Blanchard, J. L., Forster, R. M., Dulvy, N. K., and Wilson, R. W. (2008). Global-scale predictions of community and ecosystem properties from simple ecological theory. *Proceedings of the Royal Society, Series B*, **275**, 1375–1383.

Law, R., Plank, M. J., and Kolding, J. (2012). On balanced exploitation of marine ecosystems: results from dynamic size spectra. *ICES Journal of Marine Science*, **69**, 602–614.

Law, R., Plank, M. J., and Kolding, J. (2014). Balanced exploitation and coexistence of interacting, size-structured, fish species. *Fish and Fisheries*, page doi: 10.1111/faf.12098.

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

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, **46**, 323–351.

Page, R. (1968). Aftershocks and microaftershocks of the great Alaska earthquake of 1964. *Bulletin of the Seismological Society of America*, **58**, 1131–1168.

Platt, T. and Denman, K. (1977). Organisation in the pelagic ecosystem. *Helgoländer wissenschaftliche Meeresunters*, **30**, 575–581.

Platt, T. and Denman, K. (1978). The structure of pelagic marine ecosystems. *Rapport et Procès-verbaux des Réunions Conseil International pour l'Exploration de la Mer*, **173**, 60–65.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reuman, D. C., Mulder, C., Raffaelli, D., and 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.

Reuman, D. C., Gislason, H., Barnes, C., Mélin, F., and Jennings, S. (2014). The marine diversity spectrum. *Journal of Animal Ecology*, **83**, 963–979.

Rice, J. and Gislason, H. (1996). Patterns of change in the size spectra of numbers and diversity of the North Sea fish assemblage, as reflected in surveys and models. *ICES Journal of Marine Science*, **53**, 1214–1225.

Rice, J. C. (2000). Evaluating fishery impacts using metrics of community structure. *ICES Journal of Marine Science*, **57**, 682–688.

Rogers, A., Blanchard, J. L., and Mumby, P. J. (2014). Vulnerability of coral reef fisheries to a loss of structural complexity. *Current Biology*, **24**, 1000–1005.

561 Roy, S., Platt, T., and Sathyendranath, S. (2011). Modelling the time-evolution of phy-
562 toplankton size spectra from satellite remote sensing. *ICES Journal of Marine Science*,
563 **68**, 719–728.

564 San Martin, E., Irigoien, X., Harris, R. P., López-Urrutia, Á., Zubkov, M. V., and Heywood,
565 J. L. (2006). Variation in the transfer of energy in marine plankton along a productivity
566 gradient in the Atlantic Ocean. *Limnology and Oceanography*, **51**, 2084–2091.

567 Scott, F., Blanchard, J. L., and Andersen, K. H. (2014). *mizer*: an R package for mul-
568 tispecies, trait-based and community size spectrum ecological modelling. *Methods in*
569 *Ecology and Evolution*, **5**, 1121–1125.

570 Sheldon, R. W. and Parsons, T. R. (1967). A continuous size spectrum for particulate
571 matter in the sea. *Journal of the Fisheries Research Board of Canada*, **24**, 909–915.

572 Sheldon, R. W., Prakash, A., and Sutcliffe Jr., W. H. (1972). The size distribution of
573 particles in the ocean. *Limnology and Oceanography*, **17**, 327–340.

574 Shin, Y.-J., Rochet, M.-J., Jennings, S., Field, J. G., and Gislason, H. (2005). Using size-
575 based indicators to evaluate the ecosystem effects of fishing. *ICES Journal of Marine*
576 *Science*, **62**, 384–396.

577 Thorpe, R. B., Le Quesne, W. J. F., Luxford, F., Collie, J. S., and Jennings, S. (2015).
578 Evaluation and management implications of uncertainty in a multispecies size-structured
579 model of population and community responses to fishing. *Methods in Ecology and Evo-*
580 *lution*, **6**, 49–58.

581 Trebilco, R., Dulvy, N. K., Stewart, H., and Salomon, A. K. (2015). The role of habitat
582 complexity in shaping the size structure of a temperate reef fish community. *Marine*
583 *Ecology Progress Series*, **532**, 197–211.

584 Vidondo, B., Prairie, Y. T., Blanco, J. M., and Duarte, C. M. (1997). Some aspects of the
585 analysis of size spectra in aquatic ecology. *Limnology and Oceanography*, **42**, 184–192.

586 White, E. P., Ernest, S. K. M., Kerkhoff, A. J., and Enquist, B. J. (2007). Relationships
587 between body size and abundance in ecology. *Trends in Ecology and Evolution*, **22**,
588 323–330.

589 White, E. P., Enquist, B. J., and Green, J. L. (2008). On estimating the exponent of
590 power-law frequency distributions. *Ecology*, **89**, 905–912.

Table 1: Brief description of methods used to estimate the slope of a size spectrum. Two of the examples use a different logarithmic base for the regression fit to that stated, but this does not affect the estimated slope (see text).

Name	Brief description	Example reference(s)
Llin	Log-linear transform. Plot linearly binned data on log-linear axes then fit regression of log(count in bin) against midpoint of bin.	Daan *et al.* (2005)
LT	Log transform. Plot linearly binned data on log-log axes then fit regression of log(count in bin) against log(midpoint of bin).	Rice and Gislason (1996), Boldt *et al.* (2012)
LTplus1	Log transform plus 1. Plot linearly binned data on $\log_{10}$-$\log_{10}$ axes then fit regression of $\log_{10}(\text{count} + 1)$ against $\log_{10}$(midpoint of bin).	Dulvy *et al.* (2004), Graham *et al.* (2005)
LBmiz	Logarithmic binning as done by `mizer`. Bin data using $\log_{10}$ bins (but with largest bin the same arithmetic size as the penultimate bin), and regression of log(count in bin) against log(lower bound of bin).	Scott et al.'s (2014) `mizer` R package.
LBbiom	Logarithmic binning and then fit biomass size spectrum. Bin sizes using $\log_2$ bins then fit regression of $\log_{10}$(biomass in bin) against $\log_{10}$(midpoint of bin).	Maxwell and Jennings (2006), Jennings *et al.* (2007), Trebilco *et al.* (2015).
LBNbiom	Logarithmic binning with normalisation and then fit biomass size spectrum. Bin sizes using $\log_2$ bins, then fit regression of $\log_{10}$(biomass in bin divided by bin width) against $\log_{10}$(midpoint of bin).	Blanchard *et al.* (2005), Roy *et al.* (2011)
LCD	Logarithmic plotting of $1 - F(x)$; i.e. one minus the cumulative distribution. Rank data from largest (rank 1) to smallest (rank n), fit regression of $\log(\text{rank}(x)/n)$ against $\log\ x$.	Vidondo *et al.* (1997), Rogers *et al.* (2014)
MLE	Maximum likelihood estimate. No binning or plotting necessary. Calculate the maximum likelihood estimate of the parameter b. Data and fitted distribution can be plotted on a rank/frequency plot.	Arim *et al.* (2011)

Table 2: Summary statistics for each method for the 10,000 simulations of 1,000 samples from (1), corresponding to the blue histograms ($x_{\max} = 1,000$) in Figure 3. Second column indicates how the fitted slope can be translated into an estimate of b, though for the MLE method b is estimated directly. Statistics relate to the resulting estimates of b (or slope for Llin method), with the final column giving the percentage of simulations for which the estimate is below the true value of $b = -2$. See the end of the Results for the MLEbin method.

Method	Slope represents	5% quantile	Median	Mean	95% quantile	Percentage below -2
Llin	–	-0.02	-0.01	-0.01	-0.01	0
LT	b	-2.88	-2.42	-2.44	-2.09	99
LTplus1	b	-2.66	-2.20	-2.23	-1.90	82
LBmiz	$b+1$	-2.11	-2.00	-2.00	-1.89	47
LBbiom	$b+2$	-2.11	-1.99	-1.99	-1.89	45
LBNbiom	$b+1$	-2.11	-1.99	-1.99	-1.89	45
LCD	$b+1$	-2.08	-2.01	-2.01	-1.95	59
MLE	b	-2.05	-1.99	-2.00	-1.94	44
MLEbin	b	-2.05	-2.00	-2.00	-1.94	46

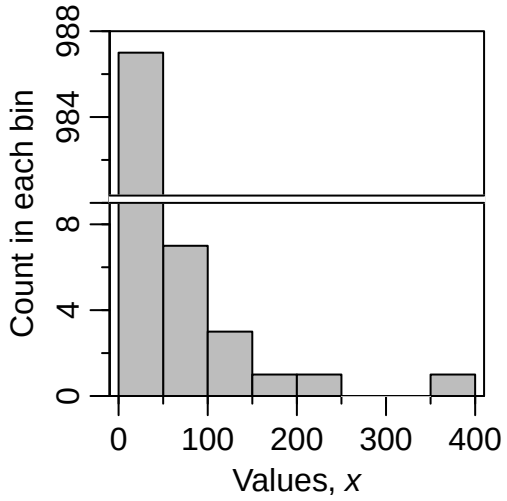

Figure 1: Standard histogram of a random sample of 1,000 values from a bounded power-law distribution (1) with $b = -2$, $x_{\min} = 1$ and $x_{\max} = 1,000$. Histogram shows the number of counts within each of the eight equally sized bins. Note the break in the y-axis to clearly show all the counts.

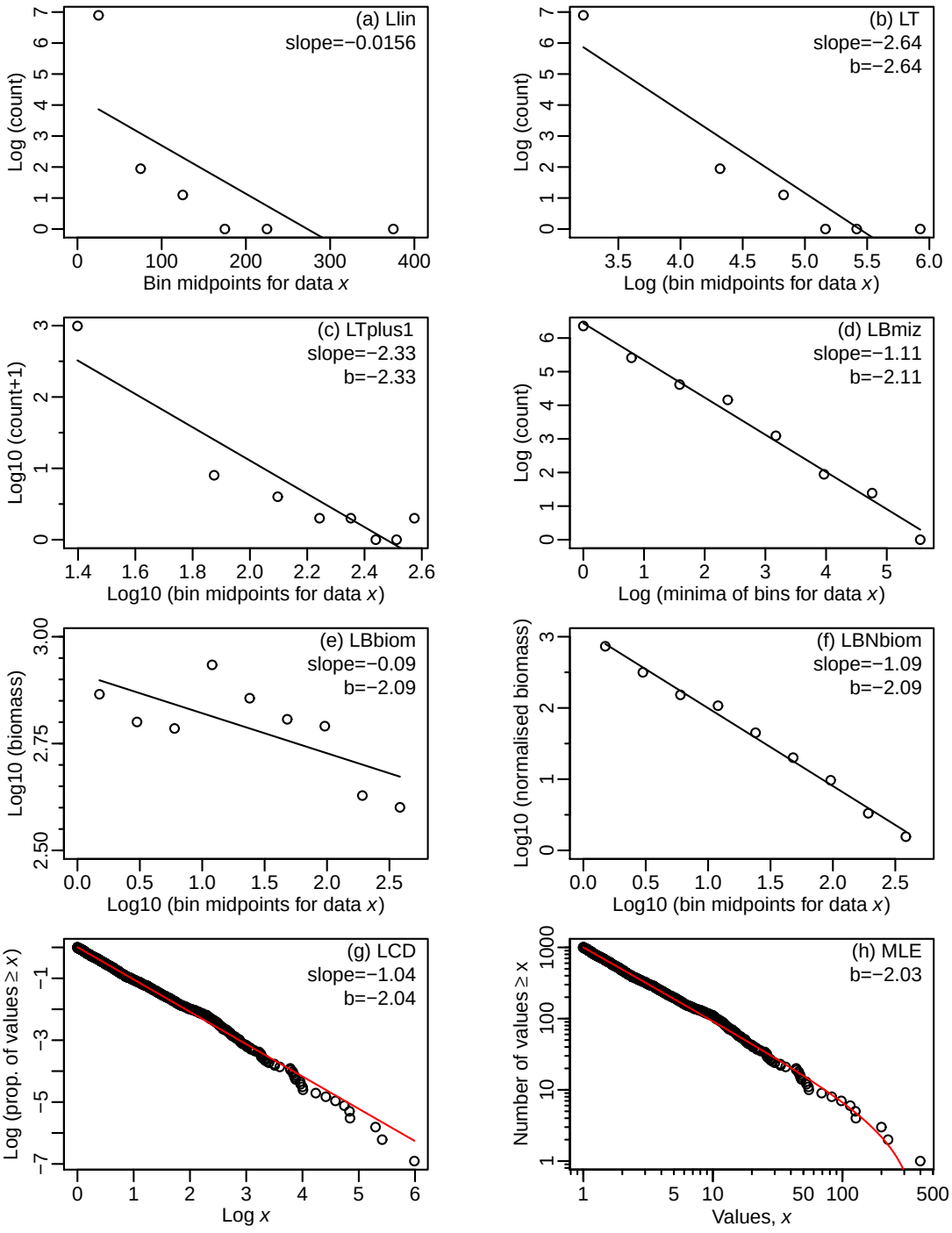

Figure 2: Results from using eight methods (Table 1) to estimate the slope or exponent of size spectra from the simulated data set of 1,000 values shown in Figure 1. The estimated slope and/or the estimated value of the ISD exponent b is given for each method, with lines showing the resulting fitted size spectra.

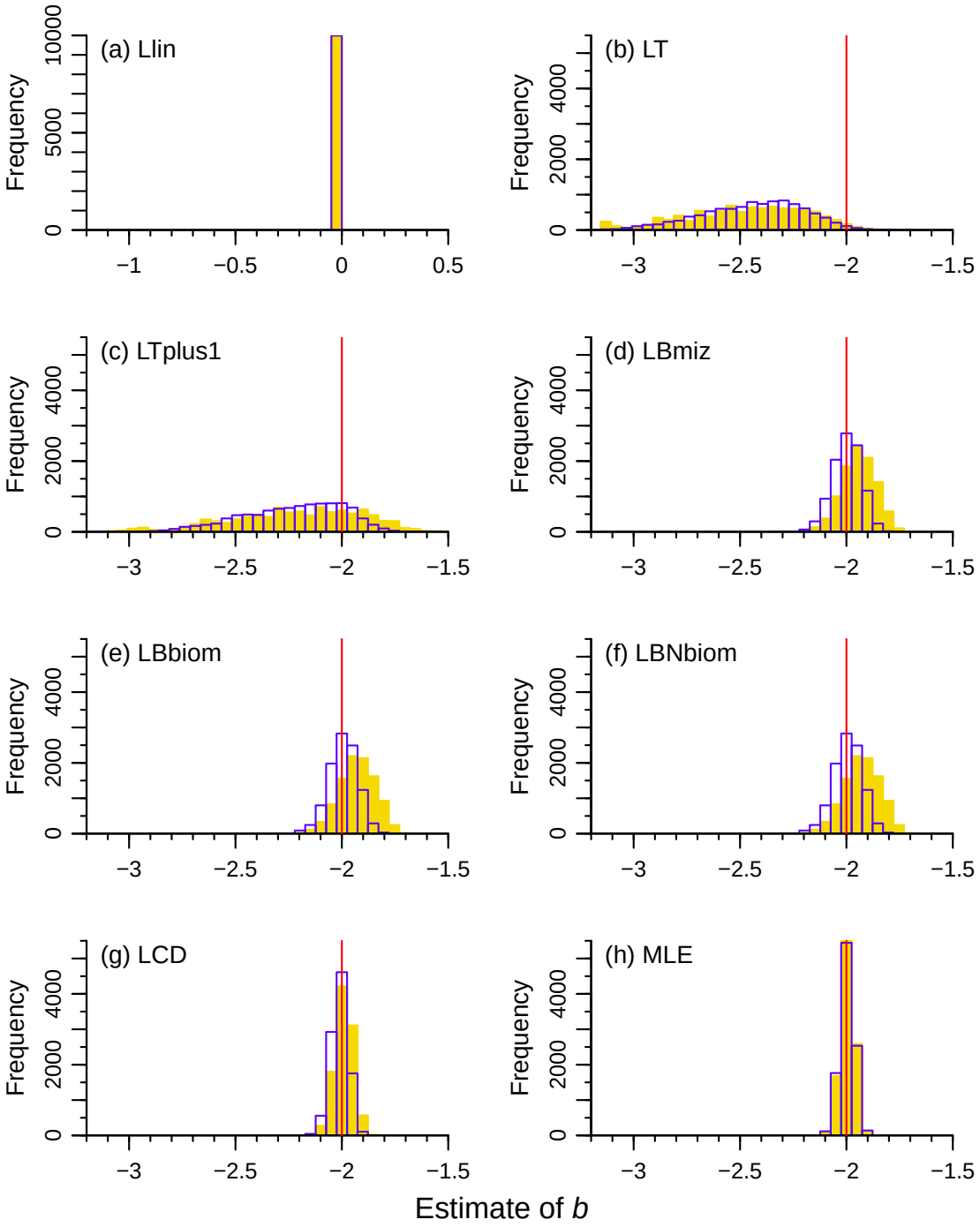

Figure 3: Histograms (in blue) of estimated exponent b for 10,000 simulated data sets, each of which contains 1,000 independent random numbers drawn from a bounded power-law distribution with $b = -2$, $x_{\min} = 1$ and $x_{\max} = 1,000$. Each panel uses the method from the corresponding panel in Figure 2. The vertical red lines indicate the known value of $b = -2$. Shaded gold histograms show results when setting $x_{\max} = 10,000$. Axes scales are the same for all panels except (a), which gives estimates of slope since the Llin method does not estimate b.

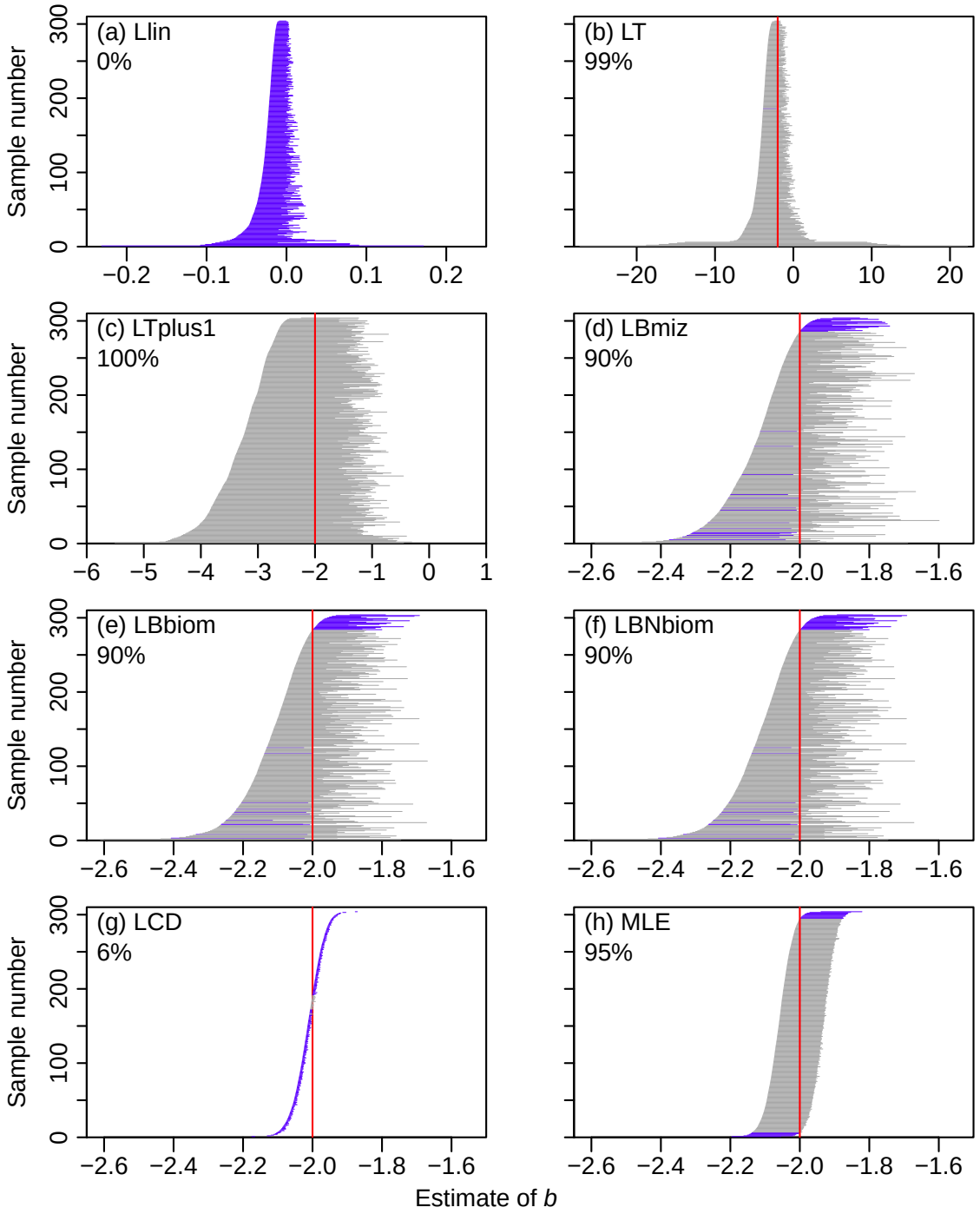

Figure 4: Confidence intervals (horizontal lines) of b obtained for each method for subsamples of the 10,000 simulated data sets (with $x_{\max} = 1,000$) used in Figure 3. For each numbered subsample on the y-axis, the 95% confidence interval of b obtained using the respective method is plotted as a horizontal line, which is coloured grey if the interval includes the true value of $b = -2$ (given by the vertical red line) or blue if it does not. Simulations are sorted in ascending order of their lower bound. The percentage for each method gives the observed coverage, namely the percentage of all 10,000 simulated data sets for which the 95% confidence interval contains the true value of b; by definition, this should ideally be 95%. Horizontal axes are the same for (d)-(h), and (a) shows confidence intervals of the slope.

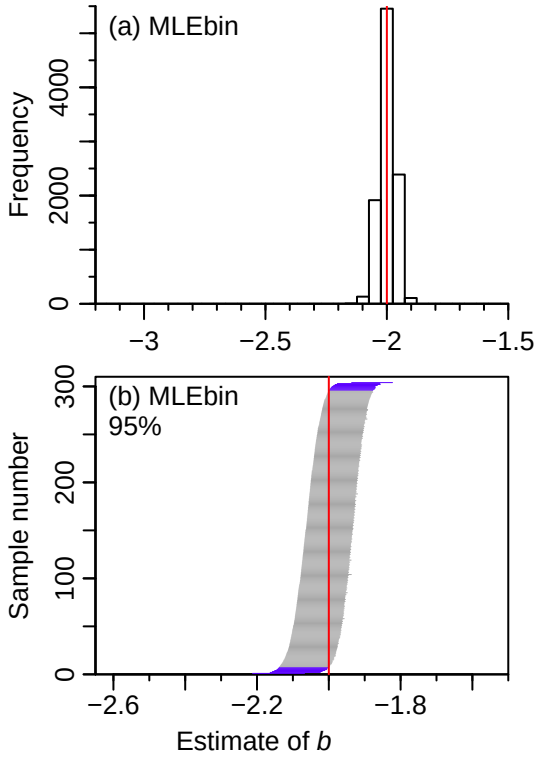

Figure 5: Testing of the MLEbin method on binned versions of each of the simulated data sets; plots as in Figures 3 and 4. Confidence intervals range in width from 0.114-0.145, only marginally wider than the range of 0.114-0.143 for the MLE method.

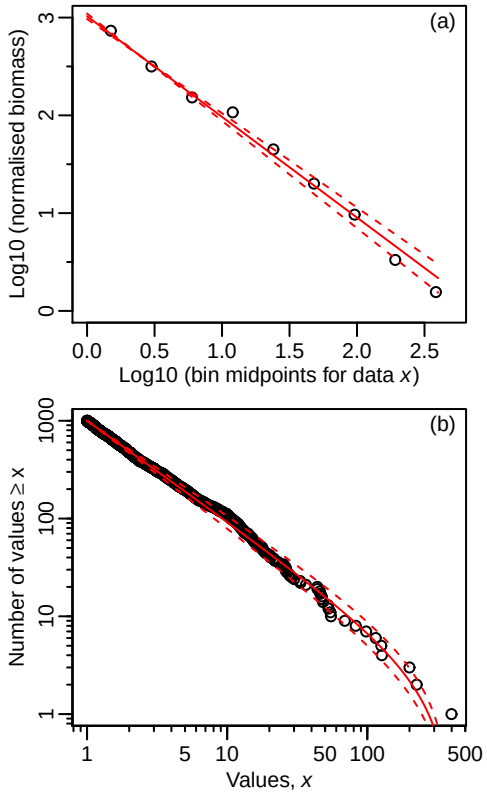

Figure 6: Suggested plots of (a) biomass size spectrum from equation (4), and (b) abundance size spectrum from equation (3), fitted using the maximum likelihood estimate -2.03 of the exponent b (red solid lines). Data are from Figure 2. For the biomass size spectrum, data are binned and normalised as for the LBNbiom method (Blanchard *et al.*, 2005). Dashed lines are from using the lower and upper bounds (-2.10 and -1.96) of the 95% confidence interval of b.