Accounting for the bin structure of data removes bias when fitting size spectra

Andrew M. Edwards[1,2,*]**, James P. W. Robinson**[2,3]**, Julia L. Blanchard**[4]**,**
Julia K. Baum[2]**, Michael J. Plank**[5,6]

[1]**Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, British Columbia V9T 6N7, Canada**
[2]**Department of Biology, University of Victoria, Victoria, British Columbia V8W 2Y2, Canada**
[3]**Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YW, UK**
[4]**Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, TAS 7001, Australia**
[5]**School of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand**
[6]**Te Pūnaha Matatini, a New Zealand Centre of Research Excellence, University of Auckland, Auckland 1011, New Zealand**

ABSTRACT: Size spectra are recommended tools for detecting the response of marine communities to fishing or to management measures. A size spectrum succinctly describes how a property, such as abundance or biomass, varies with body size in a community. Required data are often collected in binned form, such as numbers of individuals in 1 cm length bins. Numerous methods have been employed to fit size spectra, but most give biased estimates when tested on simulated data, and none account for the data's bin structure (breakpoints of bins). Here, we used 8 methods to fit an annual size-spectrum exponent, b, to an example data set (30 yr of the North Sea International Bottom Trawl Survey). The methods gave conflicting conclusions regarding b declining (the size spectrum steepening) through time, and so any resulting advice to ecosystem managers will be highly dependent upon the method used. Using simulated data, we showed that ignoring the bin structure gives biased estimates of b, even for high-resolution data. However, our extended likelihood method, which explicitly accounts for the bin structure, accurately estimated b and its confidence intervals, even for coarsely collected data. We developed a novel visualisation method that accounts for the bin structure and associated uncertainty, provide recommendations concerning different data types and have created an R package (sizeSpectra) to reproduce all results and encourage use of our methods. This work is also relevant to wider applications where a power-law distribution (the underlying distribution for a size spectrum) is fitted to binned data.

KEY WORDS: Abundance size spectrum · Biomass size spectrum   Individual size distribution · Ecosystem-based fisheries management · Ecosystem indicators · Truncated Pareto distribution · Power-law distribution

1. INTRODUCTION

Ecosystem-based fisheries management requires general indicators to accurately assess ecosystem states across space and time (Jennings 2005, Shin et al. 2005). Size spectra quantify how a property (such as abundance or biomass) varies with body size in a community (Rice & Gislason 1996), leading to a variety of applications. They are readily fitted to simple body-size data, and have therefore been widely applied to aquatic ecosystems (e.g. Dulvy et al. 2004, Daan et al. 2005, Boldt et al. 2012). A new application involves using size-spectra models and simulated carbon fluxes from an ocean-biogeochemistry model to estimate how global benthic communities may respond to climate change and ocean acidification (Kelly-Gerreyn et al. 2014, Yool et al. 2017). Size spectra are considered candidate 'ecosystem Essential

*Corresponding author: andrew.edwards@dfo-mpo.gc.ca

Ocean Variables' for monitoring changes of phytoplankton, zooplankton, benthic invertebrates, krill and fish in the Southern Ocean (Constable et al. 2016).

The slope of the size spectrum is a recommended ecosystem indicator because of its apparent responsiveness to fishing impacts. Piet & Jennings (2005) tested the response to fishing of 8 potential fish community indicators and found that only the slope of the biomass size spectrum reliably detected the effects of spatial management measures (partial closure of an area of the North Sea). Jennings & Dulvy (2005) similarly recommended the use of size spectra to provide surveillance of the status of fish communities. They pointed out difficulties in justifying reference points (targets to be aimed for and limits to be avoided), and instead recommended reference directions that monitor trends in size-spectrum slopes through time. Such trends are commonly investigated in size-spectrum studies (e.g. Blanchard et al. 2005, Boldt et al. 2012). Using 188 plausible multispecies size-structured fish community models, Thorpe et al. (2015) found that the slope of the abundance size spectrum was the most responsive indicator of the 4 indicators tested. Size-spectra parameters have been selected by the European Union as one of several indicators used to evaluate sea-floor integrity (European Commission 2010, Rice et al. 2012), and size-spectrum models continue to be widely applied (e.g. Taniguchi et al. 2014, Álvarez et al. 2016, Stasko et al. 2016, Mindel et al. 2018, van Gemert & Andersen 2018, Woodson et al. 2018, Zhang et al. 2018, Zhou et al. 2019). Such applications require consistent calculation methods, especially when used for providing practical advice to fisheries (or ecosystem) managers. However, most commonly employed methods for fitting size spectra have been shown to give biased estimates and inaccurate confidence intervals for simulated data (Edwards et al. 2017), underscoring the need for close evaluation of methods.

Data used in size-spectra studies are often only collected in binned form, whereby the 'bin structure' (breakpoints of the bins; Monnahan et al. 2016) is determined during data collection. For example, Daan et al. (2005) analysed fish lengths collected in bins of 0.5, 1 and 5 cm widths. For benthic invertebrates in the North Sea, Maxwell & Jennings (2006) sorted individuals into sixteen $\log_2$ body-mass bins, resulting in counts and total biomass for each bin. Similarly, in underwater visual surveys, fish lengths might be recorded to the nearest 1 cm (McCoy et al. 2016) or 5 cm (Richards et al. 2011),

generating size-frequency data of counts per length bin. For historical data, digitisation of points from a published figure may be all that is available, in which case the bin structure would have to be algebraically determined (e.g. Edwards 2011 in a related context).

Alternatively, data are sometimes binned during analysis. For example, Dulvy et al. (2004) and Boldt et al. (2012) binned data into 5 cm length bins to calculate size spectra for fish communities in Fiji and the eastern Bering Sea, respectively. Also, some size-spectra calculation methods require binning of data to a prescribed resolution (Edwards et al. 2017).

Here, we investigate how the bin structure of data affects the fitting of size spectra, and whether the bin structure needs to be explicitly accounted for. To motivate our work and understand issues that can arise when fitting size spectra to a complex binned data set, we first apply 8 existing methods to fit annual size spectra to 30 yr of data from the North Sea International Bottom Trawl Survey (IBTS). The data consist of the numbers of individuals of each species within each length bin that were caught per hour of trawling. We use the IBTS data set as an example data set for several reasons: (1) it is an open-access data set downloadable from the International Council for the Exploration of the Sea (ICES) website, (2) it is an extensive quality-controlled data set consisting of many species across multiple years, (3) we could use the data-processing protocols that were fully described by Fung et al. (2012) for their calculation of the Large Fish Indicator, and (4) we could use the species-specific length–weight coefficients collated by Fung et al. (2012).

For each year of data, we estimate the size-spectrum exponent, b. Explicitly, this is the exponent of the individual size distribution (ISD) and is related to the slope of the size spectrum (White et al. 2007, Edwards et al. 2017). We find that 5 methods estimate a significant declining trend in b, corresponding to a steepening size spectrum (as can be expected from fishing pressure; e.g. Rice & Gislason 1996, Daan et al. 2005). The remaining 3 methods find no significant trend in b. Thus, the ecological conclusions differ depending on the method used.

However, none of the methods properly take into account the bin structure of the data. They implicitly assign a single length to all counts within a length bin, rather than acknowledging that counts can represent a range of values within the length bin. The binned lengths are subsequently converted to body masses through species-specific length–weight rela-

tionships. We illustrate the problems that occur, which are compounded due to most of the methods requiring further binning to estimate *b*. This motivates investigation of methods to properly account for the bin structure.

The likelihood method is the only method that can be extended to properly deal with binned data. We therefore develop and test such an extended method, which deals with (1) the bin structure of length or body-mass data and (2) situations giving non-integer counts of fish. The latter occurs because the counts (for the IBTS data) are standardised to represent the number of fish caught per hour of trawling, and also arises with other types of data (e.g. Richards et al. 2011). Using simulated body-mass data, we find that applying likelihood without taking into account the bin structure of the data yields biased estimates of *b*, even for high-resolution data. However, the extended likelihood method that takes into account the bin structure is accurate even for low-resolution data. Hence, it may not be worth the effort and expense of collecting high-resolution data if the bin structure is not accounted for.

We derive the likelihood function that fully accounts for species-specific length–weight relationships. This allows the bin structure for length data to be properly accounted for when converting lengths to weights. Applying this function to the IBTS data, we find consistently greater (i.e. less negative) estimates of *b* than for the original likelihood method. Finally, we develop a new method for plotting such data and the resulting ISD that explicitly accounts for the bin structure and associated uncertainty.

Our results strengthen and extend those from Edwards et al. (2017), where we recommended the use of maximum likelihood estimation over the 7 other previous methods based on simulated (rather than real) data. Here, only the likelihood method can be properly extended to deal with the uncertainty associated with binned data.

We finish with recommendations concerning which version of the likelihood method to use for particular types of data. To encourage and enable others to implement our methods, we provide fully documented and functionalised R code in a new R package, 'sizeSpectra'. It includes instructions and vignettes for reproducing all our results, figures and tables (including those in Edwards et al. 2017), and for applying our methods to new data. It is available in Supplement 2 at www.int-res.com/articles/suppl/m636p019_supp/ or can be installed directly from https://github.com/andrew-edwards/sizeSpectra.

2. MATERIALS AND METHODS

2.1. Demersal fish data

For our example data set, we obtained data collected by the North Sea IBTS from the ICES online Database of Trawl Surveys (www.ices.dk/marine-data/data-portals/Pages/DATRAS.aspx). We downloaded the number of organisms caught per hour for each species and length class for all surveys in Areas 1–7 carried out in Quarter 1 from 1986–2015 (see Supplement 1 for full details). All surveys followed standardized gear- and data-collection protocols (ICES 2015). We followed Fung et al. (2012) in removing all non-demersal fish. During surveys, lengths were recorded as rounded down to the nearest 1 cm (or 0.5 cm for Atlantic herring *Clupea harengus* and European sprat *Sprattus sprattus*). We averaged counts of the same year/species/length-class combination across the 7 areas. For simplicity, this ignores potential differences in sampling effort due to variations in haul duration, trawl speed and net area, and fits a common size spectrum for the whole region.

For each species, *s*, the length–weight relationship is

$$w = \alpha_s l^{\beta_s}, \tag{1}$$

where *w* is the estimated body mass (g) of an individual, *l* is the known body length (cm), and α_s and β_s are species-specific parameters published by Fung et al. (2012). Following Fung et al. (2012), for *l* we initially use the 'length class' value (see Table 1 for examples), which represents the minimum of possible lengths for a given fish (i.e. a length class of 35 cm represents fish in the range 35–36 cm). This is consistent with, for example, Blanchard et al. (2005) and Daan et al. (2005), the latter also doing this for 5 and 10 cm length bins. For example, a 45 cm smallspotted catshark is assigned a calculated body mass of $\alpha_s l^{\beta_s}$ = 0.0031 × $45^{3.0290}$ = 315.46 g (first row of Table 1). In 1986 there were 0.00714 individuals h^{-1} of such fish caught, resulting in a total biomass per hour of trawling of 2.25 g h^{-1} (Table 1).

Following calculations of size spectra for earlier IBTS data (Piet & Jennings 2005) and Celtic Sea groundfish surveys (Blanchard et al. 2005), we remove all resulting body masses estimated to be <4 g, because small fish will not be effectively sampled by the gear. The resulting data set (Table 1) contains the required information to estimate *b* of the community for each year. It consists of a data frame with 42 298 rows, where each row is a unique combination of year, species and length class. Each year has between 57 and 81 unique species.

Table 1. First and last 6 rows of the International Bottom Trawl Survey data to illustrate the information available. Each row represents a unique combination of year, species and length class. 'Number' gives the number of individuals per hour of trawling observed for that combination, and can be non-integer because counts of individual fish are scaled by tow duration. Fish lengths are assigned into length classes, where each value is the minimum value of the 1 cm length bin (0.5 cm length bins for Atlantic herring and European sprat). Parameters α_s and β_s: length–weight coefficients for species s from Fung et al. (2012); 'Body mass': the resulting estimated body mass for an individual of that species and length class; 'Biomass': total biomass caught per hour of trawling for each row. Example species are smallspotted catshark *Scyliorhinus canicula*, snakeblenny *Lumpenus lampretaeformis* and thickback sole *Microchirus variegatus*. The full data set has 42 298 rows

Year	Species	Length class (cm)	Number (h^{-1})	α_s	β_s	Body mass (g)	Biomass (g h^{-1})
1986	Smallspotted catshark	45	0.007	0.0031	3.0290	315.46	2.25
1986	Smallspotted catshark	46	0.007	0.0031	3.0290	337.17	2.41
1986	Smallspotted catshark	50	0.007	0.0031	3.0290	434.05	3.10
1986	Smallspotted catshark	52	0.029	0.0031	3.0290	488.81	14.33
1986	Smallspotted catshark	53	0.011	0.0031	3.0290	517.84	5.65
1986	Smallspotted catshark	54	0.011	0.0031	3.0290	548.00	6.18
...	...	...	...	...	...	...	...
2015	Snakeblenny	34	0.028	0.0244	2.0439	32.93	0.92
2015	Thickback sole	8	0.013	0.0080	3.1410	5.49	0.07
2015	Thickback sole	14	0.039	0.0080	3.1410	31.85	1.24
2015	Thickback sole	15	0.052	0.0080	3.1410	39.55	2.05
2015	Thickback sole	16	0.065	0.0080	3.1410	48.44	3.15
2015	Thickback sole	17	0.013	0.0080	3.1410	58.60	0.76

2.2. Individual size distribution

We designate the random variable X to represent the mass (g) of an individual fish or other organism. Then, considering X to come from a bounded power-law (PLB) distribution (also called a truncated Pareto distribution), the probability density function for X is

$$f(x) = Cx^b, \ x_{min} \le x \le x_{max} \tag{2}$$

where

$$C = \begin{cases} \dfrac{b+1}{x_{max}^{b+1} - x_{min}^{b+1}}, & b \ne -1 \\[2em] \dfrac{1}{\log x_{max} - \log x_{min}}, & b = -1 \end{cases} \tag{3}$$

x represents possible values of X, b is the size-spectrum exponent, log is the natural logarithm, and x_{min} and x_{max} are the minimum and maximum possible values of the data with $0 < x_{min} < x_{max}$ (Edwards et al. 2017). The exponent b is expected to be negative, and is the quantity of interest when analysing size spectra. Eq. (2) is the ISD, and the resulting biomass size spectrum (biomass density function) for a community of n individuals is

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

For length size spectra, x would represent lengths, but Eq. (4) cannot be directly used (Edwards et al. 2017).

2.3. Estimating b for each year of the demersal fish data

We apply 8 methods for fitting size spectra that have been previously used in publications and in the R package 'mizer' (Scott et al. 2014). The methods were documented by Edwards et al. (2017). Here, Supplement 1 explains how the methods need to be adapted to deal with the fact that the data are not simply measurements of each individual fish; full results are also given. Using each method, we calculate b separately for each year of data. This gives us a time series of estimated annual values of b with confidence intervals, one time series for each method. We then fit a weighted linear regression (that uses the confidence intervals) to each time series to see whether there is a significant change ($p < 0.05$) in b through time. For clarity, we now present the results which then motivate development of further methods.

3. RESULTS

The linear regression results for each method (Fig. 1) show that (1) the absolute estimates of b are highly dependent upon the method used, (2) the confidence intervals can vary in size through time and are far broader for some methods than for others, and (3) there is no significant change in b through time when using 3 of the methods, yet a significant negative trend (steepening of the size spectrum) when using the remaining 5 methods. This latter point shows that methodological differences can lead to differing ecological conclusions.

3.1. Dealing with the bin structure of the data

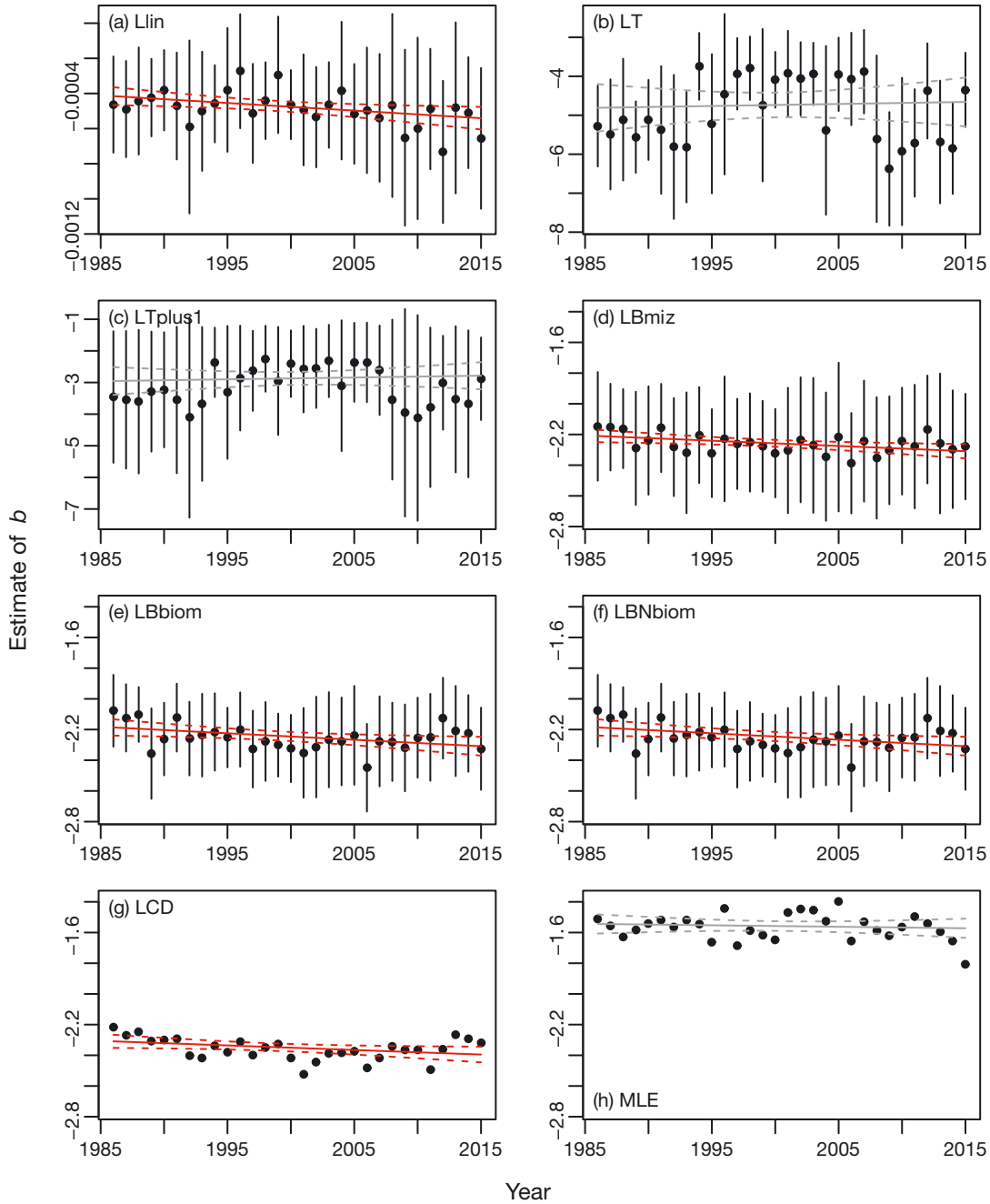However, the aforementioned analyses (Fig. 1) implicitly ignore the fact that using the minimum

Fig. 1. For the International Bottom Trawl Survey data, each of 8 methods is used to estimate the size-spectrum exponent *b* (circles with 95% confidence intervals as vertical bars) for each year. The fit of a weighted linear regression with 95% confidence interval is shown in red if the trend can be considered statistically significant from 0 ($p < 0.05$) and in grey if it cannot be ($p \geq$ 0.05). The *y*-axes are the same for (d–h). For (g) and (h) the confidence intervals for each year are too narrow to be seen. See Supplement 1 for full details of methods and acronym definitions

of a length bin and converting that to a single body mass does not properly account for the bin structure of the data (because the assigned length can actually correspond to any true length within a bin). We demonstrate this issue using length–weight param-

eters for Eq. (1) for 2 example species considered by Blanchard et al. (2005), namely $\alpha_s = 0.0255$ and $\beta_s = 2.7643$ for lemon sole *Microstomus kitt* and $\alpha_s = 0.001$ and $\beta_s = 3.4362$ for common ling *Molva molva*.

Fig. 2 shows how six 5 cm length bins are converted into 6 body-mass bins via Eq. (1). For clarity, we use length bins of width 5 cm and use the midpoint of a length bin to calculate the resulting body mass (note that the IBTS data mostly uses 1 cm bins and was analysed above using the minima of length bins rather than midpoints). The resulting body-mass bins in Fig. 2 are not of equal width because of the nonlinear nature of the length–weight relationships, and the body-mass bins differ between species because of the species-specific length–weight parameters. Such information regarding the body-mass bins was not explicitly accounted for in the above analyses of the IBTS data; all methods used a body mass obtained from applying Eq. (1) to the minimum of a length bin.

Fig. 3 shows the consequence of using the resulting body-mass values, converted from the midpoints of the length bins, when using methods to fit size

spectra that require further binning of the data. For example, the 'logarithmic binning to fit biomass size spectrum' (LBbiom) and 'logarithmic binning with normalisation to fit biomass size spectrum' (LBNbiom) methods were used by Jennings & Dulvy (2005) and Piet & Jennings (2005), respectively, to calculate trends in size-spectra slopes for earlier IBTS data. Explicitly shown is the assignment of the body-mass values into bins with bin breaks of 4, 8, 16, 32, ..., as would be used in the LBbiom and LBNbiom methods.

For example, consider blue length-bin number 5 that is highlighted in Fig. 2 for lemon sole and covers the range 30–35 cm. Say there are 20 individual lemon sole that have lengths in this range. All 20 individuals would be represented by the length equal to the midpoint of the length bin, namely 32.5 cm, which gets converted to a body-mass value of 385 g (even though the 20 body masses really lie in the range 309–473 g). This value of 385 g lies within the $\log_2$ bin that spans 256–512 g (see Bin 5 in Fig. 3). Therefore, all 20 individuals are assigned a body mass of 384 g, which is the midpoint of the $\log_2$ bin. Thus, in the LBbiom and LBNbiom methods, all 20 individuals with body masses in the range 309–473 g are assigned body masses of 384 g. The uncertainty in the original lengths (which is often an unavoidable consequence of measurement methods) is not explicitly accounted for.

Furthermore, the length range 20–25 cm (Bin 3 in Figs. 2 & 3) represents fish in the range 101–187 g. The midpoint of 22.5 cm gets converted to a body-mass value of 139 g, which lies within the $\log_2$ bin that spans 128–256 g (Fig. 3). All individuals are then assigned a body mass of 192 g (the midpoint of the $\log_2$ bin), but this is *greater* than the body mass of any of the individuals (101–187 g). A similar undesirable situation occurs for Bin 6 (Fig. 3). None of the remaining bins yield counts in $\log_2$ bins that accurately represent the original length bins (since there is no reason that they should).

Similar situations also occur for the remaining 4 methods that involve further binning of data. The 'logarithmic plotting of one minus the cumulative distribution' (LCD) and maximum likelihood estimate (MLE) methods do not bin data and so do not suffer from the extra binning described in Fig. 3, but do suffer from assigning the same body mass to all individuals in a length bin. It is not clear how
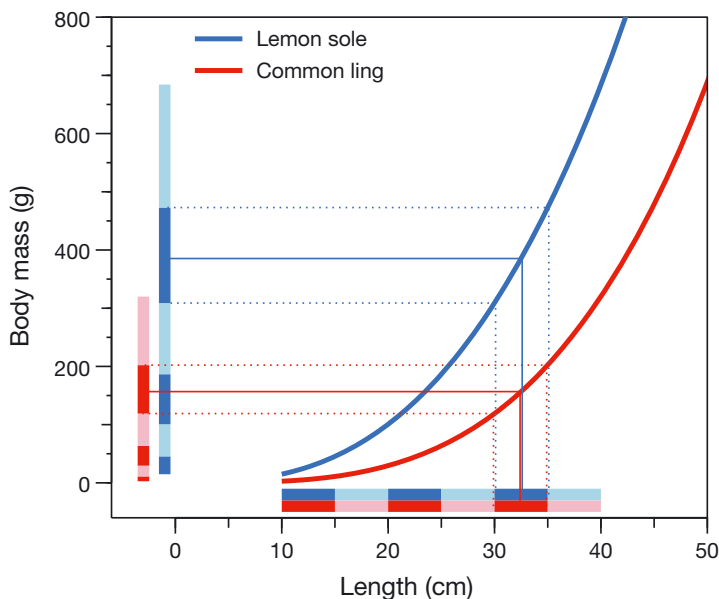

Fig. 2. Example length–weight relationships (solid curves) for lemon sole *Microstomus kitt* and common ling *Molva molva*, showing how binned length measurements yield ranges of estimated body masses. Horizontal bars: 6 example 5 cm length bins, the same bins for both species. Using the length–weight relationships, these length bins convert to the body-mass bins shown by the vertical bars, giving 6 bins of unequal width for each species. The conversion for the 30–35 cm length bin is explicitly shown as dotted lines for the endpoints and solid lines for the midpoint of 32.5 cm. For lemon sole, the 30–35 cm bin with midpoint 32.5 cm converts to a body-mass bin of 309–473 g (values rounded to nearest g for clarity); for common ling the converted body-mass bin is 119–202 g. Note that converting the midpoint of 32.5 cm gives respective body masses of 385 and 157 g, which are slightly less than the respective midpoints of the new body-mass bins (391 and 161 g) because of the nonlinearity of the length–weight relationships
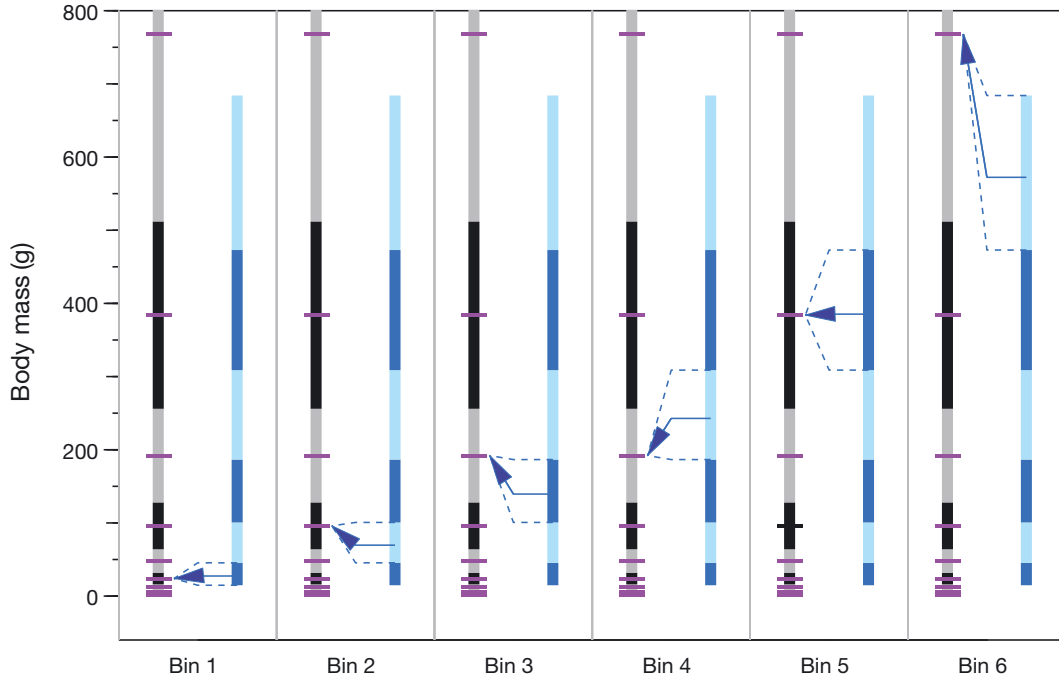
Fig. 3. Demonstration of how binned body-mass values (blue bars) are assigned to logarithmic size-class bins (black and grey bars), as occurs for the 'logarithmic binning to fit biomass size spectrum' (LBbiom) and 'logarithmic binning with normalization to fit biomass size spectrum' (LBNbiom) methods that prescribe the logarithmic bins. Blue vertical bars: the resulting 6 converted body-mass bins for lemon sole from Fig. 2. Black and grey vertical bars: body-mass bins that have equal width on a logarithmic scale. Explicitly these are at 4, 8, 16, 32, … , as used in the LBbiom and LBNbiom methods. Short horizontal purple lines: midpoints of those bins. Each panel (Bin 1 to Bin 6) highlights one blue body-mass bin. For each blue bin, the blue arrow starts from the converted value of the midpoint of the original length bin (see Fig. 2 for Bin 5), and ends at the purple midpoint of the $\log_2$ bin in which the converted value lies. Blue dashed lines indicate that all body masses within the blue bin are assigned to the same purple midpoint value. This ignores uncertainties in body masses, and for Bins 3 and 6 assigns completely incorrect values (greater than the body mass of any individual within the bin). But such values get used in the LBbiom or LBNbiom methods

the LCD method could be extended to account for the bin structure. The MLE method is the only method to accurately estimate b and its confidence intervals for unbinned data (Edwards et al. 2017) — this motivates our adaptation of the MLE method to explicitly account for the uncertainty inherent in binned data.

3.2. Likelihood methods for binned data

3.2.1. MLE based on mid-points

The simplest way to deal with binned data in a likelihood framework is to just use the midpoints of the body-mass bins, which we call the 'MLE based on mid-points' (MLEmid) method. If there is, say, a count of 10 individuals assigned to the bin that spans the body-mass range 4−8 g, then all 10 individuals are assumed to have a body mass of 6 g (the midpoint; though recall that for the IBTS data, we fol-

lowed Fung et al. 2012 in using the minimum value for all 8 methods, including MLE). In Supplement 1, we derive the likelihood function for count data in general and explain how this is used for the MLEmid method.

3.2.2. MLE based on binning

The 'MLE based on binning' (MLEbin) method calculates the MLE for b by explicitly accounting for the fact that each count represent values within its respective size bin. So for 10 individuals in the body-mass bin 4−8 g, the MLEbin method uses a likelihood function that explicitly accounts for the uncertainty in the 10 individual body masses. The body masses can be anywhere in the range 4−8 g, rather than assuming that they are all 6 g as for the MLEmid method. The likelihood function for the MLEbin method was derived as Eqs. (A.70) & (A.75) in Edwards et al. (2017).

3.2.3. Testing the MLEmid and MLEbin methods using simulated data

We now simulate data to test the 2 methods. Following Edwards et al. (2017), we generate 10 000 simulated data sets, each containing 1000 independent random numbers (body masses) drawn from the PLB distribution (Eq. 2) with $b = -2$, $x_{min} = 1$ g and $x_{max} =$ 1000 g. We bin each simulated data set to represent the effect of collecting data in a binned form. We then test how well the MLEmid and MLEbin methods estimate $b = -2$ only using the information from the binned data set (i.e. the bin definitions and the count in each bin), by maximising the appropriate likelihood function and using the profile likelihood-ratio test (Hilborn & Mangel 1997) to calculate confidence intervals.

We test 4 binning types. Type 'Linear 1' uses bins that all have a constant bin width of 1 g and that span the data. The first bin starts at the integer below the lowest data point. Similarly, 'Linear 5' and 'Linear 10' use bins of constant widths 5 and 10 g, respectively. The binning type '2k' refers to bin widths that double in size, span the data, and have bin breaks that are integer powers of 2 (i.e. are at 2^k g where k is an integer, giving bin breaks at 1, 2, 4, 8, 16, ...).

Fig. 4 shows that the MLEbin method clearly performs better than the MLEmid method for all binning types (the histograms are centered around the true value of $b = -2$; summary statistics are given in Supplement 1. The MLEmid method always overestimates b, whereas MLEbin gives roughly the same results regardless of the binning (with mean and median estimates of b of -1.99 or -2.00). Thus, the MLEbin method removes the bias associated with the MLEmid method (note that we cannot formally prove this is always true as there is no analytical solution for the MLE of b, but it also holds for all simulation results in Supplement 1).

Somewhat surprisingly, the MLEmid method is biased even for the 'Linear 1' binning type. So, binning the data using bin breaks of 1, 2, 3, ..., 999, 1000 g and then using the midpoint of a bin to represent all counts in that bin in a likelihood context gives a biased estimate of b (99 % of the simulations over-estimated b). This means that for, say, body-mass data that span the range 1 g to 1 kg and are
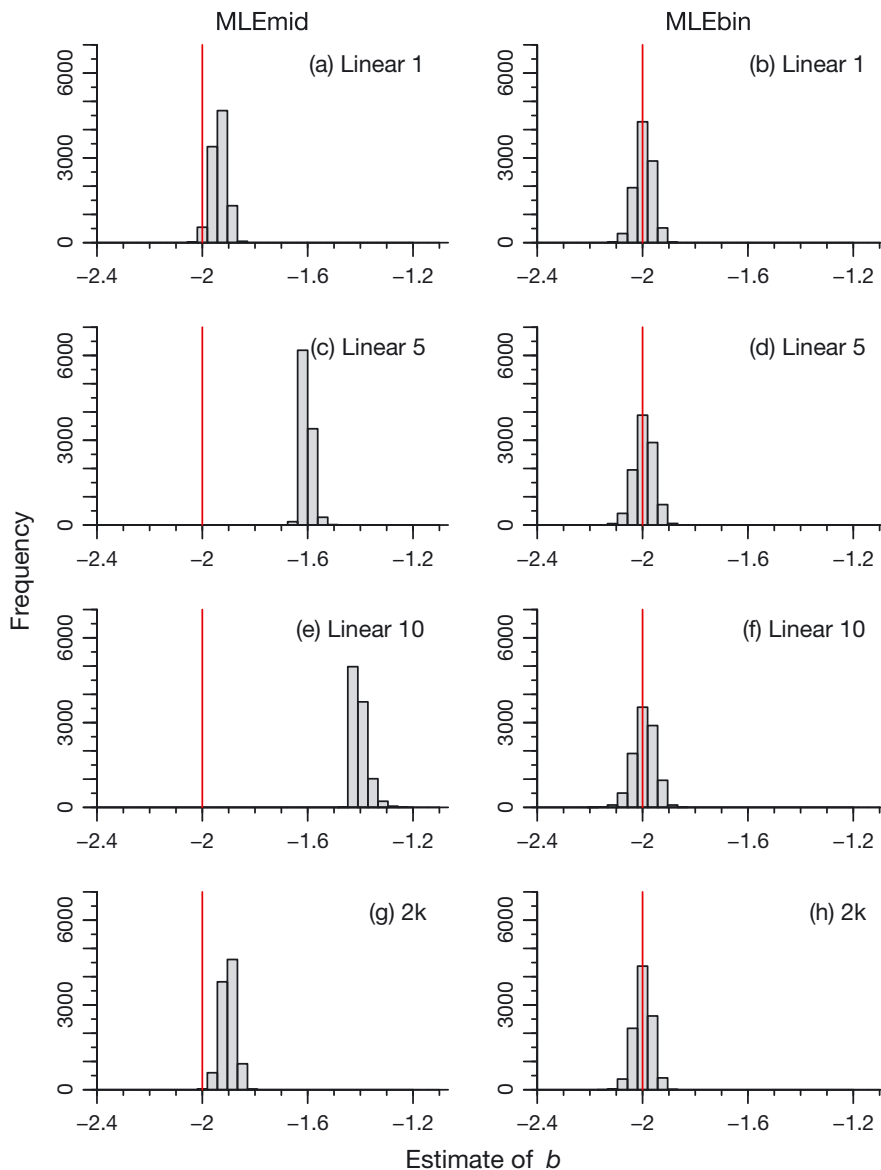

Fig. 4. Histograms of estimated size-spectrum exponent b for 10 000 simulated data sets, each of which contains 1000 independent random numbers drawn from a bounded power-law distribution with $b = -2$, $x_{min} = 1$ and $x_{max} = 1000$; (a–f) using linear bins of constant bin width 1, 5 or 10 (binning type 'Linear 1' etc.); (g–h) using bin widths that double in size (binning type '2k'). Each binned data set is fitted using either the maximum likelihood estimate based on mid-points (MLEmid) (a,c, e,g) or on binning (MLEbin) (b,d,f,h). Vertical red lines: known value of $b = -2$

measured to a resolution of 1 g, ignoring the fact that the binned data represent values in a range (i.e. using the MLEmid method) will lead to biased values of b. The reason for this is the very skewed nature of the power-law distribution. For example, for the simulated data set shown in Fig. S.41 in Supplement 1, the bin covering the range 1–2 g has a count of 528, but 348 (66%) of these values are actually <1.5 g. The MLEmid method assumes that all values equal 1.5 g, whereas the MLEbin method, by definition, accounts for the fact that the values within a bin are assumed to exhibit a power-law distribution.

For the 'Linear 10' binning type, the MLEmid method performs particularly poorly (Fig. 4e). But, perhaps unexpectedly, the corresponding MLEbin method in Fig. 4f leads to accurate estimation of b, and furthermore has better accuracy than the MLEmid method with the finer 'Linear 1' binning type in Fig. 4a. Therefore, the MLEbin method outperforms the MLEmid method, even when it is using data collected at a coarser resolution (10 vs. 1 g) than the MLEmid method. This shows that expending additional time and money to collect high-resolution data is not worthwhile if the bin structure is unaccounted for — it is better to collect low-resolution data and use the MLEbin method to account for the bin structure.

The '2k' binning type performs slightly better than 'Linear 10' when using the MLEbin method (Fig. 4f,h) because it has higher resolution for the data-rich lower bins (bin breaks are 1, 2, 4, 8, ..., compared to 1, 11, 21, 31, ..., for 'Linear 10').

The results regarding confidence intervals (Fig. 5) give a similar conclusion. For the MLEmid method and 'Linear 1' binning type, only 46% of the 95% confidence intervals contain the true value of b, less than half of the desired observed coverage of 95%. The results are worse as the bin width increases ('Linear 5' and 'Linear 10') and for bin widths that double ('2k'). However, the MLEbin gives reliable confidence intervals (observed coverage of 94 or 95%) regardless of binning type.

Thus, these results demonstrate that for binned data the MLEmid method is inaccurate and the MLEbin method should be used, even for seemingly narrow bins.
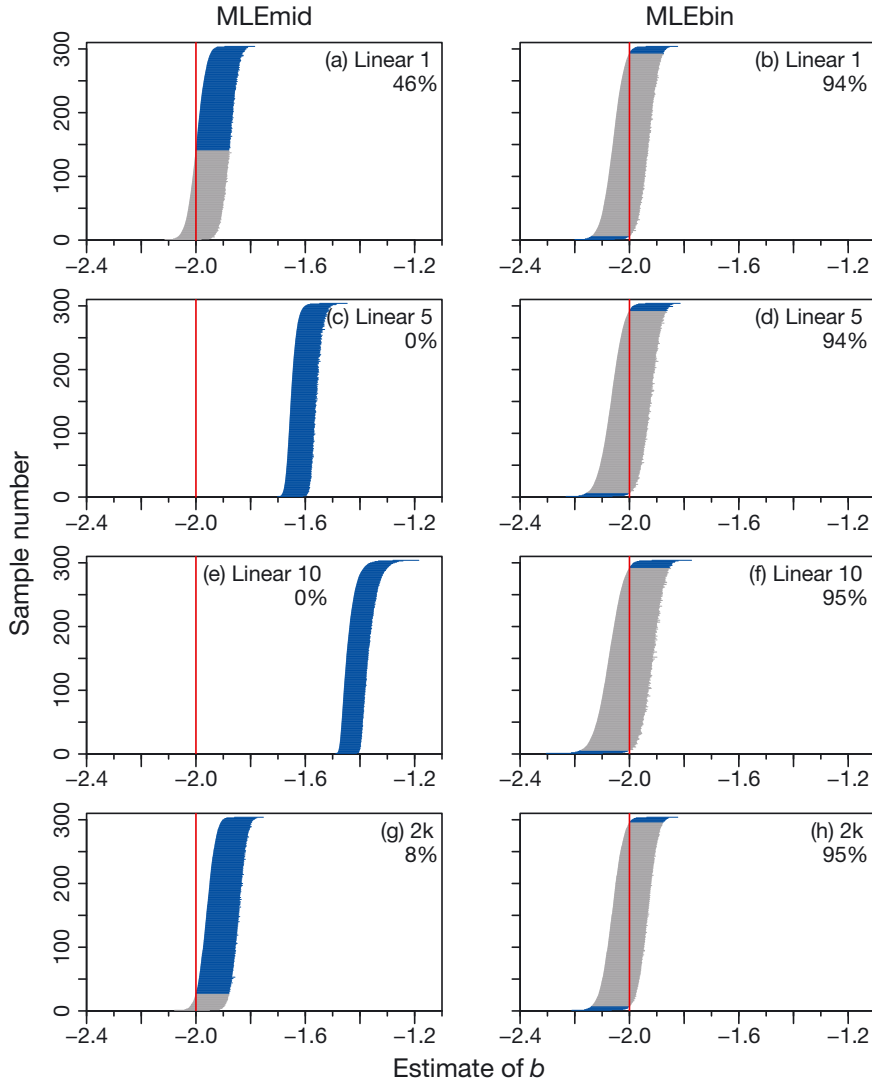

Fig. 5. Confidence intervals (horizontal lines) of the size-spectrum exponent, b, obtained for each combination of data-binning and maximum likelihood estimation method for subsamples of the 10 000 simulated data sets used in Fig. 4. Panels are the same as in Fig. 4. For each numbered subsample on the y-axis, the 95% confidence interval of b obtained using the respective method is plotted as a horizontal line, which is 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 all panels here and in Fig. 4

3.2.4. The MLEbins method: accounting for species-specific body-mass bins

The IBTS data has species-specific length–weight relationships such that the body-mass bins are different for each species, as occurs for many data sets; e.g. Mindel et al. (2018) had lengths of 1 cm resolution that were converted into body masses (and then used for the LBNbiom method). To account for this, we extend the MLEbin method to the MLEbins method, using a multinomial log-likelihood function (Lawless 2003); see Supplement 1 for details.

3.3. MLEbins method applied to IBTS data

To apply the MLEbins method to the IBTS data, we first need to calculate species-specific body-mass bins that result from the species-specific

length–weight coefficients (as for Fig. 2). Fig. 6 shows how the 1 cm (or 0.5 cm) length bins for 45 of the species get converted to body-mass bins. The conversions are different for each species because of the different values of the length–weight coefficients. Thus, the vertical bars are analogous to those in Fig. 2. Fig. 6 shows that even though the length bins are only 1 cm wide, the resulting individual body-mass bins can be quite large; the widest bin across all species is 832 g for Atlantic cod *Gadus morhua* (see Supplement 1).

In Fig. 7, we introduce a plotting technique that visualises the binning structure of the data and the fit of the ISD (from using the MLEbins method). The fitted PLB model shows a reasonable fit to the 1999 data (the red curve passing through the grey rectangles) for body masses <100 g; for larger body masses the PLB model generally overestimates the numbers of individuals. This pattern is seen for many of the
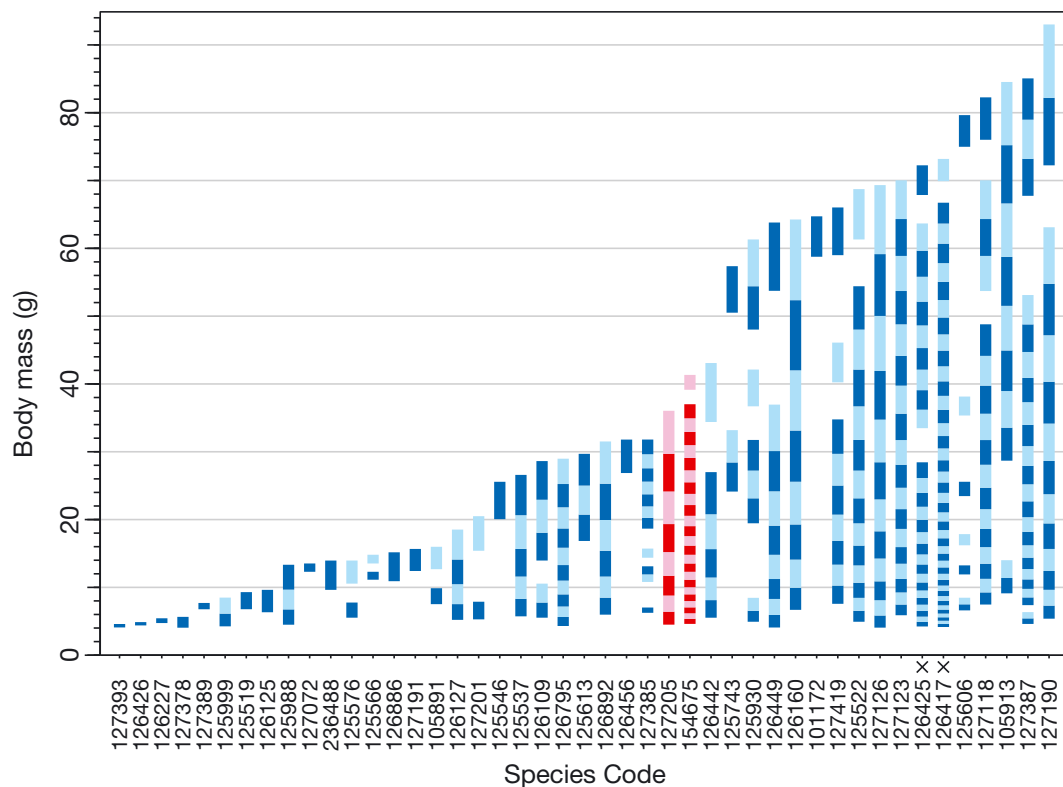

Fig. 6. Vertical bars show the body-mass bins that result from converting the 1 cm length bins into body-mass bins using the species-specific length–weight parameters. These are the first 45 species of the International Bottom Trawl Survey data, ordered by the maximum value of the highest bin. Atlantic herring and European sprat use 0.5 cm length bins (indicated by ×). Any vertical gaps between vertical bars indicate that there were no body masses of that species in that range; i.e. there were no individuals in the length bin(s) that would result in the body-mass bin(s) that would fill the gap. For example, there were no ~30 g European sprat (code 126425). Two example species are shown in red. They have similar ranges of body masses, but moustache sculpin *Triglops murrayi* (code 127205) is short and fat with lengths from 8–15 cm and length–weight relationship $w = 0.0088l^{3.00}$, whereas snakeblenny *Lumpenus lampretaeformis* (code 154675) is long and skinny with lengths from 13–37 cm and $w = 0.0244l^{2.04}$, and consequently the 1 cm length bins translate to much narrower body-mass bins than for moustache sculpin. Equivalent figures for the remaining 90 species are given in Supplement 1
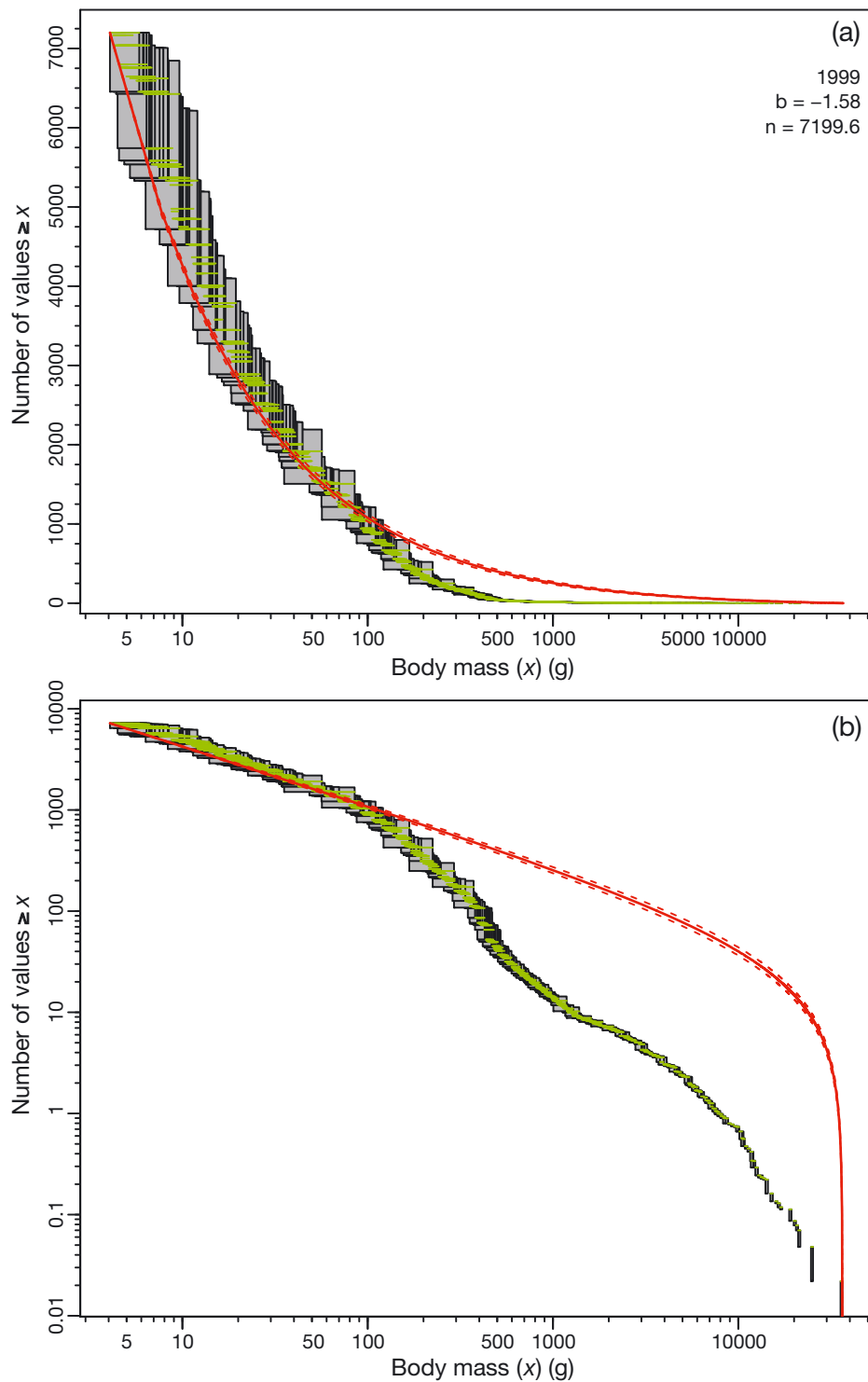
Fig. 7. Suggested plots of individual size distribution and MLEbins fit (red solid curve) with 95% confidence intervals (red dashed curves, may be hard to see) for International Bottom Trawl Survey data in 1999; *y*-axis is (a) linear and (b) logarithmic. For each bin, the horizontal green line shows the range of body sizes, with its value on the *y*-axis corresponding to the total number of individuals (per hour of trawling) in bins whose minima are ≥ the bin's minimum. The vertical span of each grey rectangle shows the possible range of the number of individuals with body mass ≥ the body mass of individuals in that bin (horizontal span is the same as for the green lines). The text in (a) gives the year, the MLE for the size-spectrum exponent *b*, and the sample size *n*. See Supplement 1 for full details and results for all years. Note that in (b) the logarithmic *y*-axis compresses most of the data into the top-left corner

years (see Supplement 1) and may potentially be a consequence of fishing pressure.

Using the species-specific MLEbins method applied to each year of the IBTS data, we obtain consistently higher estimates of b than for the original MLE method (Fig. 8). The differing estimates of b show that the likelihood method did need to be properly adapted (the MLEbins method) to deal with the binned nature of the data. In this case, both methods conclude that there is no significant temporal trend in b through time (see Supplement 1). Nevertheless, the results show the potential for different conclusions about community structure to be reached if the bin structure of data is not properly considered.

3.4. Recommendations

In Table 2 we recommend fitting methods based on the type and resolution of the available data. If data are available in unbinned form (e.g. Trebilco et al. 2015) then the MLE method can be used, but this still depends on the accuracy of the raw data. Because unbinned data are essentially (to some resolution) binned data, given the results in Figs. 4 & 5 we rec-

ommend testing the MLE versus MLEbin method for the particular bin structure and range of a data set. Alternatively, one could just use the MLEbin method, with the bin structure defined by the accuracy of the measurements. Our methods may need to be tailored for explicit data sets depending on the survey design — in Table 2 we give an example concerning monitoring of Pacific reefs.

4. DISCUSSION

We analysed a rich 30 yr data set from the North Sea, and found that ecological conclusions regarding apparent trends in size-spectra exponents can depend on the method used to estimate the exponent. Furthermore, the likelihood method needs to properly account for the binned nature of data (giving the MLEbin method). Using simulated data, even for seemingly high-resolution data, the likelihood method that does not account for the binning (i.e. the MLEmid method) is biased, whereas MLEbin is not. For data such as the IBTS data, where lengths have to be converted to body masses using species-specific length–weight coefficients, we derived the MLEbins method to estimate b and developed a technique for plotting the data and the resulting fit.

Following Edwards et al. (2017) we used a PLB distribution for the ISD because power laws are commonly used models for size spectra (Platt & Denman 1978, Boudreau & Dickie 1992) and alternative models of size-structured predator–prey dynamics predict that the aggregate community ISD may indeed be close to a power law (Andersen & Beyer 2006). Future work could focus on relaxing the assumption of a power-law distribution if other distributions are deemed feasible (Edwards et al. 2017). Using a bounded (rather than unbounded) distribution avoids extrapolation beyond the range of the data — this regrettably prevents estimation of the population density of Loch Ness monsters (as per Sheldon & Kerr 1972). Since power-law distributions have many applications in
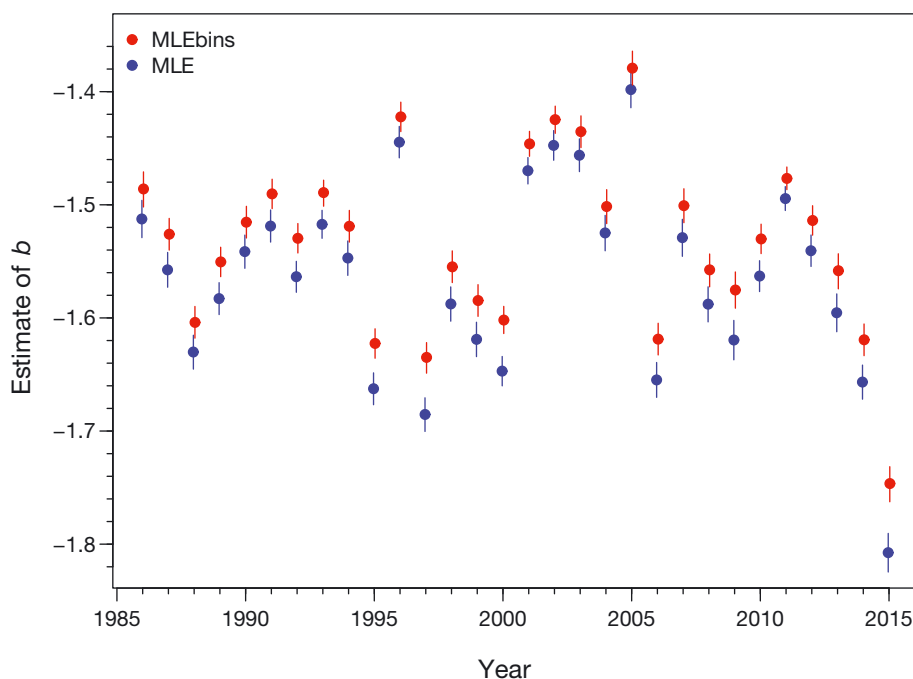

Fig. 8. Comparison of the annual estimates of the size-spectrum exponent, b, for the International Bottom Trawl Survey data using the original maximum likelihood estimate (MLE) method and the 'MLE accounting for species-specific body-mass bins' (MLEbins) method, showing medians and 95 % confidence intervals (each year is slightly offset to clearly show the confidence intervals)

Table 2. Recommendations for calculating the size-spectrum exponent, *b*, of the individual size distribution for body masses (Eq. 2), depending on the type and resolution of the data. MLE: maximum likelihood estimate; MLEbin: MLE based on binning; MLEbins: MLE accounting for species-specific body-mass bins

Data type	Resolution	Issue	Solution
Body mass	Individual body masses	Accuracy of measurements may require them to be considered as bins	Simulate data as in Fig. 4 to see if MLE method sufficient or MLEbin is needed
Body mass	Individuals assigned to common body-mass bins	Data are counts in bins	Use MLEbin method
Length	Individual lengths	Lengths have to be converted to body masses	Calculate individual body masses and proceed as per 'individual body masses'
Length	Individuals assigned to length bins	Resulting converted body-mass bins are species specific	Use MLEbins method
Lengths from under-water visual census (e.g. point counts or belt transects)	Individual lengths	May need to standardise counts when combining different survey designs (e.g. Heenan et al. 2016); surveys may cover different sized areas or count small and large organisms differently	Potential need to adapt MLEbins likelihood function to account for particular survey design

ecology (White et al. 2008), our simulation results suggest that bin structure may need to be accounted for in other contexts, and that our novel plotting method (Fig. 7) would be similarly applicable.

Ideally, our methods could be extended to incorporate estimated uncertainties of the species-specific length–weight coefficients (see Froese 2006). This would provide further uncertainty regarding the body mass associated with each length bin (see Benoist et al. 2019 for an explicit example from raw data). However, practical applications could be limited since published tabulations of length–weight relationships do not generally include confidence intervals for the parameters (e.g. Love et al. 2002, Robinson et al. 2010, Fung et al. 2012).

Following Piet & Jennings (2005) and Blanchard et al. (2005), we used a (minimum) cutoff of 4 g to analyse the IBTS data. This was necessarily done by converting the minima of the length bins into species-specific body-mass bins, and then only using the resulting bins with minima ≥4 g. For a resulting body-mass bin of 3–5 g, for example, this approach removes some individuals ≥4 g, even though we wish to impose a minimum of 4 g. This may be unavoidable, although note that the likelihood function for the MLEbins method properly assumes no data (rather than no individuals) <5 g for such a species with lowest bin 3–5 g ($w_{s,1} = 5$ g for this species in Eq. S.18 in Supplement 1). Also note that we have estimated x_{max} separately for each year although it could be set to a fixed maximum value across all years.

There are techniques available for finding the value of x_{min} that gives the best fit of an unbounded power law (Clauset et al. 2009, Gillespie 2015). But these are focussed on finding power-law *tails* (given the optimum x_{min}), rather than finding a distribution that fits the full size range of the sampled community, which is what is desired when fitting size spectra. Such techniques would have to be adapted to use our binned likelihood methods (and to use the bounded rather than unbounded power-law distribution). Rather, we recommend that the range of data be justified from biological and sampling perspectives, rather than adjusting the range to most closely resemble a power-law distribution. This is particularly so when comparing fits across multiple years or locations because x_{min} should be consistent; e.g. Robinson et al. (2017) fixed $x_{min} = 20$ g for all data sets from 38 Pacific islands. We re-ran our analyses of the IBTS data with a cutoff value of 100 g (much higher than the original 4 g). The exponent *b* was much lower for all years, and increased through time rather than not changing (see Supplement 1). This emphasises the need for careful biological consideration of the cutoff value in such applications.

We have used the IBTS data to motivate development of widely applicable methods. We caution against using our current results for management purposes concerning the North Sea, because we have not fully explored assumptions (such as those just discussed) and have used all the sampled species rather than only a well-sampled subset (as recom-

mended by Jennings & Dulvy 2005). Also, the power analyses conducted by Jennings & Dulvy (2005) on earlier IBTS data could be repeated, in particular to determine the lower cutoff value that would provide the greatest power to detect trends in *b*. Note that Jennings & Dulvy (2005) were restricted to such cutoffs being powers of 2 because of the use of the LBbiom method, but this would not be necessary with the MLEbins method.

Note that the confidence intervals calculated for the IBTS data from the MLEbins method are not strictly true confidence intervals, because the commonly used unit of 'numbers of individuals caught per hour of trawling' is arbitrary. If numbers were instead expressed as 'per day of trawling', then all counts would be divided by 24, and the resulting confidence intervals would be larger than those in Fig. 8. True confidence intervals could be obtained by using the absolute numbers of individual fish caught summed across all areas (rather than numbers per hour of trawling averaged across areas); this should be investigated if conducting further analyses of such types of data. This would change the width of the confidence intervals, but not the MLE for *b*, which is unaffected by any change in units.

All results, figures and tables presented here are reproducible in our R package 'sizeSpectra'. The code is properly documented and functionalised. The included vignettes will enable users to fit size spectra to their own data using our recommended and tested likelihood methods that properly account for the bin structure of the data.

LITERATURE CITED

Álvarez E, Morán XAG, López-Urrutia Á, Nogueira E (2016) Size-dependent photoacclimation of the phytoplankton community in temperate shelf waters (southern Bay of Biscay). Mar Ecol Prog Ser 543:73–87

Andersen KH, Beyer JE (2006) Asymptotic size determines species abundance in the marine size spectrum. Am Nat 168:54–61

Benoist NMA, Bett BJ, Morris KJ, Ruhl HA (2019) A generalised volumetric method to estimate the biomass of photographically surveyed benthic megafauna. Prog Oceanogr 178:102188

Blanchard JL, Dulvy NK, Jennings S, Ellis JR, Pinnegar JK, Tidd A, Kell LT (2005) Do climate and fishing influence size-based indicators of Celtic Sea fish community structure? ICES J Mar Sci 62:405–411

Boldt JL, Bartkiw SC, Livingston PA, Hoff GR, Walters GE (2012) Investigation of fishing and climate effects on the community size spectra of eastern Bering Sea fish. Trans Am Fish Soc 141:327–342

Boudreau PR, Dickie LM (1992) Biomass spectra of aquatic ecosystems in relation to fisheries yield. Can J Fish Aquat Sci 49:1528–1538

Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. SIAM Rev 51:661–703

Constable AJ, Costa DP, Schofield O, Newman L and others (2016) Developing priority variables ('ecosystem Essential Ocean Variables' – eEOVs) for observing dynamics and change in Southern Ocean ecosystems. J Mar Syst 161:26–41

Daan N, Gislason H, Pope JG, Rice JC (2005) Changes in the North Sea fish community: evidence of indirect effects of fishing? ICES J Mar Sci 62:177–188

Dulvy NK, Polunin NVC, Mill AC, Graham NAJ (2004) Size structural change in lightly exploited coral reef fish communities: evidence for weak indirect effects. Can J Fish Aquat Sci 61:466–475

Edwards AM (2011) Overturning conclusions of Lévy flight movement patterns by fishing boats and foraging animals. Ecology 92:1247–1257

Edwards AM, Robinson JPW, Plank MJ, Baum JK, Blanchard JL (2017) Testing and recommending methods for fitting size spectra to data. Methods Ecol Evol 8: 57–67

European Commission (2010) Commission decision of 1 September 2010 on criteria and methodological standards on good environmental status of marine waters. Off J Eur Union L232:14–24

Froese R (2006) Cube law, condition factor and weight-length relationships: history, meta-analysis and recommendations. J Appl Ichthyology 22:241–253

Fung T, Farnsworth KD, Reid DG, Rossberg AG (2012) Recent data suggest no further recovery in North Sea Large Fish Indicator. ICES J Mar Sci 69:235–239

Gillespie CS (2015) Fitting heavy tailed distributions: the poweRlaw package. J Stat Softw 64:1–16

Heenan A, Gorospe K, Williams I, Levine A and others (2016) Ecosystem monitoring for ecosystem-based management: using a polycentric approach to balance information trade-offs. J Appl Ecol 53:699–704

Hilborn R, Mangel M (1997) The ecological detective: confronting models with data. Monographs in Population Biology, Vol 28. Princeton University Press, Princeton, NJ

ICES (2015) Manual for the International Bottom Trawl Surveys. Series of ICES Survey Protocols SISP 10 – IBTS IX. International Bottom Trawl Survey Working Group, ICES, Copenhagen

Jennings S (2005) Indicators to support an ecosystem approach to fisheries. Fish Fish 6:212–232

Jennings S, Dulvy NK (2005) Reference points and reference directions for size-based indicators of community structure. ICES J Mar Sci 62:397–404

Kelly-Gerreyn BA, Martin AP, Bett BJ, Anderson TR and others (2014) Benthic biomass size spectra in shelf and deep-sea sediments. Biogeosciences 11:6401–6416

Lawless JF (2003) Statistical models and methods for lifetime data, 2nd edn. Wiley Series in Probability and Statistics, Hoboken, NJ

Love MS, Yoklavich M, Thorsteinson L (2002) The rockfishes of the Northeast Pacific. University of California Press, Berkeley, CA

Maxwell TAD, Jennings S (2006) Predicting abundance-body size relationships in functional and taxonomic subsets of food webs. Oecologia 150:282−290

McCoy K, Heenan A, Asher J, Ayotte P and others (2016) Pacific Reef Assessment and Monitoring Program: data report. Ecological monitoring 2015 — reef fishes and benthic habitats of the main Hawaiian Islands, Northwest Hawaiian Islands, Pacific Remote Island Areas, and American Samoa. PIFSC Data Report DR-16-002. NOAA Pacific Islands Fisheries Science Center, Honolulu, HI

Mindel BL, Neat FC, Webb TJ, Blanchard JL (2018) Size-based indicators show depth-dependent change over time in the deep sea. ICES J Mar Sci 75:113−121

Monnahan CC, Ono K, Anderson SC, Rudd MB and others (2016) The effect of length bin width on growth estimation in integrated age-structured stock assessments. Fish Res 180:103−112

Piet GJ, Jennings S (2005) Response of potential fish community indicators to fishing. ICES J Mar Sci 62:214−225

Platt T, Denman K (1978) The structure of pelagic marine ecosystems. Rapp P-v Reun Cons Int Explor Mer 173: 60−65

Rice J, 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 J Mar Sci 53:1214−1225

Rice J, Arvanitidis C, Borja A, Frid C and others (2012) Indicators for sea-floor integrity under the European Marine Strategy Framework Directive. Ecol Indic 12:174−184

Richards BL, Williams ID, Nadon MO, Zgliczynski BJ (2011) A towed-diver survey method for mesoscale fishery-independent assessment of large-bodied reef fishes. Bull Mar Sci 87:55−74

Robinson LA, Greenstreet SPR, Reiss H, Callaway R and others (2010) Length-weight relationships of 216 North Sea benthic invertebrates and fish. J Mar Biol Assoc UK 90:95−104

Robinson JPW, Williams ID, Edwards AM, McPherson J and others (2017) Fishing degrades size structure of coral reef fish communities. Glob Change Biol 23:1009−1022

Scott F, Blanchard JL, Andersen KH (2014) *mizer*: an R package for multispecies, trait-based and community size spectrum ecological modelling. Methods Ecol Evol 5: 1121−1125

Sheldon RW, Kerr SR (1972) The population density of monsters in Loch Ness. Limnol Oceanogr 17:796−798

Shin YJ, Rochet MJ, Jennings S, Field JG, Gislason H (2005) Using size-based indicators to evaluate the ecosystem effects of fishing. ICES J Mar Sci 62:384−396

Stasko AD, Swanson H, Majewski A, Atchison S, Reist J, Power M (2016) Influences of depth and pelagic subsidies on the size-based trophic structure of Beaufort Sea fish communities. Mar Ecol Prog Ser 549:153−166

Taniguchi DAA, Franks PJS, Poulin FJ (2014) Planktonic biomass size spectra: an emergent property of size-dependent physiological rates, food web dynamics, and nutrient regimes. Mar Ecol Prog Ser 514:13−33

Thorpe RB, Le Quesne WJF, Luxford F, Collie JS, Jennings S (2015) Evaluation and management implications of uncertainty in a multispecies size-structured model of population and community responses to fishing. Methods Ecol Evol 6:49−58

Trebilco R, Dulvy NK, Stewart H, Salomon AK (2015) The role of habitat complexity in shaping the size structure of a temperate reef fish community. Mar Ecol Prog Ser 532: 197−211

van Gemert R, Andersen KH (2018) Implications of late-in-life density-dependent growth for fishery size-at-entry leading to maximum sustainable yield. ICES J Mar Sci 75:1296−1305

White EP, Ernest SKM, Kerkhoff AJ, Enquist BJ (2007) Relationships between body size and abundance in ecology. Trends Ecol Evol 22:323−330

White EP, Enquist BJ, Green JL (2008) On estimating the exponent of power-law frequency distributions. Ecology 89:905−912

Woodson CB, Schramski JR, Joye SB (2018) A unifying theory for top-heavy ecosystem structure in the ocean. Nat Commun 9:23

Yool A, Martin AP, Anderson TR, Bett BJ, Jones DOB, Ruhl HA (2017) Big in the benthos: future change of seafloor community biomass in a global, body size-resolved model. Glob Change Biol 23:3554−3566

Zhang C, Chen Y, Xu B, Xue Y, Ren Y (2018) Evaluating fishing effects on the stability of fish communities using a size-spectrum model. Fish Res 197:123−130

Zhou S, Kolding J, Garcia SM, Plank MJ and others (2019) Balanced harvest: concept, policies, evidence, and management implications. Rev Fish Biol Fish 29:711−733