Supplementary Material – Extended Methods and Results

Supplementary Material for 'Accounting for the bin structure of data removes bias when fitting size spectra' by Andrew M. Edwards, James P. W. Robinson, Julia L. Blanchard, Julia K. Baum and Michael J. Plank. *Marine Ecology Progress Series*.

Table of Contents

S.1 Extended methods and results

S.1.1 Outline

In Section S.1.2 we give further details regarding the International Bottom Trawl Survey (IBTS) data. In Section S.1.3 we analyse the IBTS data by assuming the 'Length class' value to be the exact length of the fish (as is often done, e.g. Blanchard *et al.* 2005; Daan *et al.* 2005), ignoring the fact that a length class of, say, 35 cm actually represents fish in the range 35-36 cm. We explain how to extend each of the eight methods for estimating the size-spectrum exponent to deal with the non-integer counts of the numbers of fish. For the likelihood method we derive the likelihood function for counts data (Section S.1.4) and show that it is valid for non-integer counts (Section S.1.5).

We then derive the likelihood function for the MLEbins method, which properly accounts for the species-specific body-mass bins (Section S.1.6). The other methods cannot be extended in a similar way.

In Section S.1.7 we show the body-mass bins for the IBTS data for the species not shown in Figure 6. In Section S.1.8 we develop a method for plotting the data and resulting ISD in a way that accounts for the bin structure and the uncertainty in the estimate of the exponent b, and show figures for each year of the IBTS data.

Section S.1.9 gives summary statistics for the main simulation results, plus sensitivity results from setting a larger minimum body size and not sampling the smallest organisms. In Section S.1.10 we repeat the analysis of the IBTS data but with a much higher minimum body size. In Section S.1.11 we show histograms of a simulated data set from a bounded power-law distribution to illustrate the highly skewed nature of the distribution.

We give details of our R package `sizeSpectra` in Section S.2, which can be used to reproduce all results in this paper and apply the methods to other data. It can also be downloaded directly from `https://github.com/andrew-edwards/sizeSpectra`.

S.1.2 Details of the IBTS data

We obtained data collected by the North Sea IBTS from the ICES online Database of Trawl Surveys (`http://www.ices.dk/marine-data/data-portals/Pages/DATRAS.aspx`). Seven countries survey the whole of the North Sea, bounded by the eastern English Channel, northern continental margin above the Shetland archipelago, the east coast of the UK and the Skagerrak and Kattegat regions (ICES, 2015). Surveys take place in the first and third quarters of each year following standardized gear and data collection protocols (ICES, 2015). We downloaded the number of organisms caught per hour (surveyed catch per unit effort) for each species and length class for all surveys in Areas 1 to 7 carried out in Quarter 1 from 1986-2015. Following the protocol of Fung *et al.* (2012), we restricted the time coverage of the surveys to 1986 onwards because these are the years in which standardized fishing gear (the GOV trawl – chalut à grande ouverture verticale) was deployed on all vessels. We downloaded data as "cpue per length per haul" in the formats "Exchange Data" and "SMALK", and applied the classifications published in Fung *et al.* (2012) to remove all non-demersal fish. During data collection on 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 seven 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.

S.1.3 Applying eight methods to the IBTS data

We use eight methods that have been previously used to calculate size-spectra slopes or exponents. See Edwards *et al.* (2017) for full details and example references for each method. Briefly, the methods are:

- Llin (Log-linear transform) – plot linearly binned data on log-linear axes then fit regression of log(count in bin) against midpoint of bin;

- LT (log transform) – plot linearly binned data on log-log axes then fit regression of log(count in bin) against log(midpoint of bin);
- 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 $\log_{10}(\text{midpoint of bin})$;
- LBmiz (logarithmic binning as done in the R package `mizer` by Scott *et al.* 2014) – bin data using $\log_{10}$ bins, but with largest bin the same arithmetic size as the penultimate bin, then fit regression of log(count in bin) against log(lower bound of bin);
- LBbiom (logarithmic binning) and then fit biomass size spectrum – bin body masses using $\log_2$ bins then fit regression of $\log_{10}(\text{biomass in bin})$ against $\log_{10}(\text{midpoint of bin})$;
- LBNbiom (logarithmic binning with normalisation and then fit biomass size spectrum) – bin body masses using $\log_2$ bins, then fit regression of $\log_{10}$ (biomass in bin divided by bin width) against $\log_{10}(\text{midpoint of bin})$;
- LCD (logarithmic plotting of $1 - F(x)$; i.e. one minus the cumulative distribution) – rank data from largest (rank 1) to smallest (rank n) and fit regression of $\log(\text{rank}(x)/n)$ against $\log x$;
- MLE (maximum likelihood estimate) – calculate maximum likelihood estimate of b.

Using each method we calculate the size-spectra exponent b separately for each year of data. As in Table 1 (following Fung *et al.* 2012) we use the 'Length class' values (the minimum of each length bin) to calculate body masses associated with the numbers of individuals. Although some methods use counts of individuals and some use biomass, they are related through (2) and (4) and hence can estimate b (Edwards *et al.*, 2017). 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 – see below) to each time series to see whether there is a significant change ($p < 0.05$) in b through time.

However, the eight methods first have to be extended to deal with the non-integer counts of numbers of fish of each body mass. For the binning-based methods (Llin, LT, LTplus1, LBmiz, LBbiom and LBNbiom) the extension is fairly obvious. The bins used to fit the size spectra are defined by the method. Each count of a particular body mass (each row in Table 1) is assigned to a bin that is defined by the method being used, and (for each year) the total count in each body-mass bin is calculated by summing the 'Number' values for that bin. The resulting total count in each body-mass bin no longer has to be an integer (because the 'Number' values are not integers), but the resulting regression fits can be calculated as described by Edwards *et al.* (2017).

The LCD method requires ranking the body mass, x, of each individual fish in descending order from 1 (largest) to n (smallest) and then fitting a regression of log(fraction of values $\geq x$) against log x. However, for the IBTS data the body masses do not correspond to individual fish, rather we have the number of fish of a certain body mass – but that number can be non-integer. We use the simplest approach to adapt the LCD method. First, for a particular year arrange the data ('Number' for each body mass from Table 1) in descending order of body mass, and for each row calculate the cumulative sum of the 'Number' – this is akin to assigning ranks for individual body masses. Dividing the cumulative sum by the total 'Number', gives the fraction of individuals with a body mass $\geq x$ for each given body mass, x. Then plot log(proportion of values $\geq x$) against log x and fit a regression, as per Figure 2(g) of Edwards *et al.* (2017). Each plotted point no longer corresponds to an individual fish, but relates to a row in Table 1.

The original MLE method of Edwards *et al.* (2017) requires the body mass of individual fish, and so in Sections S.1.4 and S.1.5 we formally extend it to deal with count data.

So we calculate the estimated b for each of the 30 years of data, using each of the eight methods. For each method, we plot the resulting time series of estimates of b and analyse any trend by fitting a weighted linear regression to the estimates. A weighted regression is used because we can estimate the variances as the square of the standard errors from regression outputs, and the variances appear to vary over time. For the MLE method, having already calculated the

Table S.1: Results of weighted regression analyses of trend through time of the estimated exponent b for the IBTS data, as estimated using each of the eight methods shown in Figure 1, plus the MLEbins method developed here. 'Trend' is the estimated annual trend, with 95% confidence intervals given by 'Low' and 'High', p is the p-value for the probability that the trend is significantly different to 0, and R^2 is the coefficient of determination. If $p \geq 0.05$ then the trend can be considered not significantly different to 0. If $p < 0.05$ then a negative trend indicates a statistically significant decline in the exponent over time, and a positive trend indicates a statistically significant increase.

Method	Low	Trend	High	p	R^2
Llin	-0.0000	-0.0000	-0.0000	0.01	0.19
LT	-0.0313	0.0052	0.0417	0.77	0.00
LTplus1	-0.0206	0.0058	0.0321	0.66	0.01
LBmiz	-0.0060	-0.0034	-0.0009	0.01	0.21
LBbiom	-0.0076	-0.0042	-0.0008	0.02	0.19
LBNbiom	-0.0076	-0.0042	-0.0008	0.02	0.19
LCD	-0.0057	-0.0030	-0.0002	0.04	0.15
MLE	-0.0047	-0.0010	0.0027	0.60	0.01
MLEbins	-0.0043	-0.0010	0.0024	0.56	0.01

95% confidence intervals $(b_{\text{low}}, b_{\text{high}})$ using the profile likelihood method Hilborn and Mangel (1997), the standard error, γ, can be calculated from the confidence interval being approximately $b_{\text{MLE}} \pm 1.96\gamma$ where b_{MLE} is the MLE for b (Burnham and Anderson, 2002). Explicitly, we used the mean

$$\gamma = \frac{|b_{\text{low}} - b_{\text{MLE}}| + |b_{\text{high}} - b_{\text{MLE}}|}{2 \times 1.96} = \frac{b_{\text{high}} - b_{\text{low}}}{2 \times 1.96}. \tag{S.1}$$

Figure 1 and Table S.1 shows that the absolute estimates of b are highly dependent upon the method used. The confidence intervals are somewhat narrower ($\leq$0.002) in absolute size for the Llin and MLE methods than those for the other methods ($\geq$0.005), and vary in size through time. There is no significant change in b when using three of the estimation methods (LT, LTplus1 and MLE), yet a significant decline when using the remaining five methods. Thus, five methods imply a steepening of the size spectrum over time, whereas three methods imply no change. This demonstrates how methodological differences can lead to differing ecological conclusions.

S.1.4 Likelihood function for count data

Here we derive the likelihood function for count data, that is then shown in Section S.1.5 to apply to non-integer counts. The latter was used for the MLE method (that ignores the bin structure) for the IBTS data.

From Edwards *et al.* (2017), the log-likelihood function for the PLB model for $b \neq -1$ is

$$\begin{aligned} \log[\mathrm{L}(b|\text{data } \mathbf{x})] &= \sum_{j=1}^{n} \log f(x_j) && \text{(S.2)} \\ &= n \log\left(\frac{b+1}{x_{\max}^{b+1} - x_{\min}^{b+1}}\right) + b \sum_{j=1}^{n} \log x_j, && \text{(S.3)} \end{aligned}$$

where $\mathrm{L}(b|\text{data } \mathbf{x})$ is the likelihood of a particular value of the unknown parameter b given the known data $\mathbf{x} = \{x_1, x_2, x_3, ..., x_n\}$, and $f(\cdot)$ is the probability density function (2). The maximum likelihood estimates for $x_{\min}$ and $x_{\max}$ are simply the minimum and maximum values of the data.

Now, if some of the x_j are repeated values, we can represent the data in a more concise form. For example, the data set $\mathbf{x} = \{4, 6, 6, 6, 9, 10, 10\}$ can be represented as counts $\{1, 3, 1, 2\}$ for each of the unique x values $\{4, 6, 9, 10\}$. More generally, let c_k be the count (number of repetitions) of x'_k, where $k = 1, 2, 3, ..., K \leq n$, and $\sum_{k=1}^{K} c_k = n$. The vector $\mathbf{x}' = \{x'_k\}$ represents the unique values of the original x_j, and $K = n$ only when all $c_k = 1$ (i.e. the original x_j values are all unique).

To obtain the log-likelihood function, note that

$$\sum_{j=1}^{n} \log f(x_j) = \sum_{k=1}^{K} c_k \log f(x'_k), \qquad \text{(S.4)}$$

because if x_j is repeated c_k times then the corresponding contribution of the x_j values to the overall $\sum \log f(x_j)$ is just $c_k \log f(x'_k)$ with $x'_k = x_j$. Hence, the equivalent log-likelihood function to (S.3) is

$$\begin{aligned} \log[\mathrm{L}(b|\text{data } \mathbf{x}')] &= \sum_{k=1}^{K} c_k \log f(x'_k) && \text{(S.5)} \\ &= n \log\left(\frac{b+1}{x_{\max}^{b+1} - x_{\min}^{b+1}}\right) + b \sum_{k=1}^{K} c_k \log x'_k. && \text{(S.6)} \end{aligned}$$

Similarly, for $b = -1$ the log-likelihood function is

$$\log[\mathrm{L}(b = -1|\text{data } \mathbf{x}')] = -n\log(\log x_{\max} - \log x_{\min}) - \sum_{k=1}^{K} c_k \log x'_k. \tag{S.7}$$

Note that if a value x_j is repeated because it represents a value that is rounded due to the resolution of the measurement process, (e.g. $x_j = 10$ g really represents a value in the range $9.5 - 10.5$ g and then occurs multiple times in the data set) then the MLEbin method should be used (Edwards *et al.*, 2017). If the values $\{x_j\}$ each represent true discrete measurements, then the *continuous* power-law distribution (2) is not appropriate and a *discrete* distribution should be used. But for length and weight measurements used for size spectra, the underlying variables are indeed continuous, and any discrete values are due to measurement resolution.

S.1.5 Likelihood function for data with non-integer counts

For the IBTS data, the counts c_k corresponding to each body mass x'_k are not all integers – they can take non-integer values. The duration of individual research trawls was not constant. So the integer counts of the number of a particular species of fish within a particular length bin from a particular trawl needed to be divided by that trawl's duration, to give the number of fish per hour within each length bin. This standardised the unit of measurement across trawls of different duration. The resulting counts per hour, c_k, are therefore not restricted to being integers. Similarly, Daan *et al.* (2005) analysed bottom-trawl survey data that were 'catch in number by size or Lmax [maximum length] class per hour fishing' that resulted in non-integer counts, as did Blanchard *et al.* (2005) in their analysis of groundfish surveys from the Celtic Sea.

We ascertain that the above likelihood functions (S.6) and (S.7) can still be applied to data containing non-integer counts. In the above derivations, each count c_k of the number of fish of size x'_k was assumed to be integer-valued. But we now show that this condition can be relaxed, and therefore allow each c_k to be non-integer. If c_k are non-integer then they can be scaled by a constant factor z to give counts $c'_k = zc_k$ such that all the c'_k are integers. Using c'_k in place of c_k and $n' = zn$ in place of n simply multiplies the log-likelihood function in (S.6) by a constant

z, which does not change the location of its maximum. Hence, using non-integer densities c_k in (S.6) gives the same MLE for b as scaling the densities up to integer values.

However, any such rescaling, by multiplying (S.6) by z, changes the curvature of the log-likelihood function, and so changes the width of the resulting confidence interval. This is because the interval is calculated using the profile likelihood-ratio test as the range of parameters for which the log-likelihood is within 1.92 of the maximum value of the log-likelihood (page 163 of Hilborn and Mangel 1997). The 1.92 is based on a chi-squared distribution with one degree of freedom (and is more precisely calculated as `qchisq(0.95,1)/2` in our R code), and is an absolute value that does not change when the log-likelihood function is scaled by z.

Thus, calculated confidence intervals will depend on the units of the counts – changing the counts of fish in the IBTS data from numbers per hour of trawling to numbers per day of trawling will change the confidence intervals of b (but not the MLEs). An analogy is tossing 10 identical (but unfair) coins and getting 7 heads and 3 tails (the MLE for the probability of getting a head is 0.7, but with a wide confidence interval), compared to tossing 1,000 coins and getting 700 heads and 300 tails (the MLE is still 0.7, but the uncertainty will be much narrower given the larger sample size).

Note that the units of measurement (g vs. kg) will not affect the MLEs or the confidence intervals as this change only adds a constant value to the log-likelihood function.

S.1.6 Likelihood function for the MLEbins method

Here we derive the likelihood function for the MLEbins method. This extends the MLEbin method for when the body-mass bins are not defined the same for all species. This occurs for the IBTS data because the counts in the 1-cm (or 0.5-cm) length bins yield counts in body-mass bins via length-weight relationships. Since the length-weight relationships are species-specific, the resulting body-mass bins are species-specific, as shown in Figures 6 and S.1-S.3.

We extend and generalise the MLEbin method derived in Edwards *et al.* (2017), which in

turn extended the methods from Edwards et al. (2007) and Edwards (2011) for binned movement data. The aim is to obtain the likelihood functions to calculate the maximum likelihood estimate for the exponent b in (2). The MLEbin method in Edwards *et al.* (2017) assumed that all data were binned using the same binning protocol, such that there is just one set of bin breaks that gives the breakpoints between the bins. Here we relax that assumption to allow for species-specific bin breaks, as occurs when converting binned length data into binned body-mass data using species-specific length-weight relationships.

Consider the data to consist of counts, d_{sj}, of the number of individuals of species s that are in each size bin $j = 1,2,3,...,J_s$, where J_s is the index of the final bin for species s, and there are S species labelled $s = 1,2,3,...,S$.

Let bin j cover the values of x (which could be weight, or length for fitting a length size spectrum) in the interval $[w_{sj}, w_{s,j+1})$, such that $w_{s1}, w_{s2}, ..., w_{s,J_s+1}$ define the bin breaks. For example, for species $s = 3$, bin $j = 5$ goes from $w_{3,5}$ to $w_{3,6}$. [In terms such as w_{sj} the s and j are indices; we include commas where necessary to avoid ambiguity, such as $w_{s,j+1}$ and $w_{3,6}$]. For bin $j = J_s$ the interval is $[w_{sJ_s}, w_{s,J_s+1}]$, which includes the upper bound. The sample size (total number of counts) for species s is $n_s = \sum_{j=1}^{J_s} d_{sj}$, and the overall total number of counts is $n = \sum_{s=1}^{S} n_s$. We assume that, for each species, the first and last bins each have at least one individual in them (i.e. $d_{s,1}, d_{s,J_s} > 0$).

Similar calculations to those by Edwards et al. (2012) for unbinned data show that the known $\min\{w_{s1}\}_{s=1}^{S}$ and $\max\{w_{s,J_s+1}\}_{s=1}^{S}$ are the maximum likelihood estimates for $x_{\min}$ and $x_{\max}$, respectively. So $x_{\min}$ and $x_{\max}$ can be set to their respective maximum likelihood estimates, and we now only need to calculate the maximum likelihood estimate of b to fully specify the ISD (2).

The probability that the body size of an individual lies within the range $[w_{sj}, w_{s,j+1})$ is simply

(assume for now that $b \neq -1$)

$$\begin{aligned}
\mathrm{P}\left(\text{individual has body size in } [w_{sj}, w_{s,j+1}) | b\right) &= \int_{w_{sj}}^{w_{s,j+1}} C x^b \mathrm{d}x && \text{(S.8)} \\
&= \frac{C}{b+1}\left[x^{b+1}\right]_{w_{sj}}^{w_{s,j+1}} && \text{(S.9)} \\
&= \frac{C}{b+1}\left(w_{s,j+1}^{b+1} - w_{sj}^{b+1}\right), && \text{(S.10)} \\
&= \frac{\cancel{(b+1)}\left(w_{s,j+1}^{b+1} - w_{sj}^{b+1}\right)}{\cancel{(b+1)}(x_{\max}^{b+1} - x_{\min}^{b+1})}, && \text{(S.11)} \\
&= \frac{w_{s,j+1}^{b+1} - w_{sj}^{b+1}}{x_{\max}^{b+1} - x_{\min}^{b+1}}. && \text{(S.12)}
\end{aligned}$$

The individual can be from any species (not necessarily species s).

Recall that the w_{sj} are all known, since they are the species-specific bin breaks. Thus the only unknown in (S.12) is b. The remaining known quantities are the counts, d_{sj}, of the number of individuals of species s that are in each bin $[w_{sj}, w_{s,j+1})$.

Given these counts, we develop a multinomial log-likelihood function (Lawless, 2003) as follows. The log-likelihood function for the parameter b (the only unknown), given the counts $\{d_{sj}\}_{j=1}^{J_s}$ corresponding to bin breaks $\{w_{sj}\}_{j=1}^{J_s+1}$ for each species $s = 1, 2, 3, ..., S$, is

$$\begin{aligned}
l(b|\{d_{sj}\}, \{w_{sj}\}) &= \log\left[\prod_{s=1}^{S}\prod_{j=1}^{J_s}\left(\mathrm{P}(\text{individual has body size in } [w_{sj}, w_{s,j+1})|b)\right)^{d_{sj}}\right] && \text{(S.13)} \\
&= \sum_{s=1}^{S}\sum_{j=1}^{J_s} d_{sj}\log\left(\mathrm{P}(\text{individual has body size in } [w_{sj}, w_{s,j+1})|b)\right) && \text{(S.14)} \\
&= \sum_{s=1}^{S}\sum_{j=1}^{J_s} d_{sj}\log\left(\frac{w_{s,j+1}^{b+1} - w_{sj}^{b+1}}{x_{\max}^{b+1} - x_{\min}^{b+1}}\right) && \text{(S.15)} \\
&= \sum_{s=1}^{S}\sum_{j=1}^{J_s} d_{sj}\left(\log\left|w_{s,j+1}^{b+1} - w_{sj}^{b+1}\right| - \log\left|x_{\max}^{b+1} - x_{\min}^{b+1}\right|\right) && \text{(S.16)} \\
&= -\log\left|x_{\max}^{b+1} - x_{\min}^{b+1}\right| \sum_{s=1}^{S}\sum_{j=1}^{J_s} d_{sj} + \sum_{s=1}^{S}\sum_{j=1}^{J_s} d_{sj}\log\left|w_{s,j+1}^{b+1} - w_{sj}^{b+1}\right| && \text{(S.17)} \\
&= -n\log\left|x_{\max}^{b+1} - x_{\min}^{b+1}\right| + \sum_{s=1}^{S}\sum_{j=1}^{J_s} d_{sj}\log\left|w_{s,j+1}^{b+1} - w_{sj}^{b+1}\right|. && \text{(S.18)}
\end{aligned}$$

The two terms inside the absolute symbols $|\cdot|$, i.e. $w_{s,j+1}^{b+1} - w_{sj}^{b+1}$ and $x_{\max}^{b+1} - x_{\min}^{b+1}$, are both positive for $b < -1$ and both negative for $b > -1$ (because $w_{s,j+1} > w_{sj}$ and $x_{\max} > x_{\min}$ by definition), such that taking their absolute values ensures that (S.15) and (S.16) are equivalent. Equation (S.18) is analogous to (A.70) in the Appendix of Edwards *et al.* (2017) for which the data have a single non-species-specific set of bin breaks, but extended here to account for the species-specific bin breaks. It cannot be analytically solved to give the maximum likelihood estimate of b (by differentiating with respect to b and setting to 0), and so numerical methods are required. It is formulated in the function `negLL.PLB.bins.species()` in our `sizeSpectra` R package.

Note that (S.13) necessarily assumes that individual body sizes are independently distributed, which may be an assumption that could be investigated in later work. However, we have not had to make any assumptions about the distribution of body sizes within each species, only about the community-level ISD.

For the case where $b = -1$, we have $C = 1/(\log x_{\max} - \log x_{\min})$ from (3) and, analogous to (S.12) we have

$$\begin{aligned} \mathrm{P}\left(\text{individual has body size in } [w_{sj}, w_{s,j+1}) | b = -1\right) &= \int_{w_{sj}}^{w_{s,j+1}} C x^{-1} \mathrm{d}x & \text{(S.19)} \\ &= C \left[\log x\right]_{w_{sj}}^{w_{s,j+1}} & \text{(S.20)} \\ &= \frac{\log w_{s,j+1} - \log w_{sj}}{\log x_{\max} - \log x_{\min}}. & \text{(S.21)} \end{aligned}$$

Analogous to (S.14), the log-likelihood function is then just

$$l(b=-1|\{d_{sj}\},\{w_{sj}\}) = \sum_{s=1}^{S}\sum_{j=1}^{J_s} d_{sj}\log\Big(\mathrm{P}\left(\text{individual has body size in } [w_{sj},w_{s,j+1})|b=-1\right)\Big) \tag{S.22}$$

$$= \sum_{s=1}^{S}\sum_{j=1}^{J_s} d_{sj}\log\left(\frac{\log w_{s,j+1}-\log w_{sj}}{\log x_{\max}-\log x_{\min}}\right) \tag{S.23}$$

$$= \sum_{s=1}^{S}\left[\sum_{j=1}^{J_s} d_{sj}\left(\log\left(\log w_{s,j+1}-\log w_{sj}\right)-\log\left(\log x_{\max}-\log x_{\min}\right)\right)\right] \tag{S.24}$$

$$= -\log\left(\log x_{\max}-\log x_{\min}\right)\sum_{s=1}^{S}\sum_{j=1}^{J_s} d_{sj} + \sum_{s=1}^{S}\sum_{j=1}^{J_s} d_{sj}\log\left(\log w_{s,j+1}-\log w_{sj}\right) \tag{S.25}$$

$$= -n\log\left(\log x_{\max}-\log x_{\min}\right) + \sum_{s=1}^{S}\sum_{j=1}^{J_s} d_{sj}\log\left(\log w_{s,j+1}-\log w_{sj}\right). \tag{S.26}$$

Note that the equivalent single-species equation (A.75) in Edwards *et al.* (2017) was incorrect – it contained the log terms that results from integrating x^{-1}, but not the log term resulting from taking the log-likelihood. The correct equation is

$$l(b=-1|\text{data}) = -n\log(\log w_{J+1}-\log w_1)+\sum_{j=1}^{J} d_j\log(\log w_{j+1}-\log w_j). \tag{S.27}$$

This is equivalent to (S.26) with $S=1$ and dropping the s subscript (and replacing $x_{\min}$ and $x_{\max}$ by their maximum likelihood estimates of w_1 and w_{J+1}, respectively). This error does not occur in the `sizeSpectra` package (which is the easiest way to use the code now), and, for completeness, has been corrected in the `negLL.PLB.binned()` function in `PLBfunctions.r` in the GitHub code repository associated with Edwards *et al.* (2017), and is documented there as Issue 7 at `https://github.com/andrew-edwards/fitting-size-spectra/issues/7` – thank you to Philip Wallhead (Norwegian Institute for Water Research, pers. comm.) who independently noticed this error and that we had not originally fully corrected it. In practice this error is very unlikely to matter, as it will only occur when the log-likelihood function `negLL.PLB.binned()`

is called for a value of precisely $b = -1$, which is unlikely to occur (due to machine precision) when, for example, minimising the negative log-likelihood function. This omission did not occur for the unbinned data (equation (A.11) in Edwards *et al.* 2017), as can be seen by the `log( log(xmax) - log(xmin) )` term in the function `negLL.PLB()` in `PLBfunctions.r`.

S.1.6.1 Non-integer counts

We have assumed in (S.13) that the counts $\{d_{sj}\}$ are integer valued. However, using the same argument developed for (S.4) and (S.6) we contend that this assumption can be relaxed, i.e. that the $\{d_{sj}\}$ can take non-integer values. This is what we have for the IBTS data, since the counts are counts per hour of trawling (so that the data are standardised). For example, if three individual Atlantic Cod whose size falls in bin j are caught during a trawl of 17 minutes, then the standardisation yields a count $d_{sj} = 3 \times 17/60 = 0.85$ which is non-integer. Averaging counts per hour across areas also generates non-integer counts.

Our approach (and R code) is applicable to any data set that has species-specific bin breaks. It can be used directly for length data or for body-mass data. For the IBTS data, calculations of the individual size distribution based on lengths (the length size spectrum), rather than body masses, would still require this approach because two of the species (herring and sprat) used 0.5-cm bins, compared to 1-cm bins for all the others. In practice, to simplify the numerical computations, the analysis could just consider two pseudo-species, $s = 1$ representing all species for which 1-cm bins were used (and the counts within each length bin just summed across all such species), and $s = 2$ for the 0.5-cm species (herring and sprat).

S.1.7 Resulting body-mass bins for IBTS data

Figures S.1 and S.2 show the equivalent of Figure 6 for the remaining 90 species of the IBTS data (and Figure S.3 shows an enlargement of part of Figure S.2).

The body-mass bin with the biggest ratio of its width to its lower bound occurs for the Black-

belly Rosefish (*Helicolenus dactylopterus*) which is species code 127251, indicated by $\times$ in Figure S.3. The bin goes from 10.29 g to 20.31 g, such that the ratio of bin-width to the lower bound is 0.97 (though this is hard to see in Figure S.3). Consequently, in our earlier eight-methods analysis that did not account for the bin structure, all individuals of this species with a body-length of 4-5 cm would get assigned a body length of 4 cm, and consequently a body mass of 10.29 g. However, the true possible range of body masses is 10.29-20.31 g. Thus, individuals at the upper end of this range would be assigned body masses of about half of their actual body mass.

A similar effect occurs for *all* body masses in the data set. This systematic rounding down of body masses impacts all the methods except for the MLEbins method that explicitly accounts for the bin structure.

The widest resulting body-mass bin occurs for Atlantic Cod (*Gadus morhua*), which is species code 126436, the rightmost species in Figure S.2. The widest bin goes from 35.630 kg to 36.462 kg, with a width of 832 g.

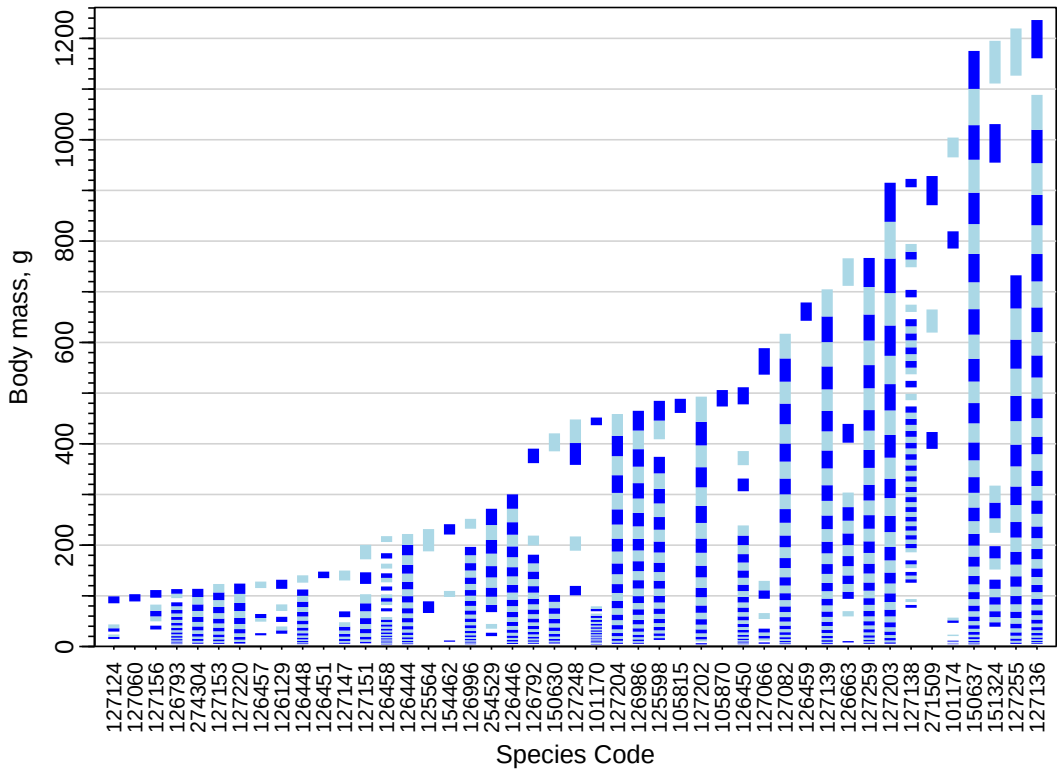

Figure S.1: As for Figure 6 but for the next 45 species – note the different vertical axis scale from Figure 6.

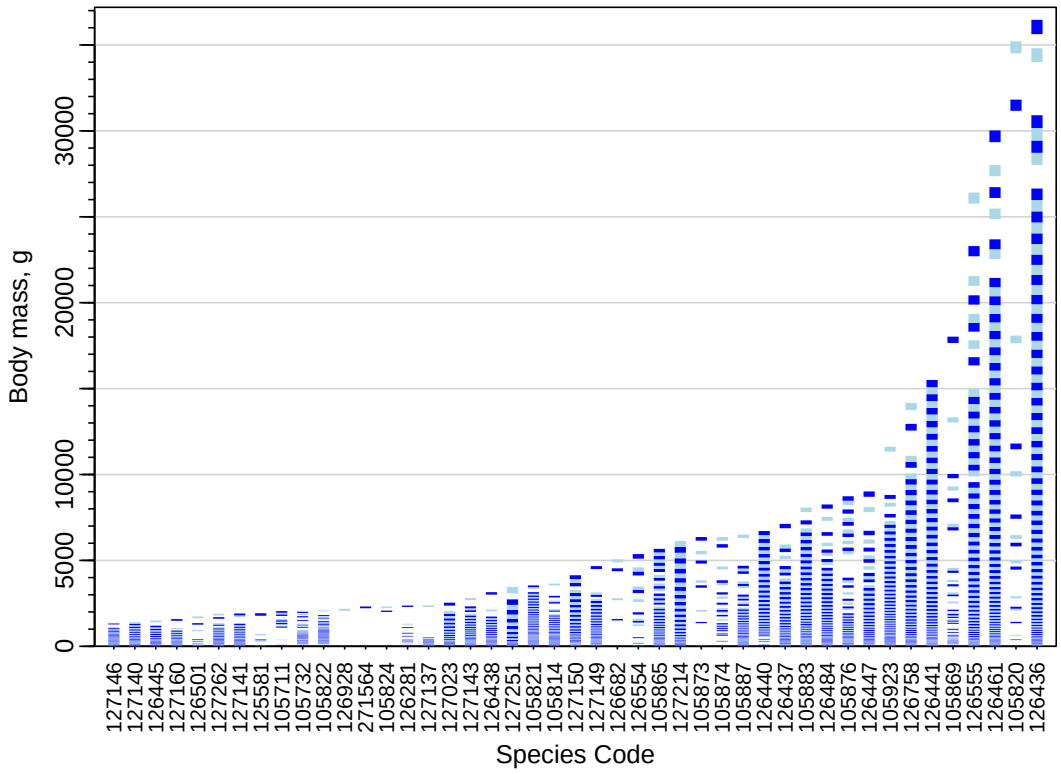

Figure S.2: As for Figure 6 but for the final 45 species – note the different vertical axis scale from Figure 6. If the figure is rotated 90° clockwise it is somewhat characteristic of a bounded power-law distribution, with a very few species growing much larger than the remaining species.

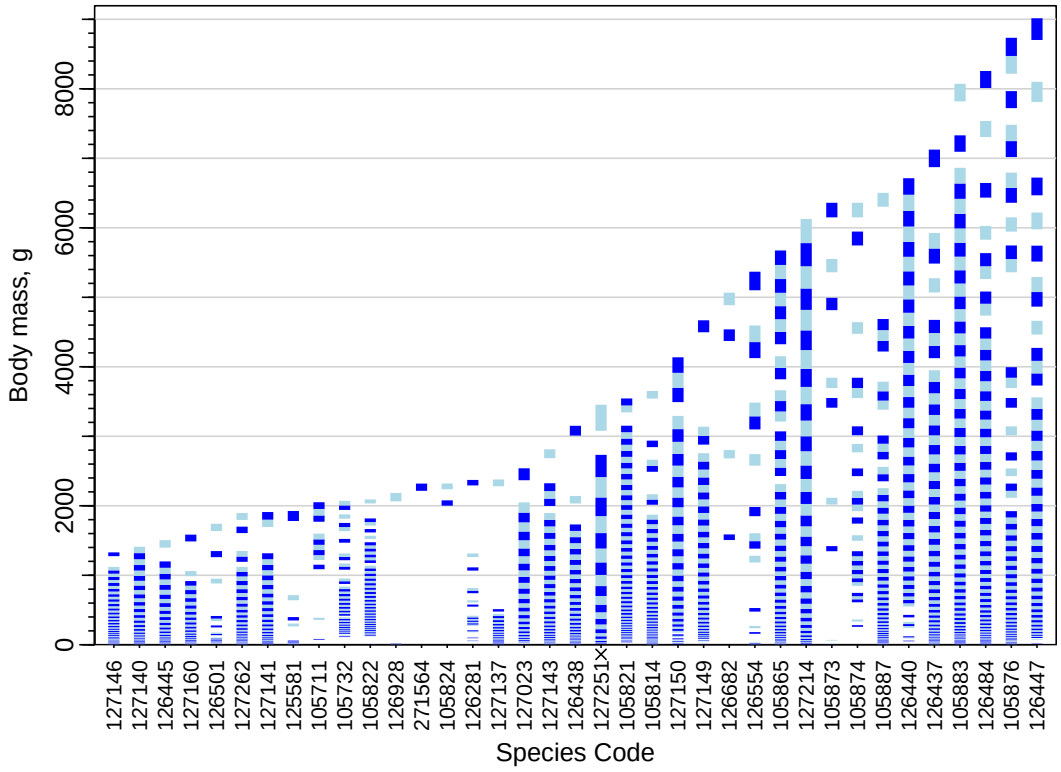

Figure S.3: As for Figure S.2 but just for species with maximum body mass up to 10 kg to more clearly show their body-mass bins. The × for species code 127251 indicates Blackbelly Rosefish (*Helicolenus dactylopterus*) which has the body-mass bin with the largest ratio of width to lower bound (see text).

S.1.8 Plotting the IBTS data and the resulting PLB fit for each year

Having accounted for the species-specific body-mass bins when fitting the ISD to the IBTS data, we now describe how to plot the data (and resulting fit) in a way that also accounts for the binning. In Figure 6 of Edwards *et al.* (2017) we recommended plotting both the biomass size spectrum (in binned form) and the abundance size spectrum. However, since the biomass size spectrum requires further binning (which is problematic, as shown in Figure 3) here we focus on the abundance size spectrum.

For unbinned data, in Edwards *et al.* (2017) we suggested plotting the abundance size spectrum by ranking the body mass, x, of each individual fish in descending order from 1 (largest) to n (smallest), and plotting log(number of values $\geq x$) against log x for each individual body mass x. The fitted ISD is then plotted as $(1-F(x))n$, where $F(x)$ is the cumulative distribution function

$$F(x) = \begin{cases} \dfrac{x^{b+1} - x_{\min}^{b+1}}{x_{\max}^{b+1} - x_{\min}^{b+1}}, & b \neq -1 \\ \dfrac{\log(x/x_{\min})}{\log(x_{\max}/x_{\min})}, & b = -1, \end{cases} \tag{S.28}$$

and n is the total sample size.

For the binned IBTS data, we first try defining the body sizes using the minimum body mass for each of the bins. For bin j for species s the minimum body mass is, by definition, w_{sj}. We use D_{sj} to represent the equivalent to the unbinned 'number of values $\geq x$' , and define it as the total number of counts in all the bins that have minimum body mass $\geq w_{sj}$:

$$D_{sj} = \sum_{s'j' \text{ with } w_{s'j'} \geq w_{sj}} d_{s'j'}. \tag{S.29}$$

The summation is over combinations of s' and j' that satisfy $w_{s'j'} \geq w_{sj}$. We can then account for the range of body masses in each bin by plotting a horizontal line from w_{sj} to $w_{s,j+1}$, i.e. showing the range of each bin, on the x-axis, and D_{sj} on the y-axis; these are shown as the green horizontal bars in Figure 7 for the 1999 data.

However, we have overlapping bins (because of the species-specific length-weight relationships), but have calculated D_{sj} based only on the ranking of the minima of each bin. Only using the minima implies, for example, that a bin with range 20-22 g would be designated as containing individuals larger than a bin with range 19-25 g, even though the latter bin can contain individuals larger (e.g. of mass 24 g) than the former bin. And bins may overlap without one fully encompassing the other, such that using the minima means that a bin with range 30-35 g is considered to always contain individuals larger than a bin with range 29-34 g, even though the latter bin can again contain individuals larger than some in the former bin.

We account for such possibilities by calculating, for bin sj, the range $[E_{sj1}, E_{sj2}]$, where

$$E_{sj1} = \sum_{s'j' \text{ with } w_{s'j'} \geq w_{s,j+1}} d_{s'j'}, \tag{S.30}$$

$$E_{sj2} = \sum_{s'j' \text{ with } w_{s',j'+1} > w_{sj}} d_{s'j'}. \tag{S.31}$$

The low end of the range, E_{sj1}, accounts for bins whose minimum value equals or exceeds the maximum value of bin sj, i.e. bins with s' and j' such that $w_{s'j'} \geq w_{s,j+1}$, because individuals in such bins must be larger than (or equal to) those in bin sj as the bins do not overlap (see Figure S.4, further explained below). The high end of the range, E_{sj2}, accounts for bins whose maximum value exceeds the minimum value of bin sj, i.e. bins with s' and j' such that $w_{s',j'+1} > w_{sj}$, because individuals in such bins may be larger than those in bin sj as the bins overlap. For E_{sj2} we need to use the criteria of $>$, rather than $\geq$, to avoid erroneously counting bins whose maximum value is the minimum value of bin sj.

Figure S.4 shows these calculations for one 'target' bin ($s = 2, j = 7$) and two example species namely Moustache Sculpin and Snakeblenny (which are highlighted as 127205 and 154675, respectively, in red in Figure 6). The resulting minimum possible number (per hour) of individuals with body size larger than those in bin $s = 2, j = 7$ is $E_{2,7,1} = 0.41$, and the maximum possible number is $E_{2,7,2} = 2.71$, giving quite a range of possible values.

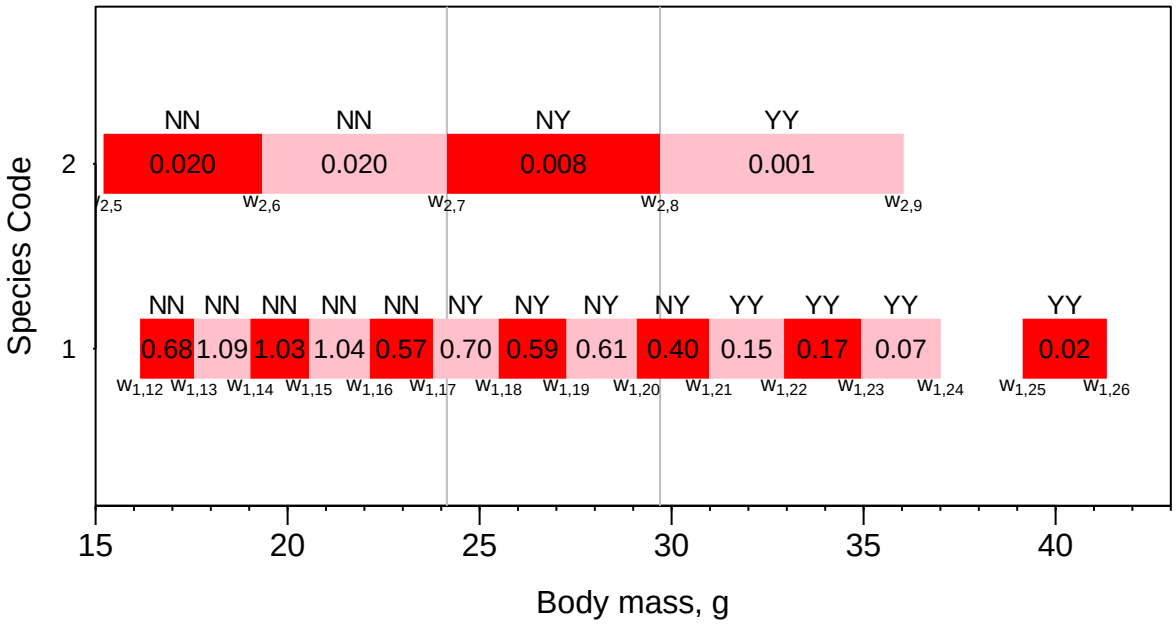

Figure S.4: Schematic diagram to explain how we calculate the range of counts of individuals that are larger than those in a given bin. Red and pink body-mass bins are those for Snakeblenny and Moustache Sculpin (here labelled species 1 and 2, respectively) from Figure 6. Bin breaks are denoted by $w_{s,j}$ and the number inside each bin is the number observed per hour of trawling. For illustration the data are combined across all years and only bins with minima > 15 g are shown. The target bin has $s = 2$ and $j = 7$ and therefore has bin breaks $w_{2,7}$ and $w_{2,8}$ and is indicated by the vertical grey lines. The first letter in each pair ('NN', 'NY', or 'YY') denotes whether or not each bin is included in the low count $E_{2,7,1}$, i.e. its lower bound is $\geq$ the upper bound of the target bin. Similarly, the second letter denotes whether or not each bin is included in the high count $E_{2,7,2}$, i.e. its upper bound is $>$ the lower bound of the target bin. Summing the respective counts as per (S.30) [first letter is 'Y'] and (S.31) [second letter is 'Y'] gives $E_{2,7,1} = 0.41$ and $E_{2,7,2} = 2.71$.

For the full data set for a particular year, similar calculations are done for each combination of s and j (i.e. all bins for all species), to give resulting ranges of counts $[E_{sj1}, E_{sj2}]$. These are shown as the vertical span of each grey rectangle in Figures 7 and S.5-S.34 (one figure for each year). The horizontal span of the rectangle depicts the range of that bin, i.e. $[w_{sj}, w_{s,j+1}]$. The figures share a common range for the x-axis.

The fitted curve, using the MLEbins method, is plotted as $(1-F(x))n$, where the total sample size is $n = \sum_{s=1}^{S} \sum_{j=1}^{J_s} d_{sj}$. The MLE for b and the low and high values of the 95% confidence intervals are also shown, but the confidence intervals are fairly narrow and their fits only just show up on the figures. Table S.2 gives, for each year, the estimated values of $x_{\min}$, $x_{\max}$, n, the MLE of b and its 95% confidence interval, and the normalisation constant C.

Visually, the PLB model seems to fit fairly well (red curve passing through the grey rectangles), at least up to 100 g, for some years (in particular 1986, 1988, 1989, 1992, 1995, 1997, 1999, 2000, 2004, 2006, 2008, 2009, 2010 and 2014). Above 100 g the PLB model generally overestimates the number of individuals. This could perhaps be indicative of fishing pressure – ideally we would have data concerning the unfished community to see how well the PLB model fit such data. Developing a formal statistical test of goodness-of-fit could be a focus of future work.

As well as no significant trend of b through time (Figure 8), Figures S.5-S.34 show no clear change in structure through the 30 years of data. For example, there is not a disappearance of larger fish through time. For providing more rigorous advice to fisheries managers, assumptions that could be investigated are the use of a minimum cutoff of 4 g (briefly investigated in Section S.1.10), and the setting of $x_{\min}$ and $x_{\max}$ to be the span of the data for each year (maybe only a smaller range is of interest, and may provide a better fit of the PLB model).

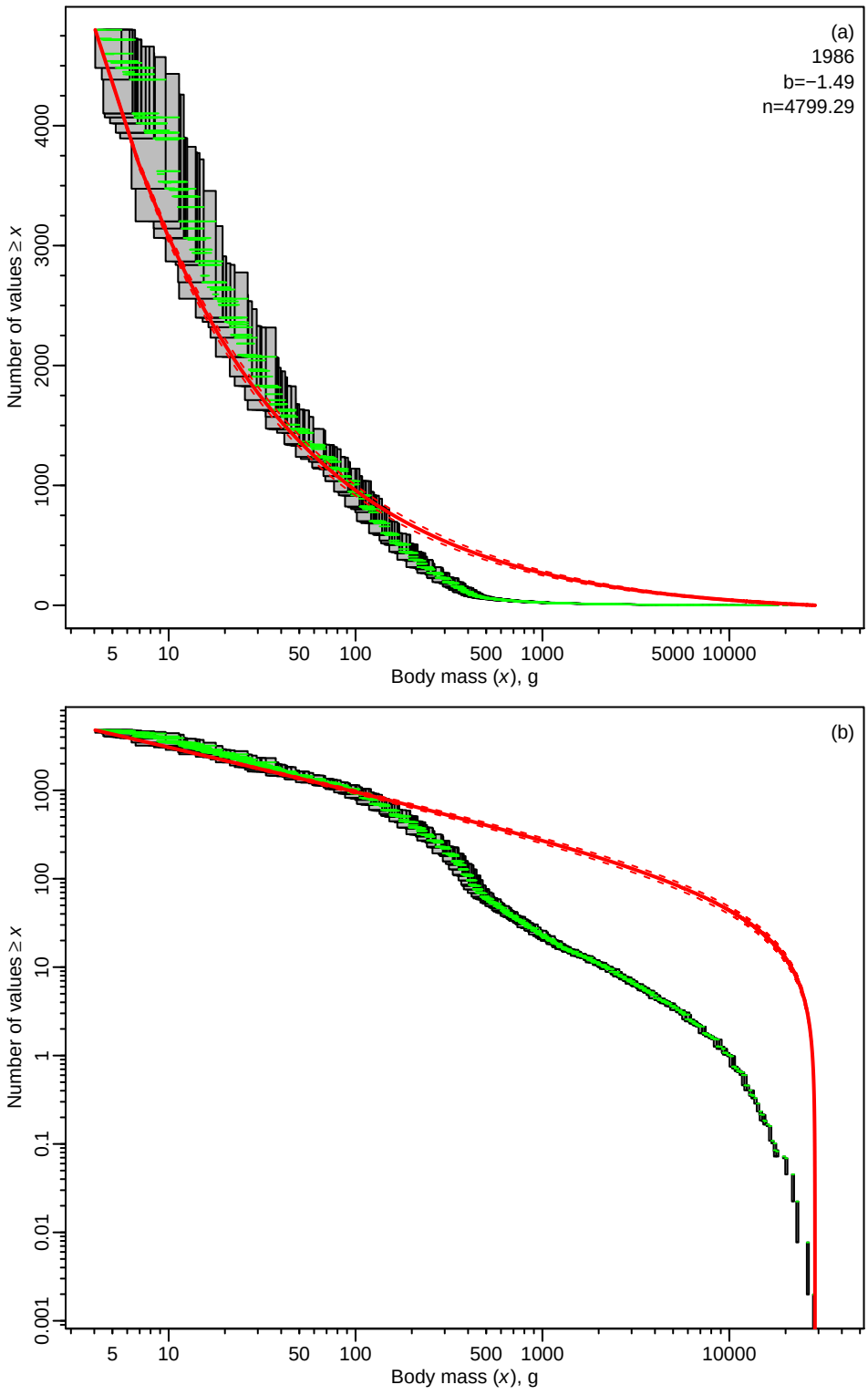

Figure S.5: Individual size distribution and MLEbins fit (red solid curve) with 95% confidence intervals (red dashed curves) for IBTS data in 1986. The y-axis is linear in (a) and logarithmic in (b). For each bin, the horizontal green line shows the range of body sizes, with its value on the y-axis corresponding to the total number of individuals (per hour of trawling) in bins whose minima are $\geq$ the bin's minimum. The vertical span of each grey rectangle shows the possible range of the number of individuals with body mass $\geq$ 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.

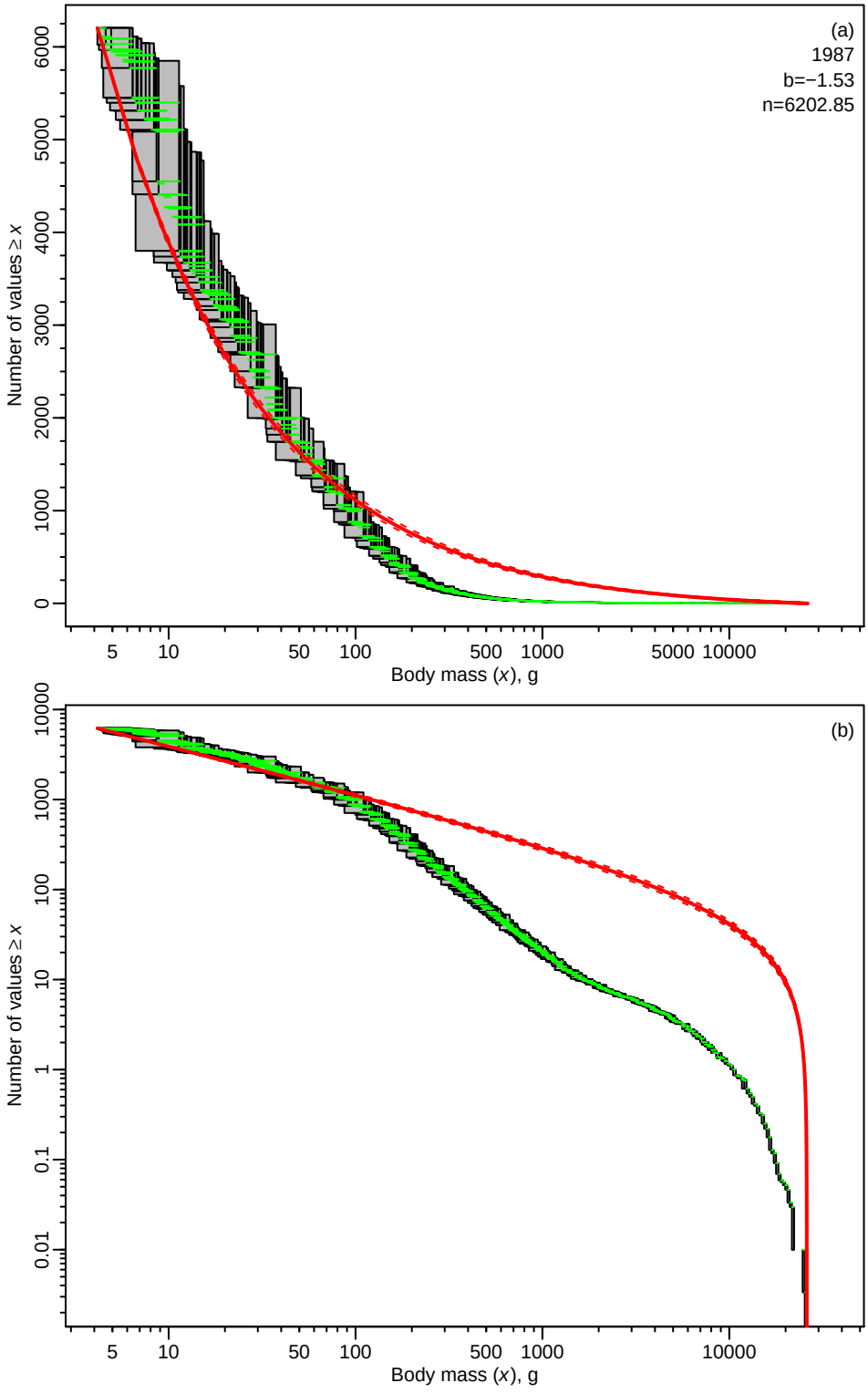

Figure S.6: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 1987. Details as in Figure S.5.

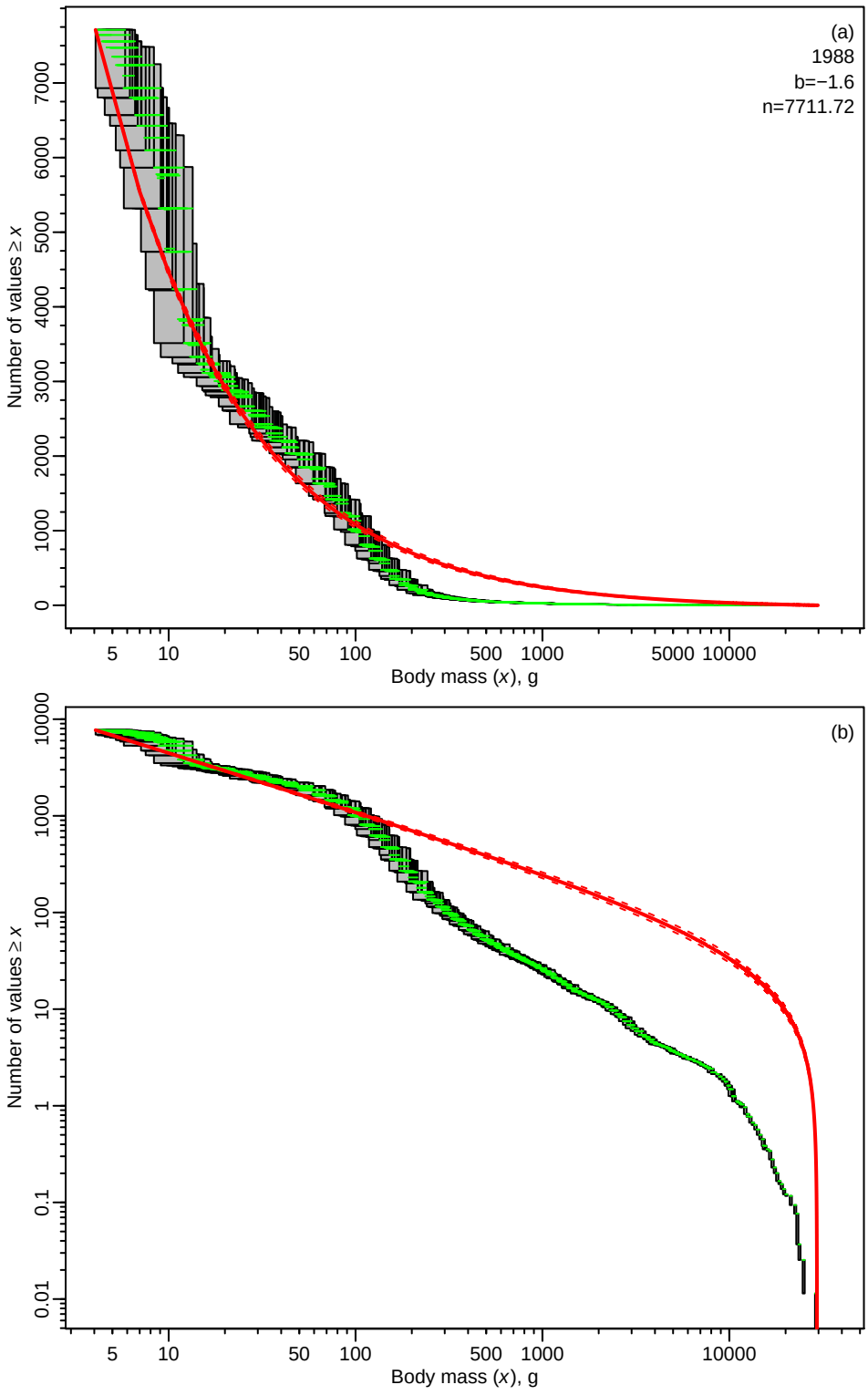

Figure S.7: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 1988. Details as in Figure S.5.

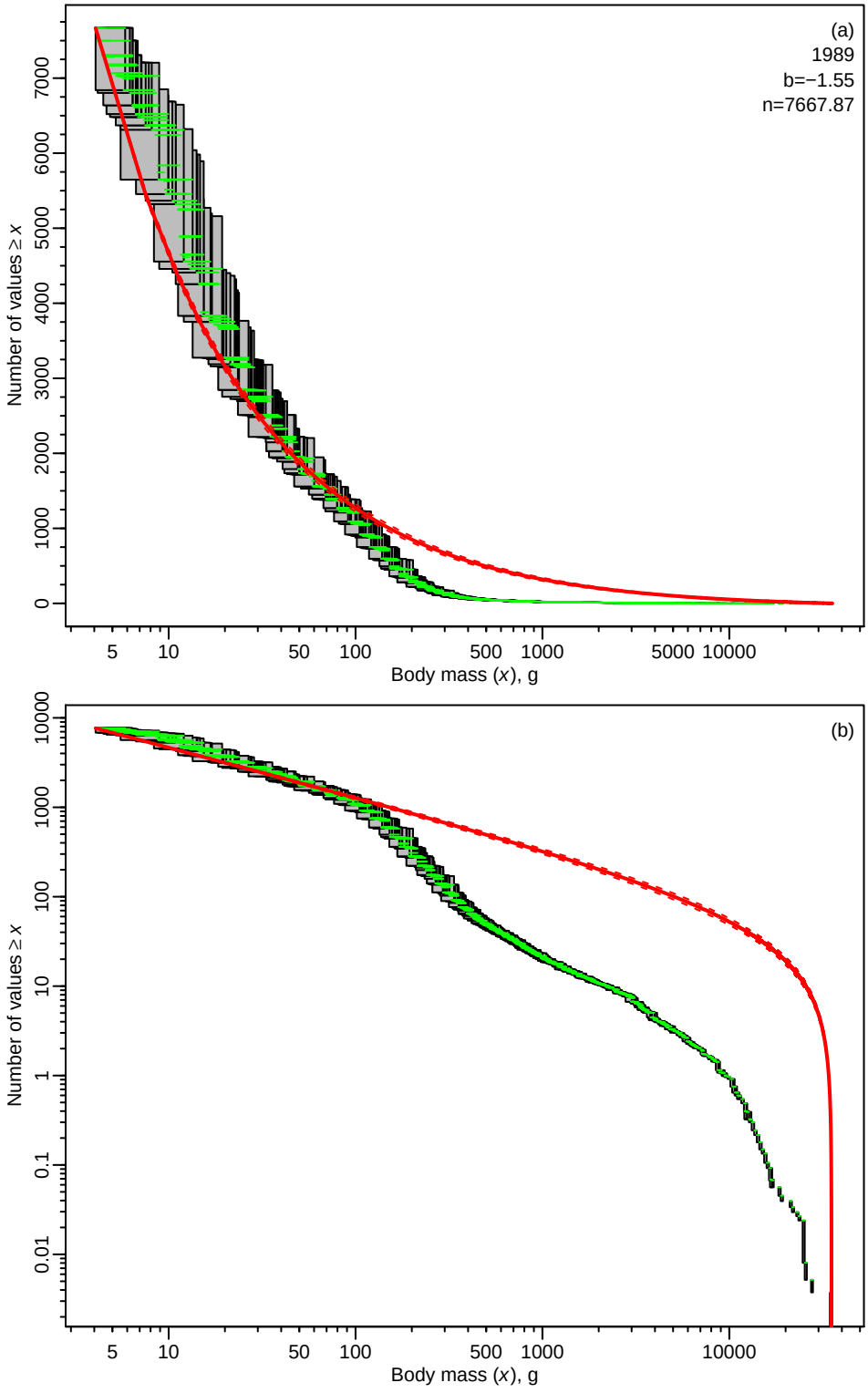

Figure S.8: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 1989. Details as in Figure S.5.

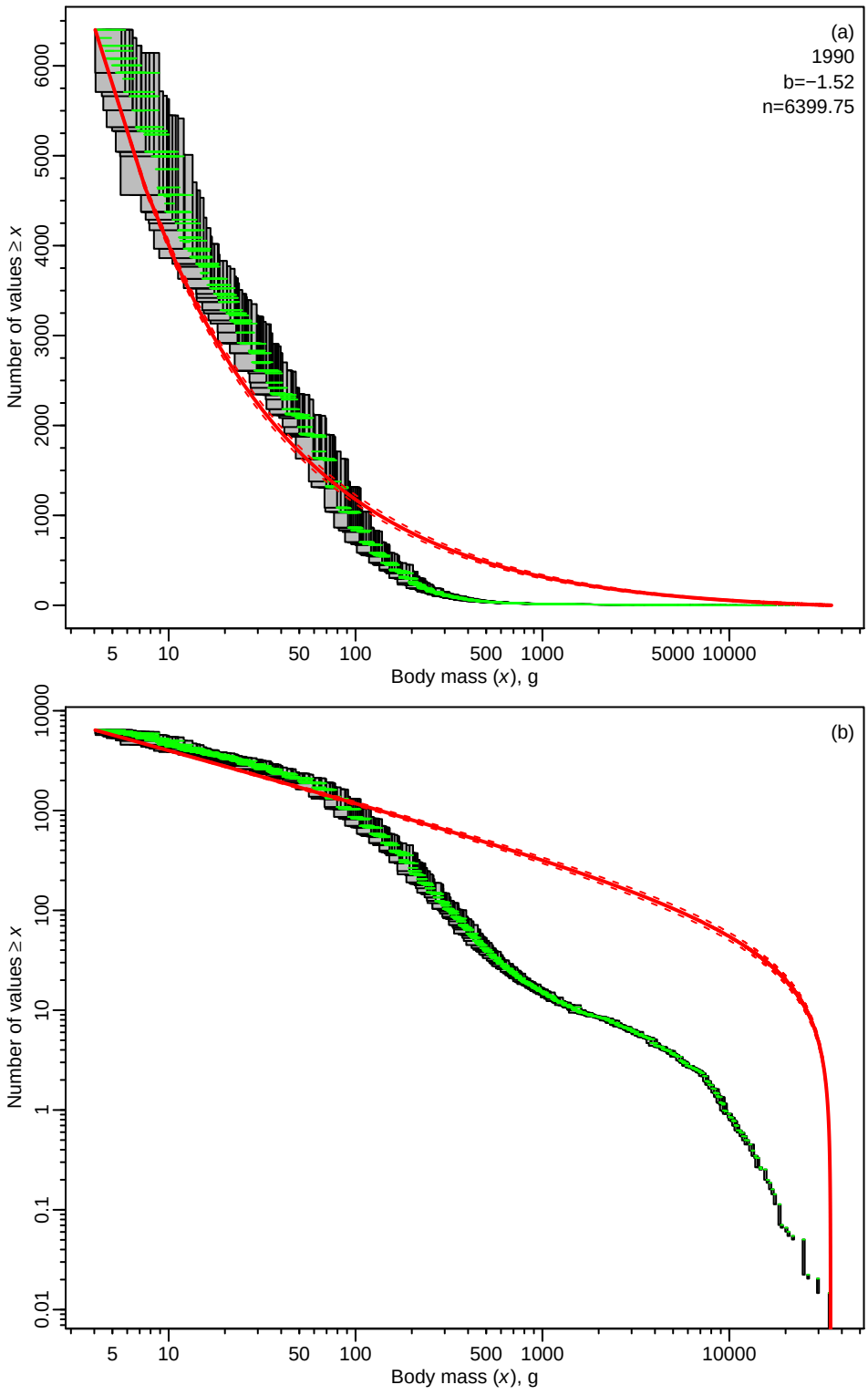

Figure S.9: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 1990. Details as in Figure S.5.

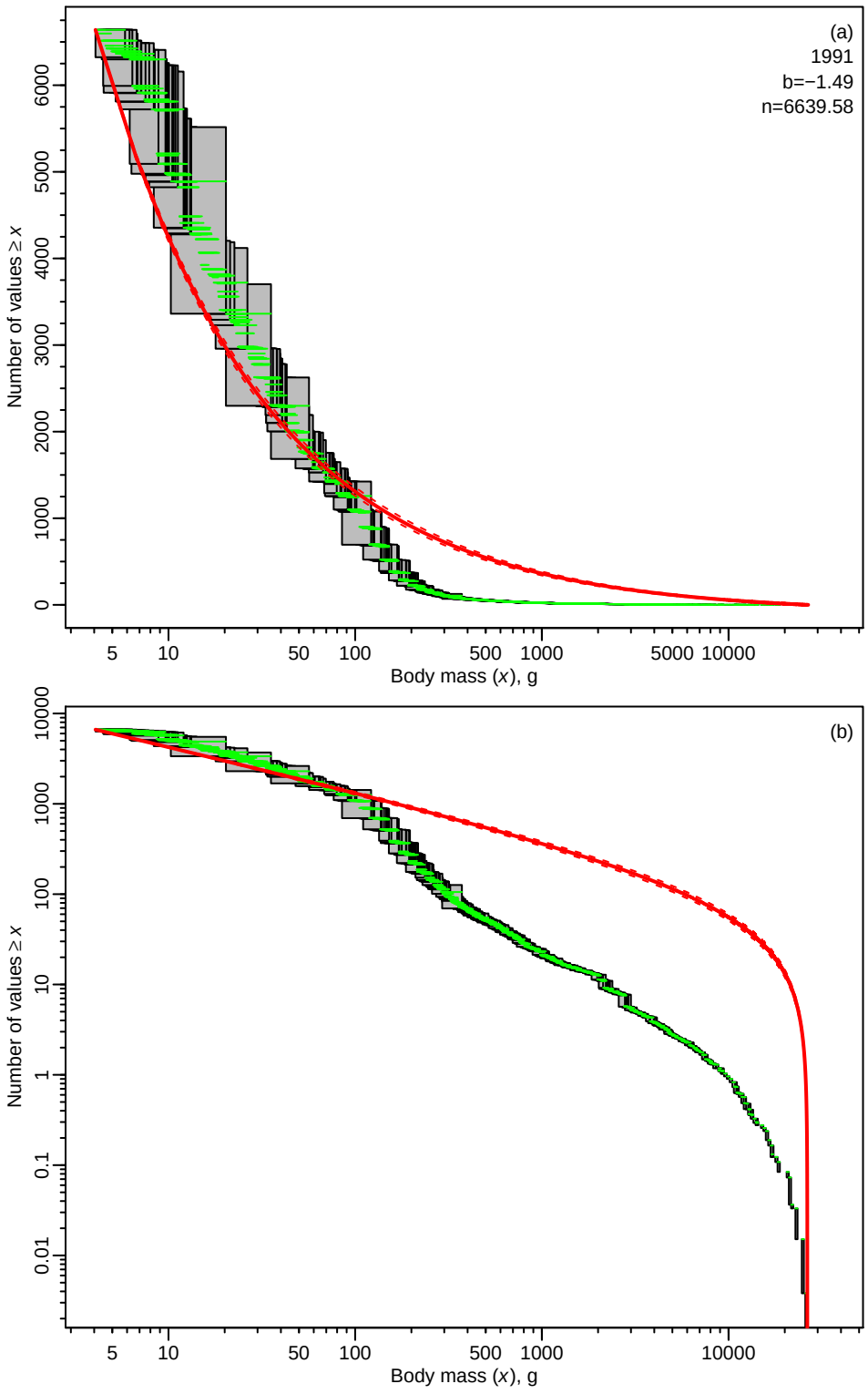

Figure S.10: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 1991. Details as in Figure S.5.

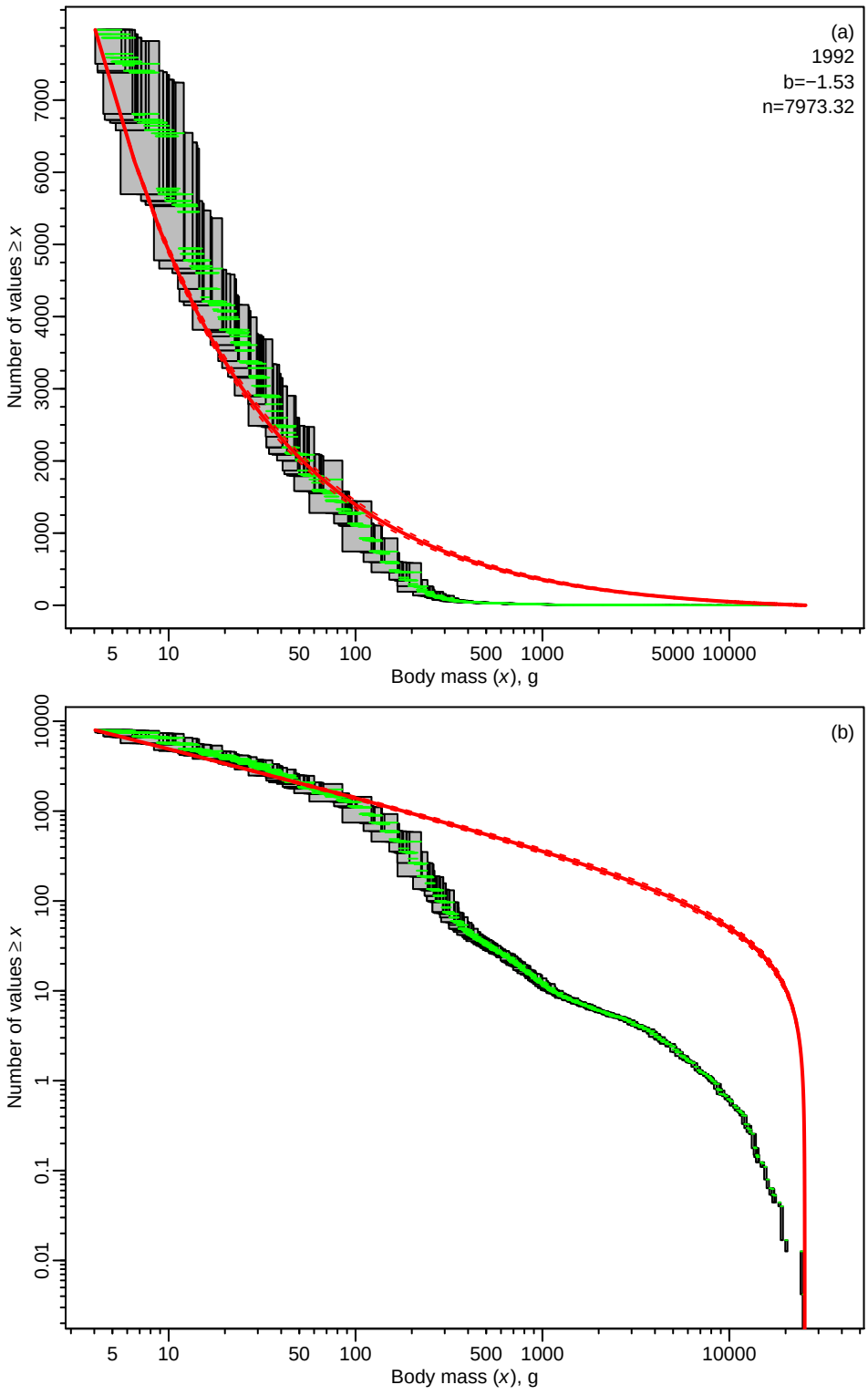

Figure S.11: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 1992. Details as in Figure S.5.

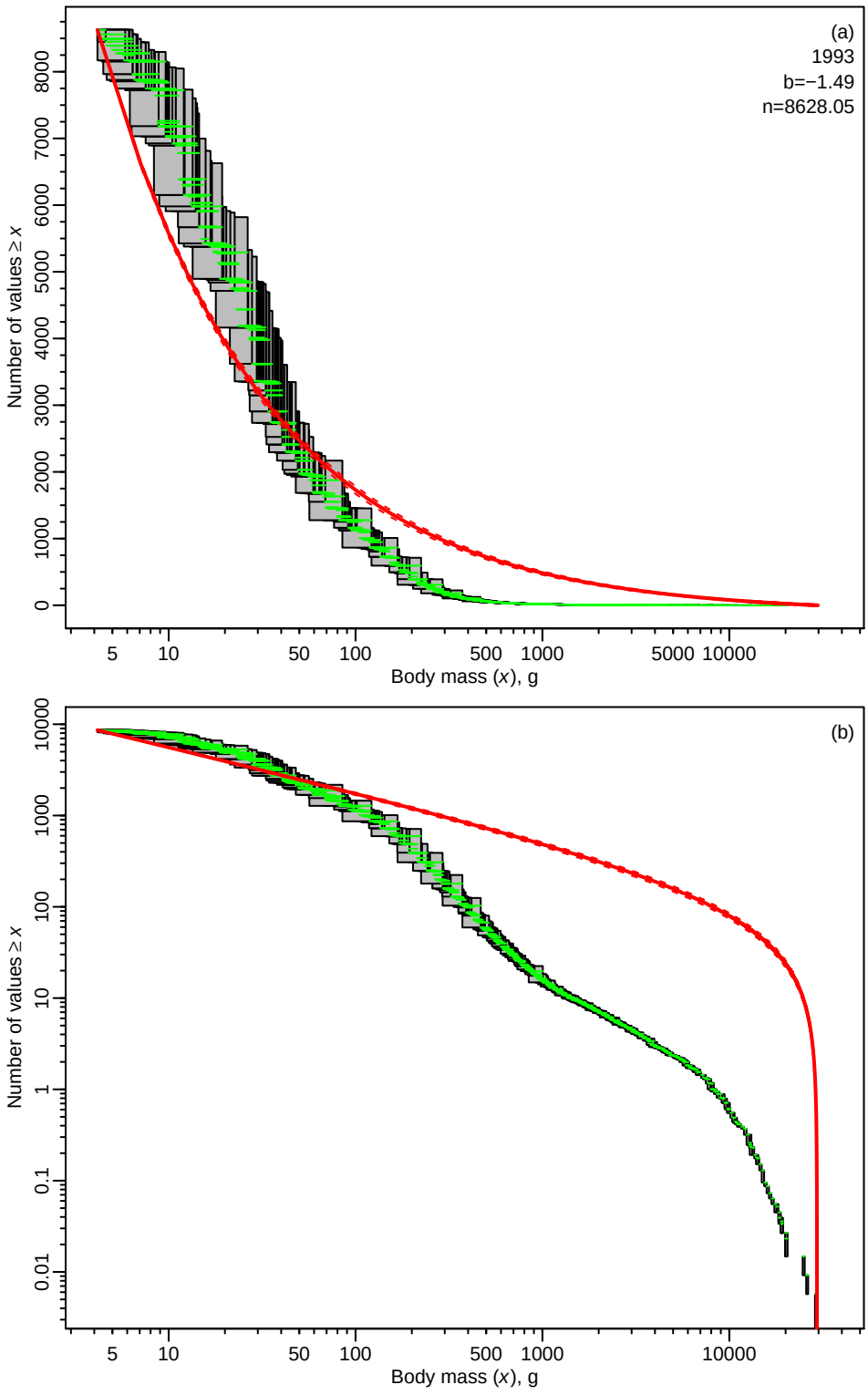

Figure S.12: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 1993. Details as in Figure S.5.

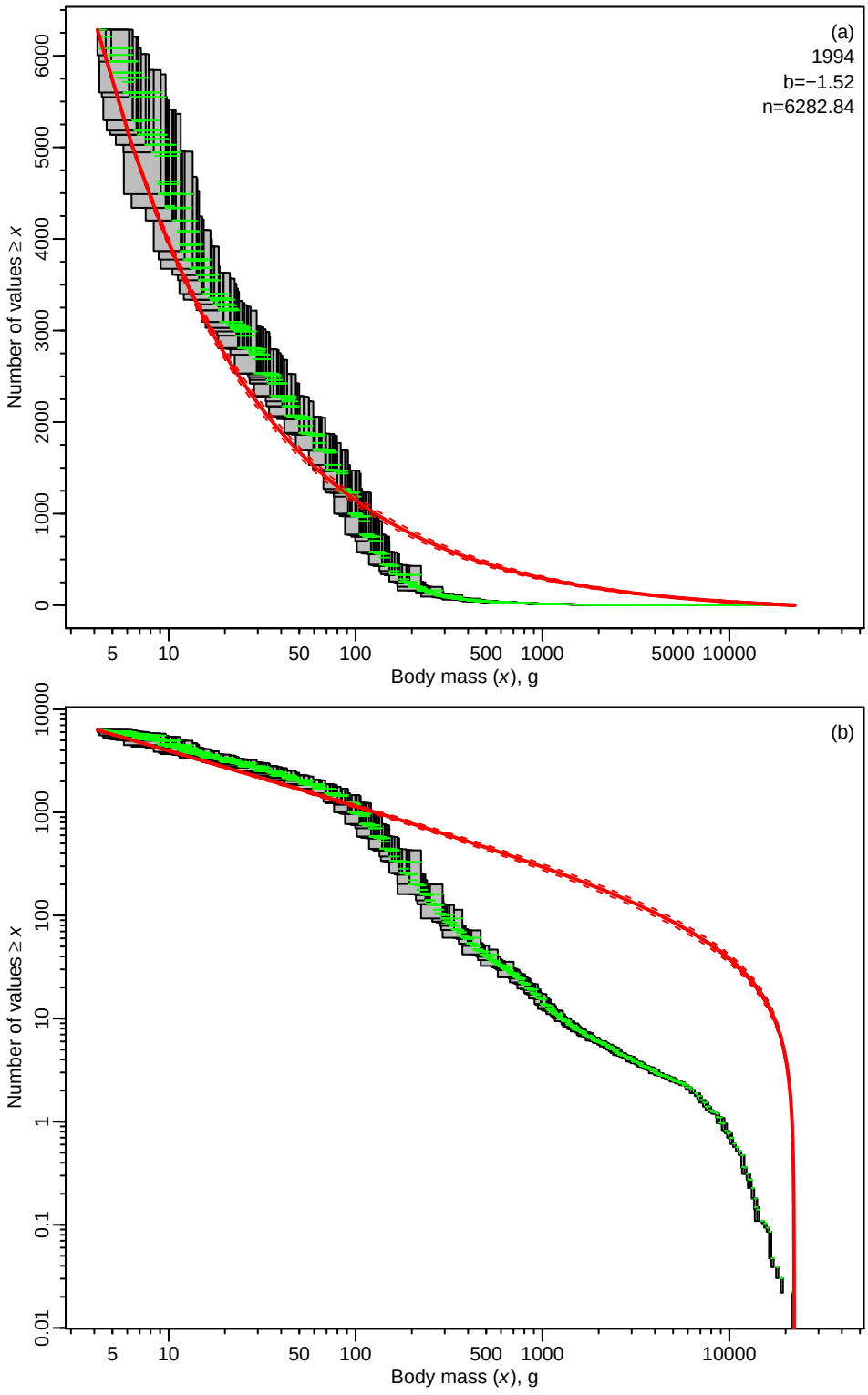

Figure S.13: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 1994. Details as in Figure S.5.

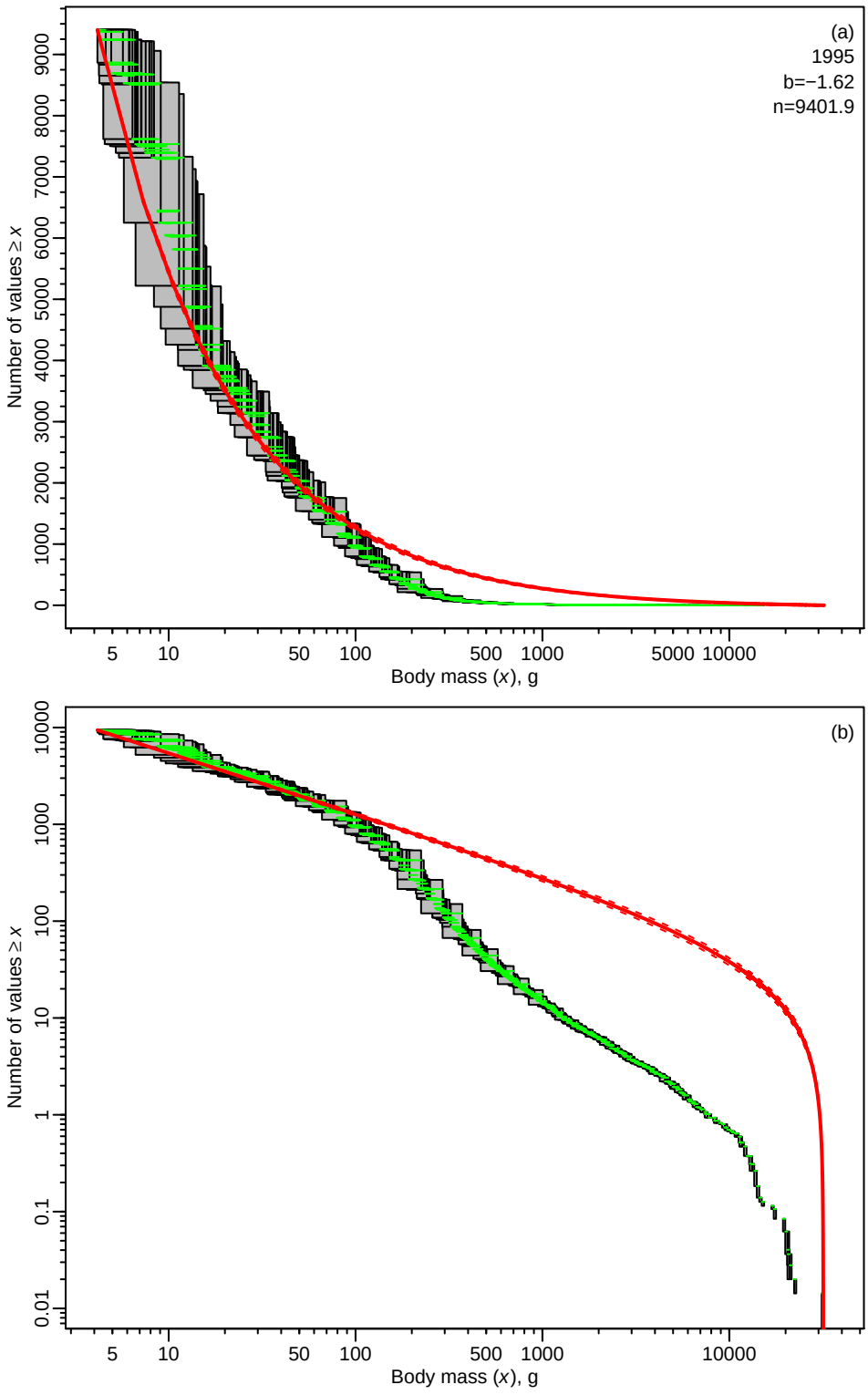

Figure S.14: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 1995. Details as in Figure S.5.

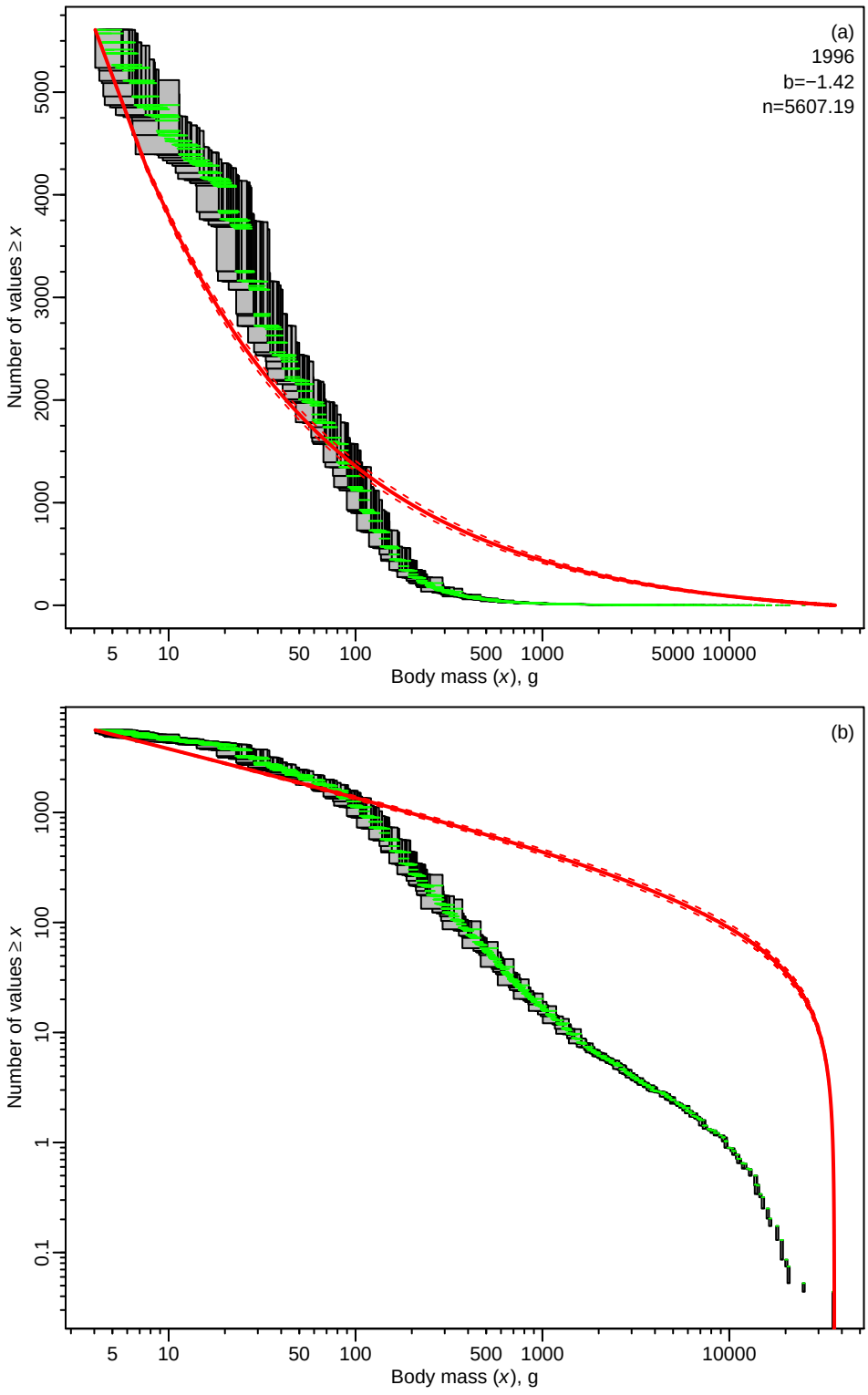

Figure S.15: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 1996. Details as in Figure S.5.

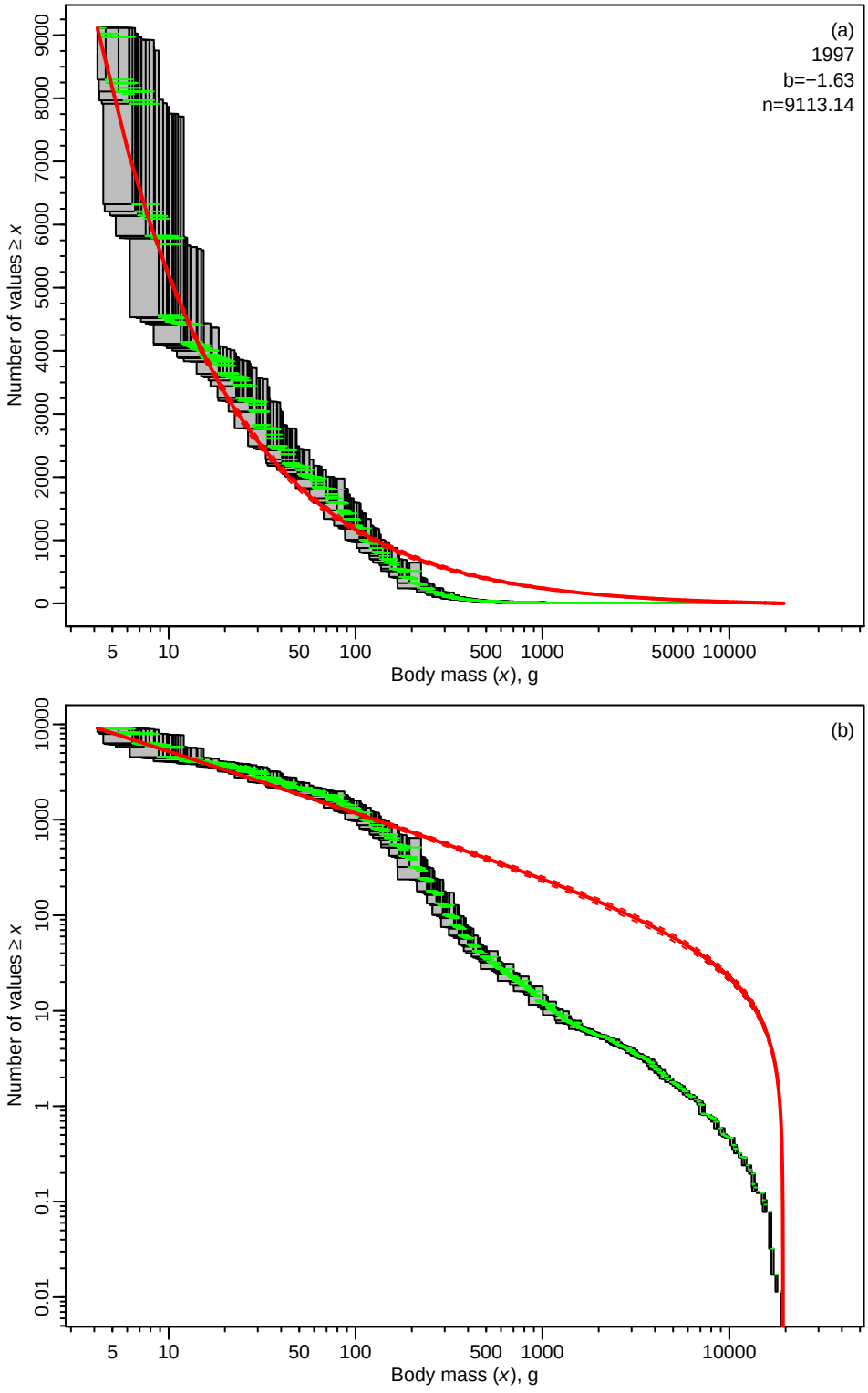

Figure S.16: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 1997. Details as in Figure S.5.

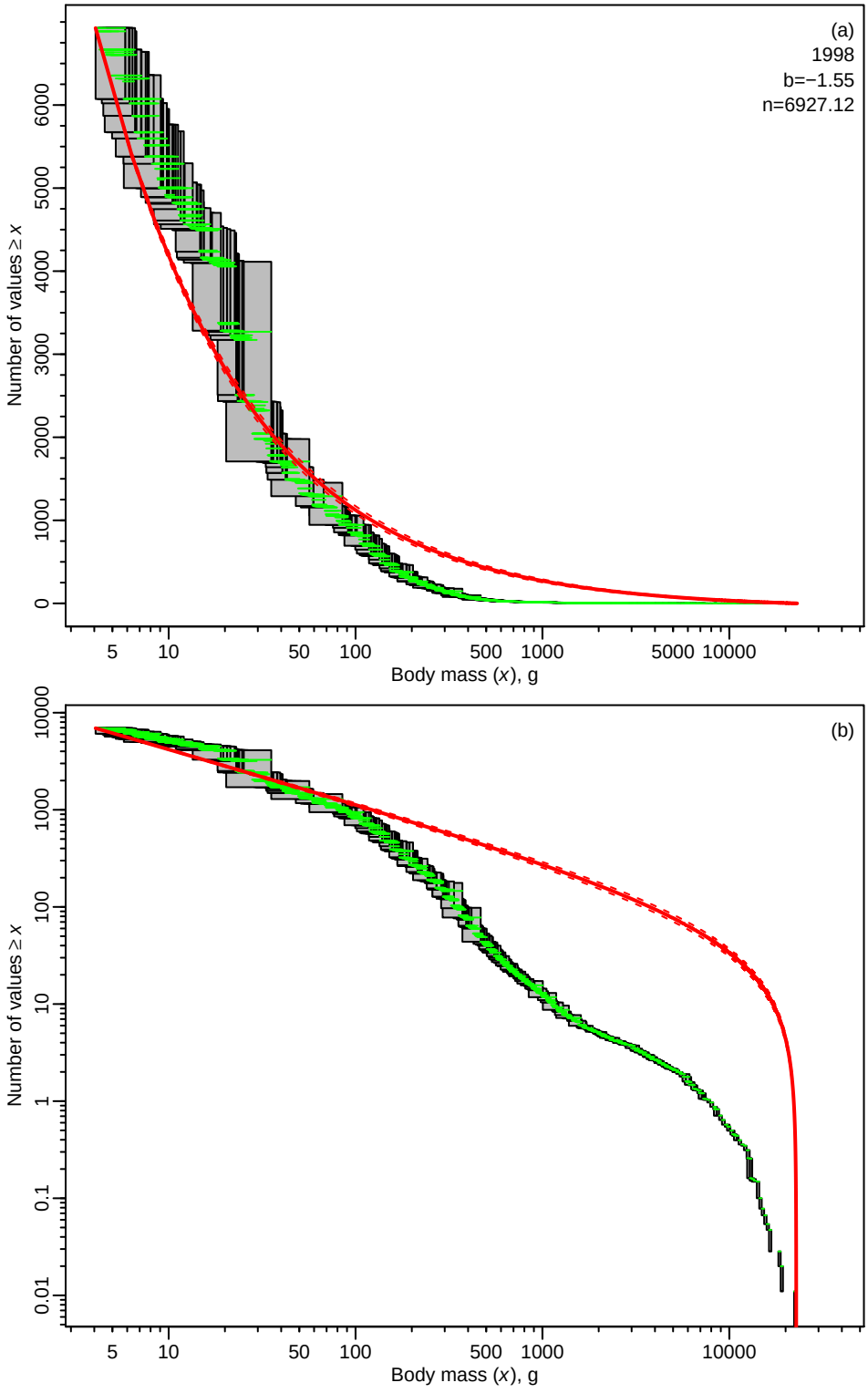

Figure S.17: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 1998. Details as in Figure S.5.

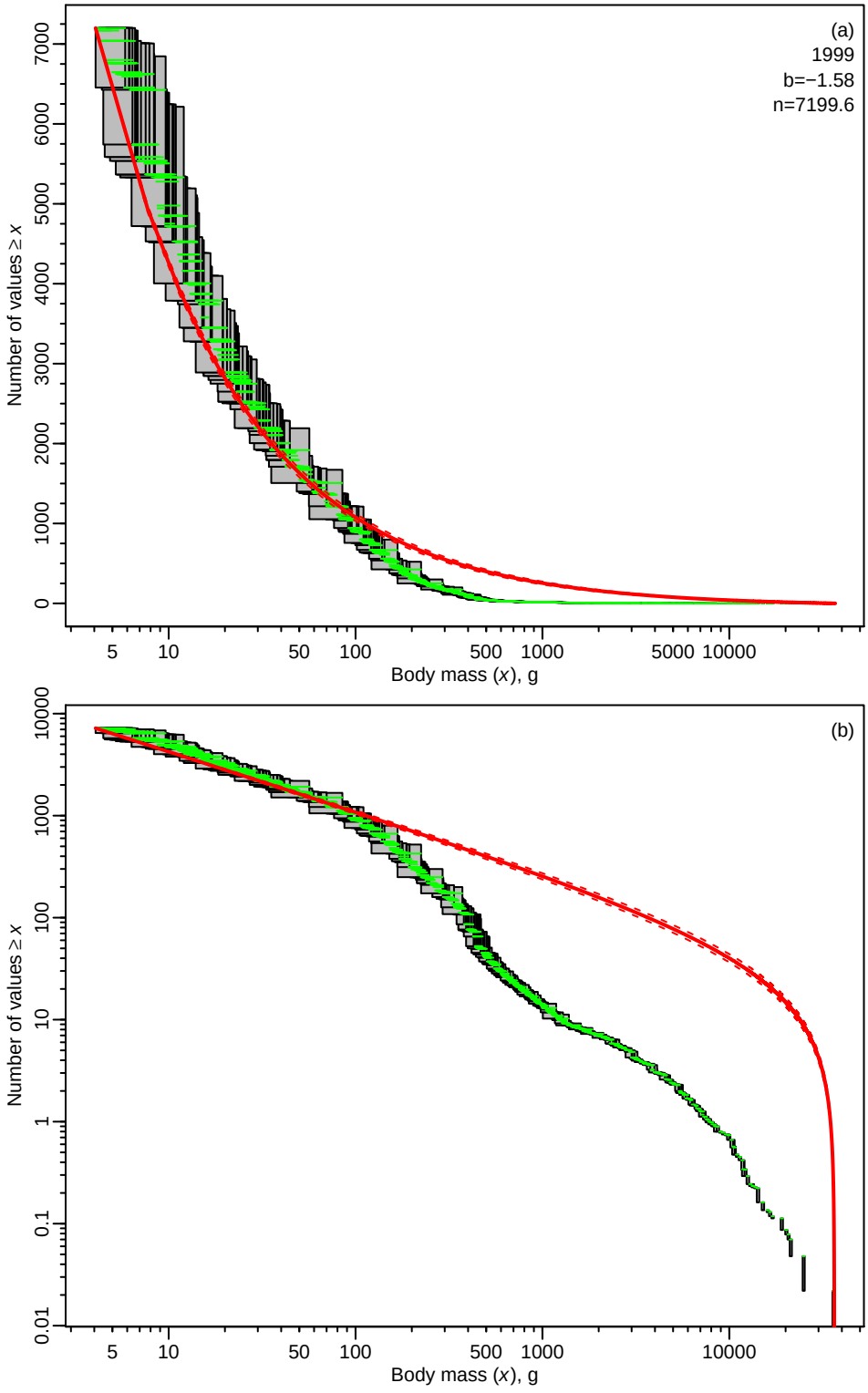

Figure S.18: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 1999. Details as in Figure S.5.

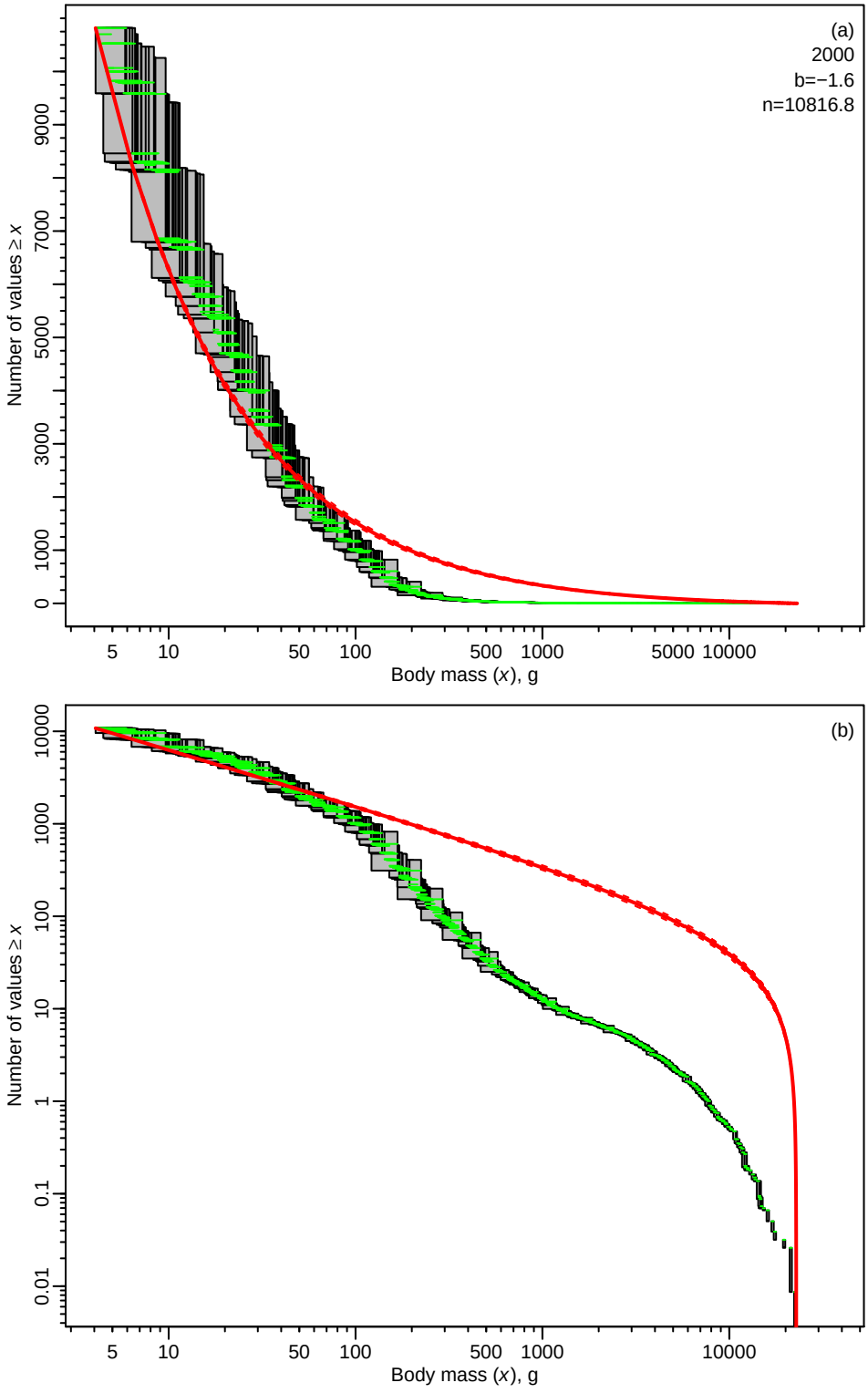

Figure S.19: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 2000. Details as in Figure S.5.

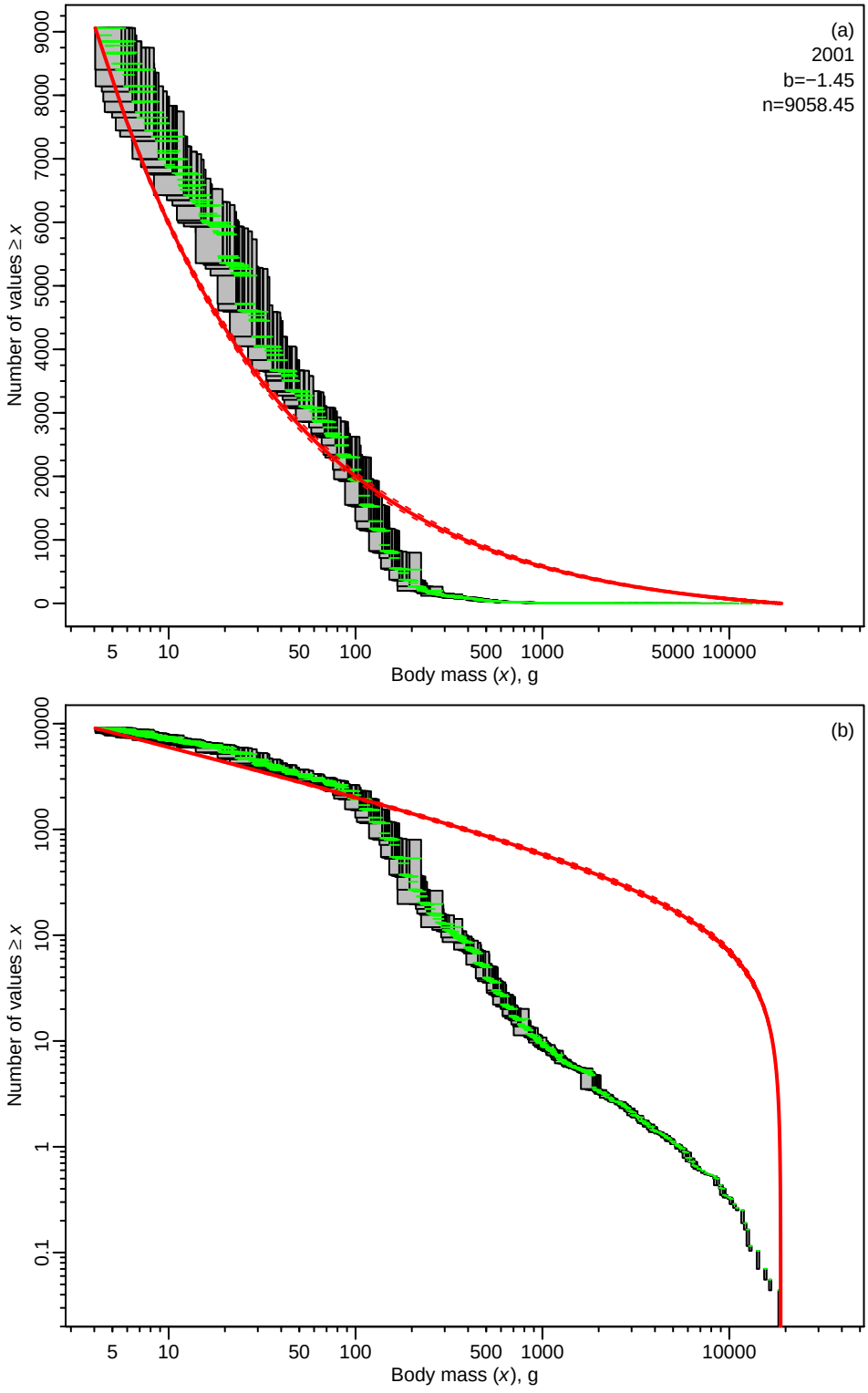

Figure S.20: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 2001. Details as in Figure S.5.

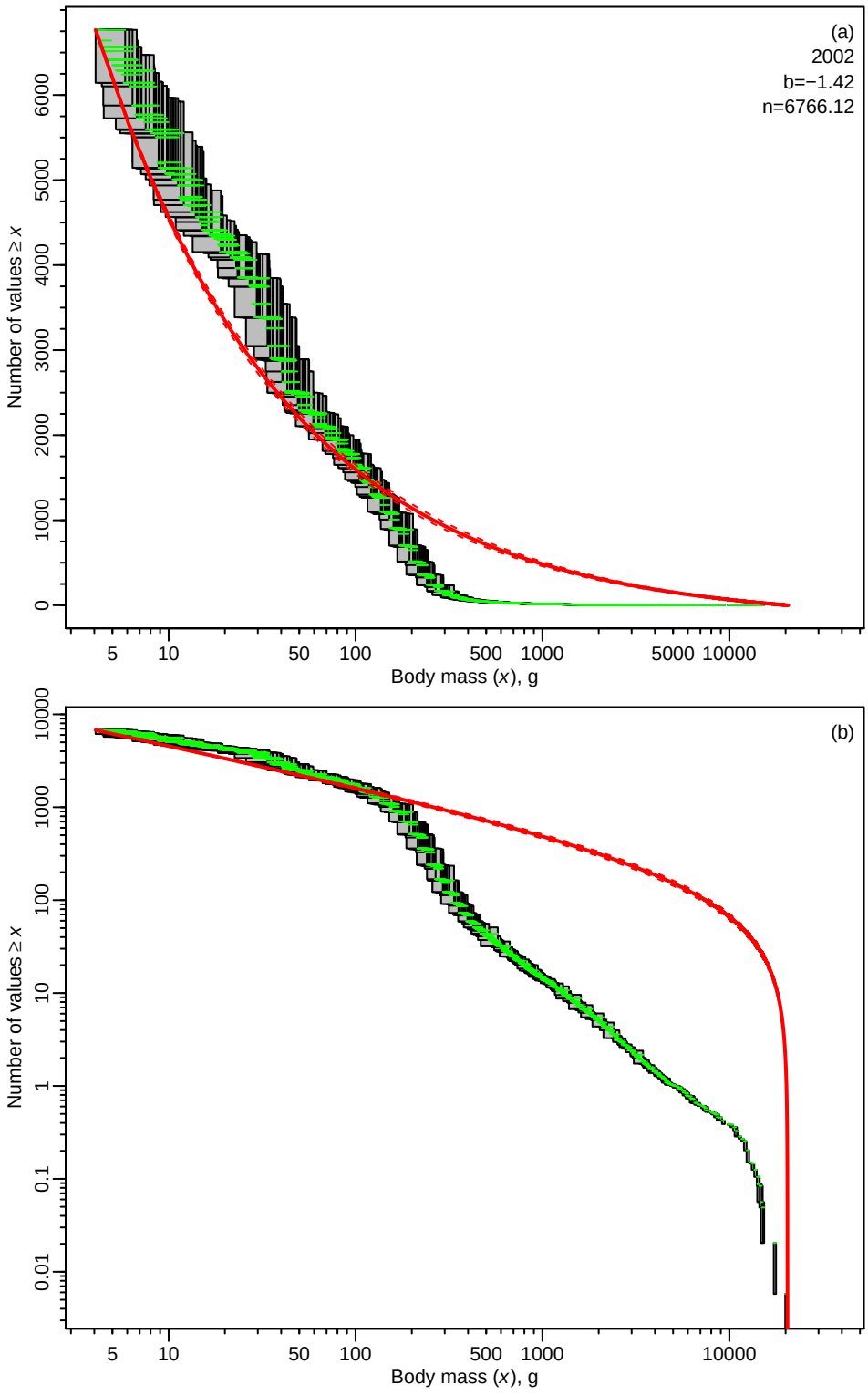

Figure S.21: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 2002. Details as in Figure S.5.

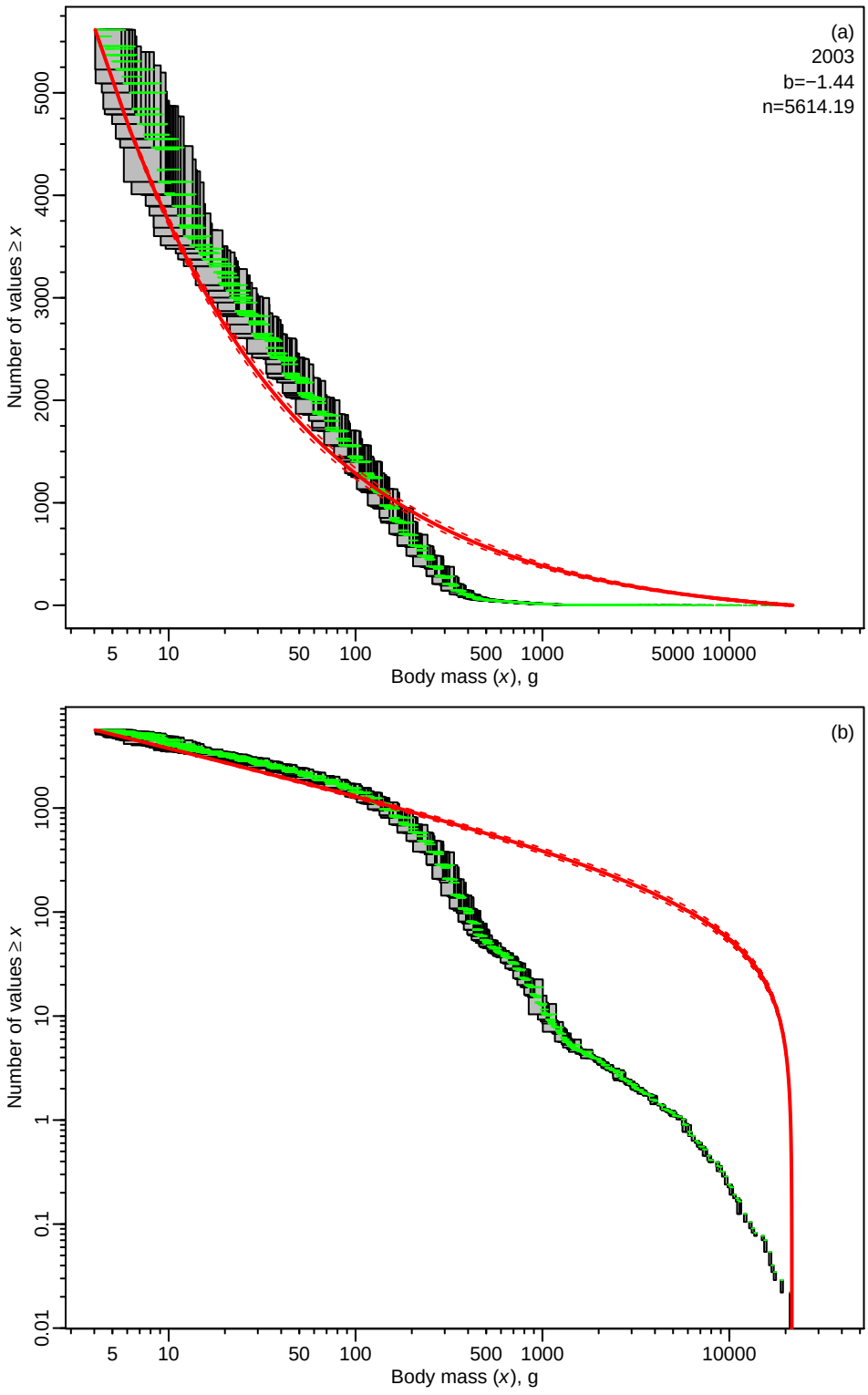

Figure S.22: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 2003. Details as in Figure S.5.

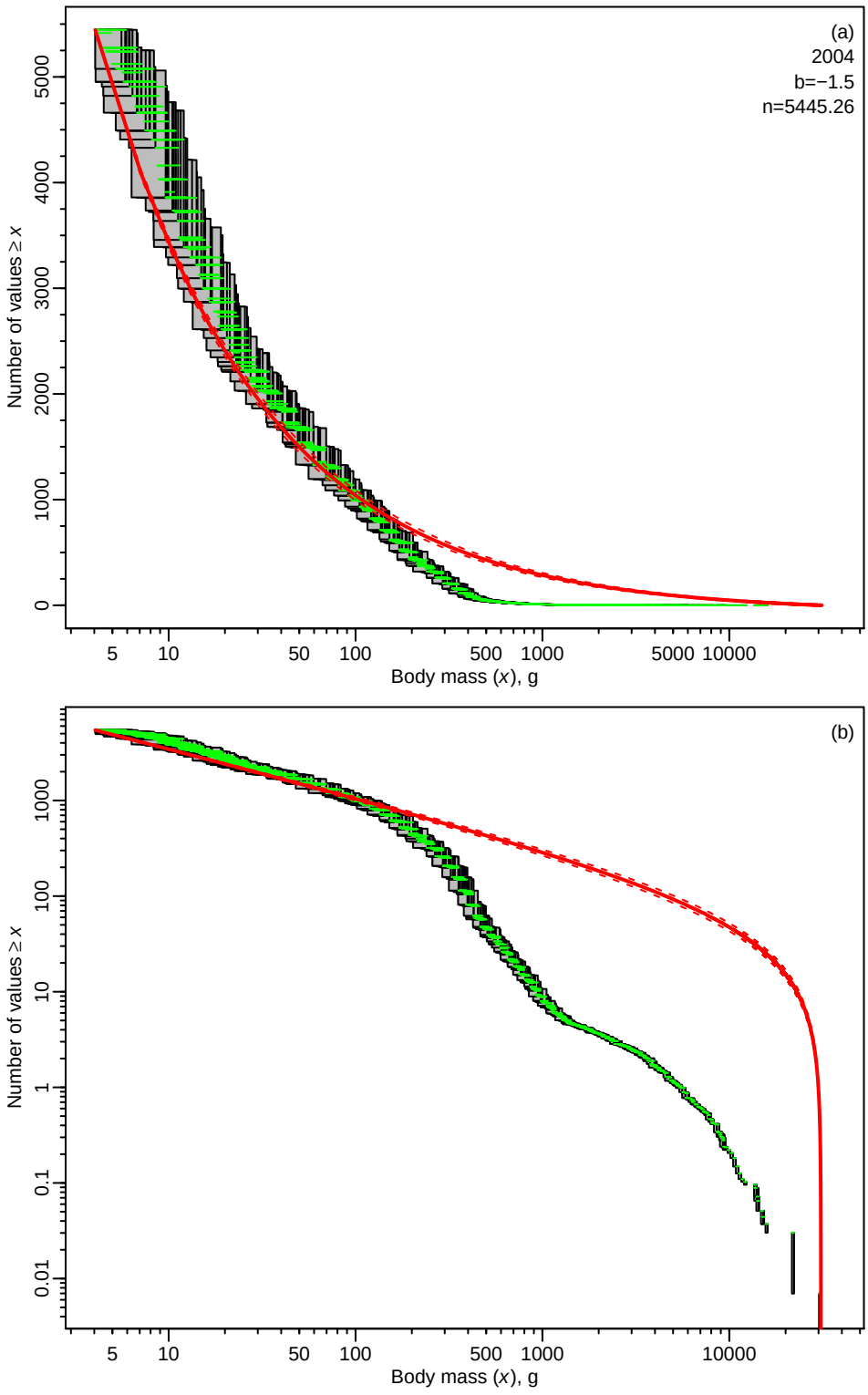

Figure S.23: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 2004. Details as in Figure S.5.

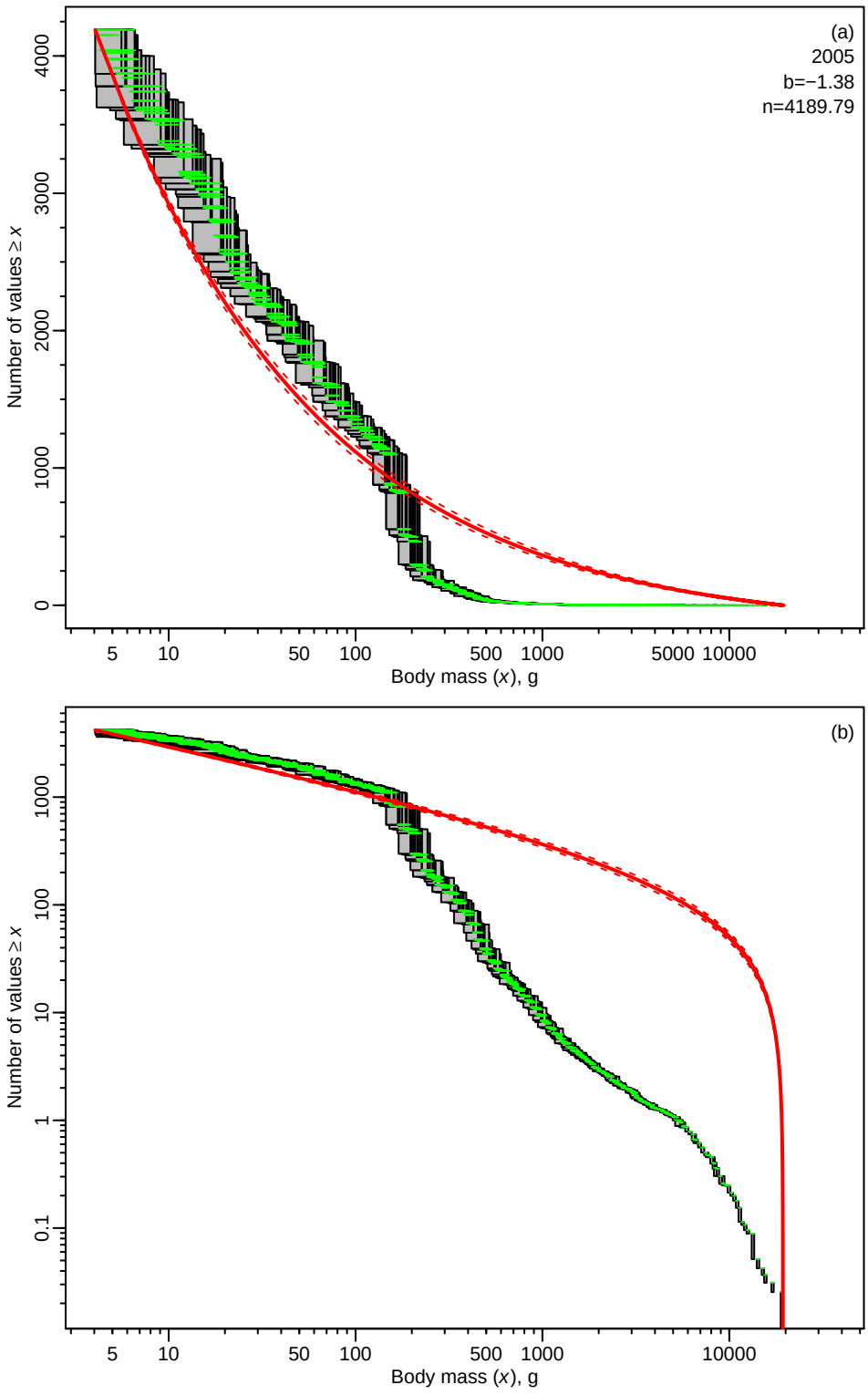

Figure S.24: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 2005. Details as in Figure S.5.

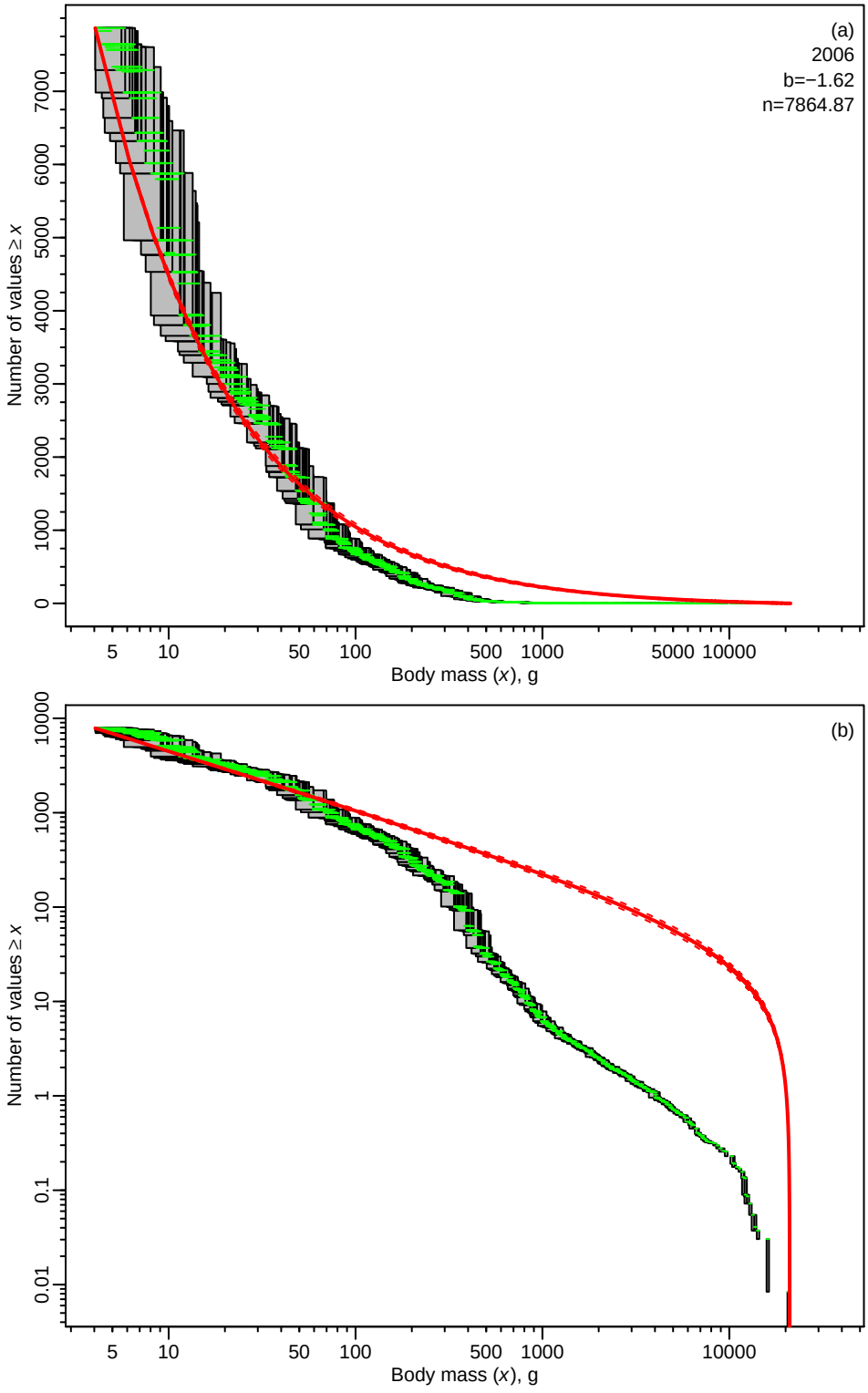

Figure S.25: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 2006. Details as in Figure S.5.

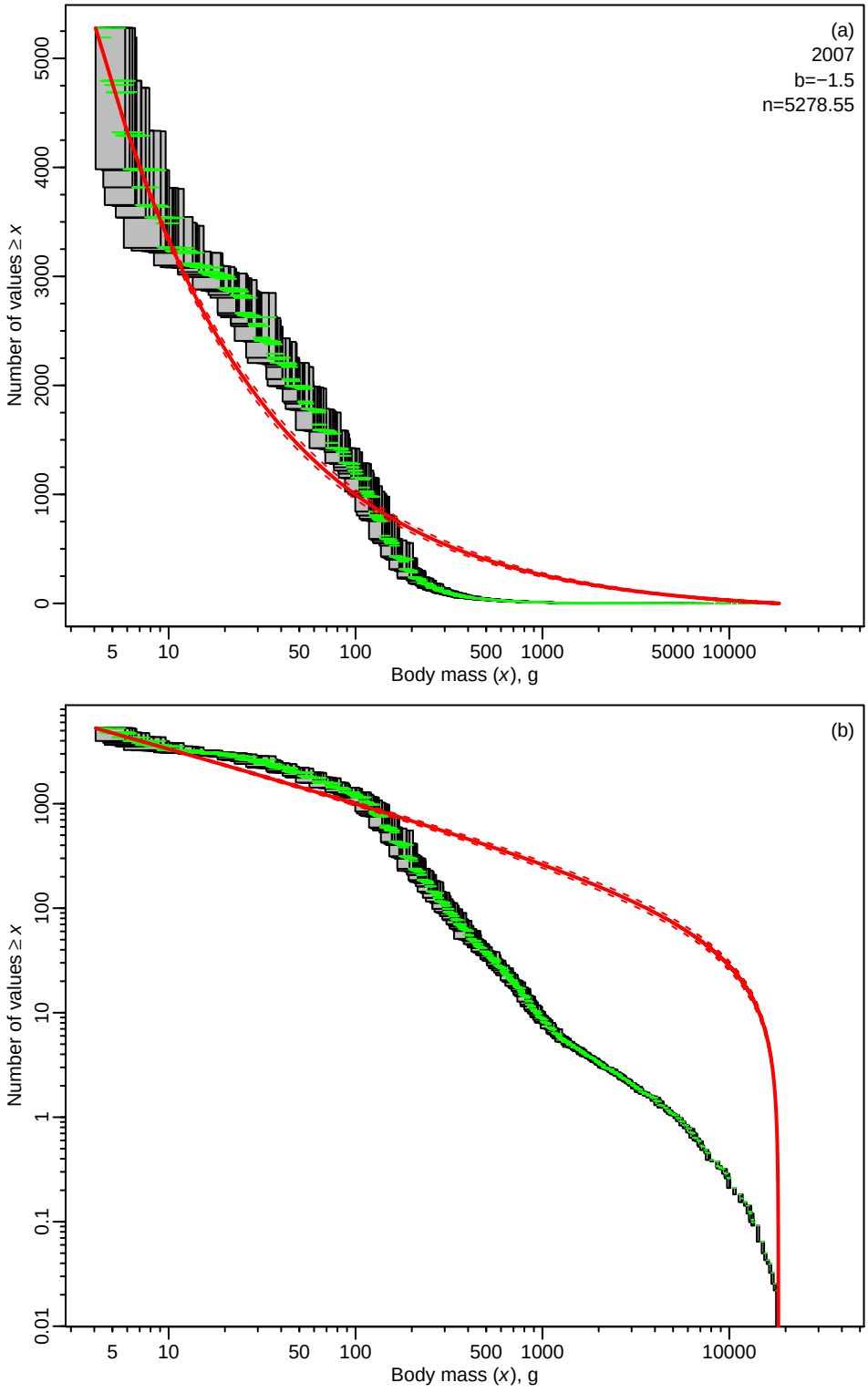

Figure S.26: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 2007. Details as in Figure S.5.

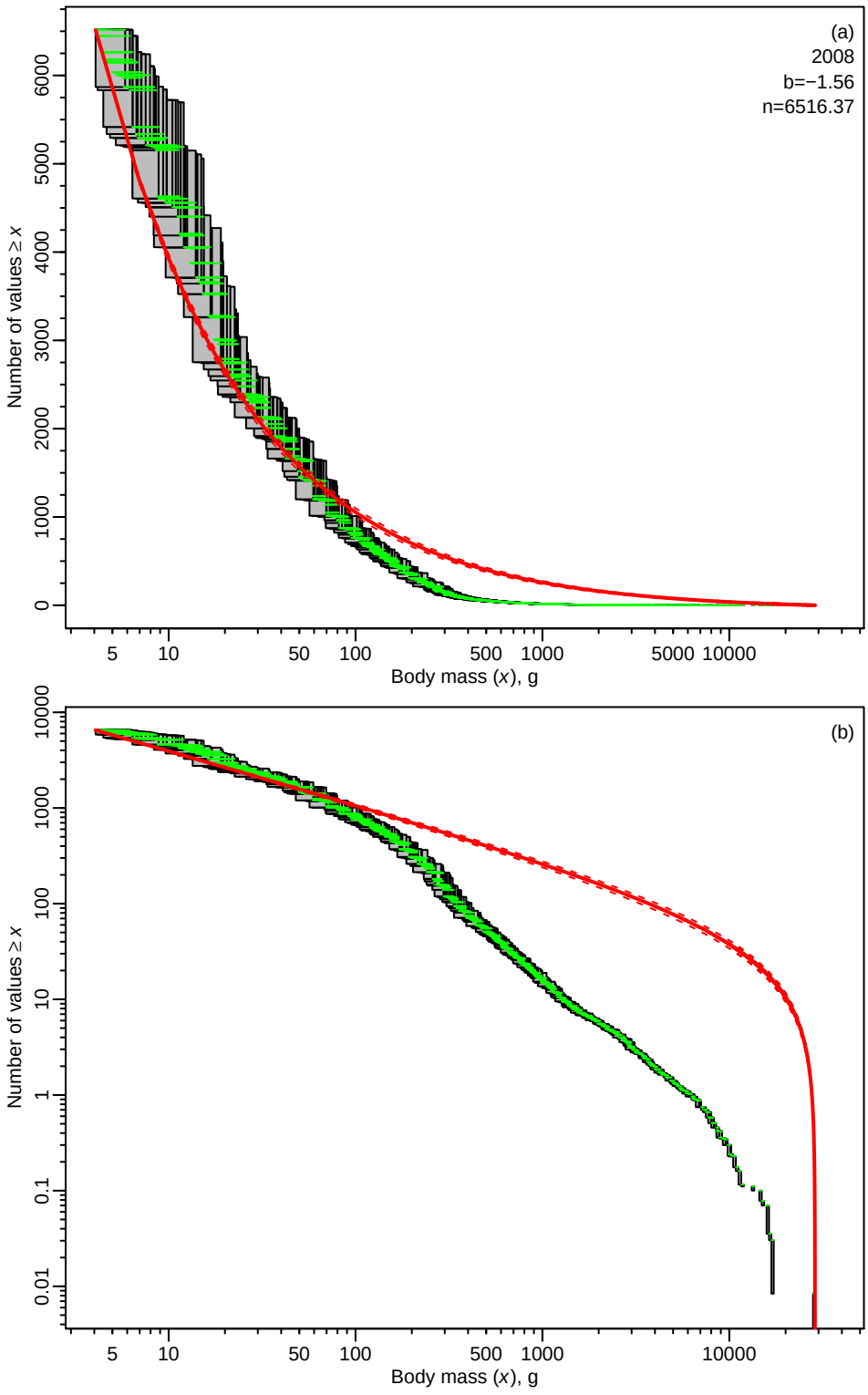

Figure S.27: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 2008. Details as in Figure S.5.

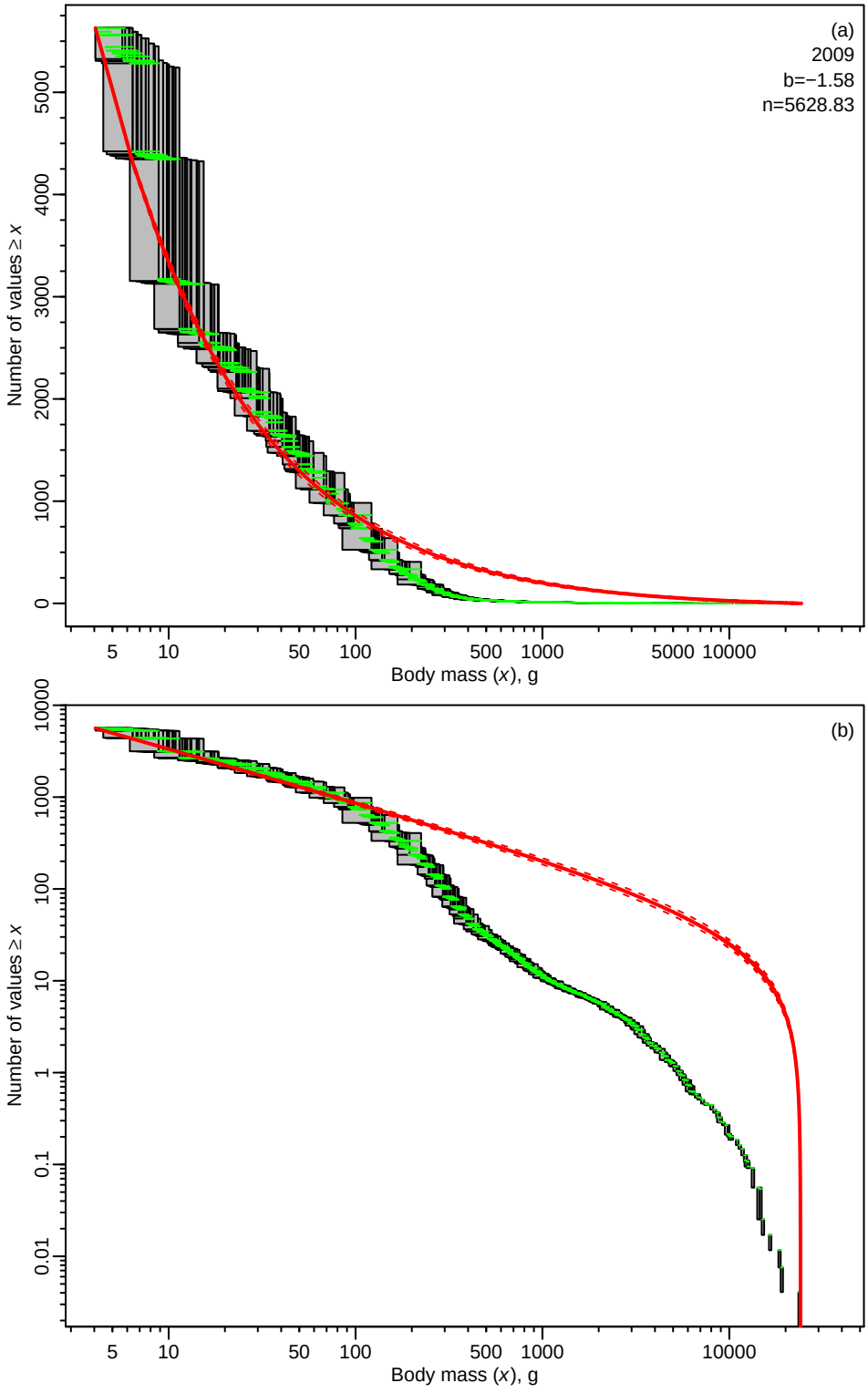

Figure S.28: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 2009. Details as in Figure S.5.

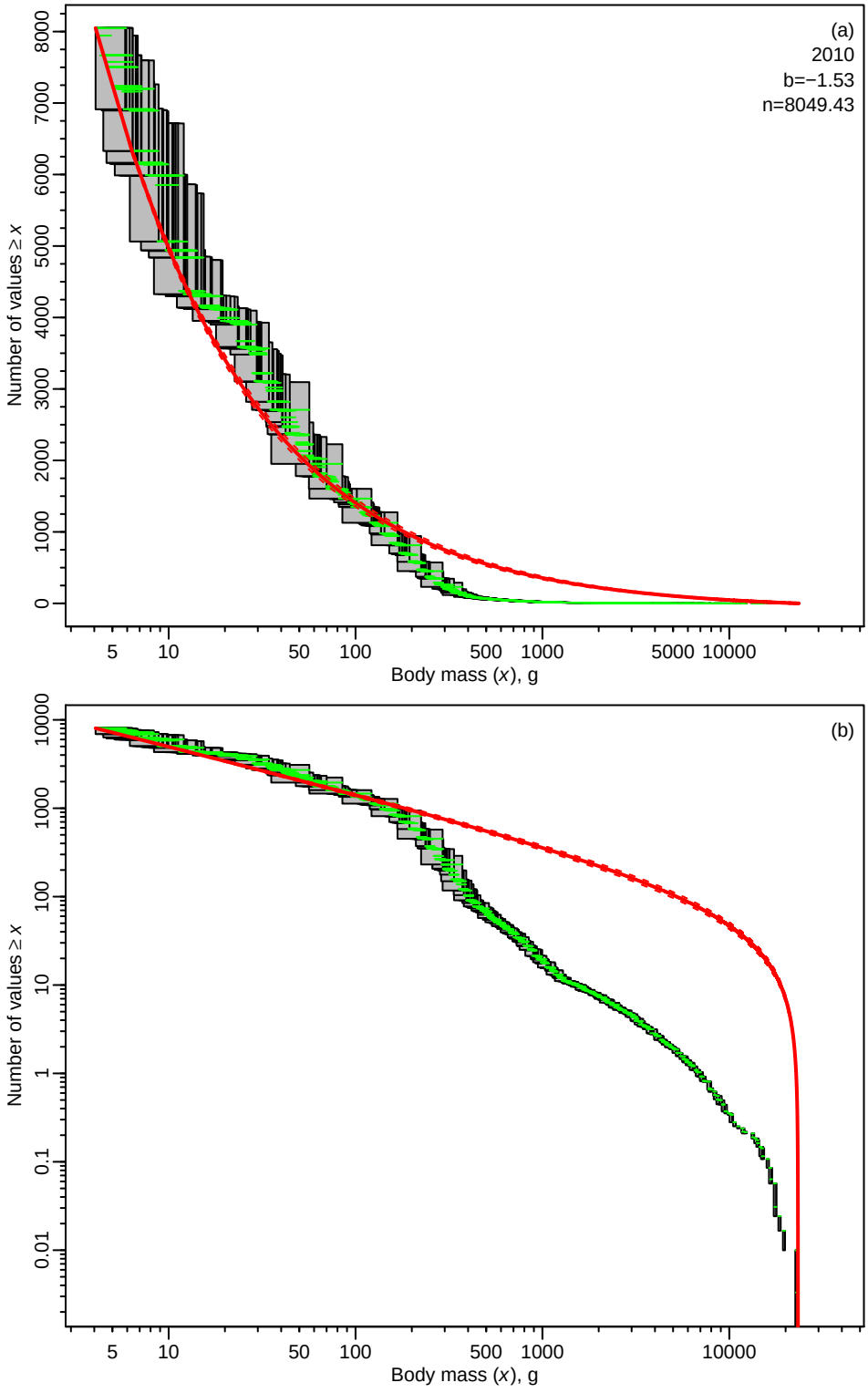

Figure S.29: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 2010. Details as in Figure S.5.

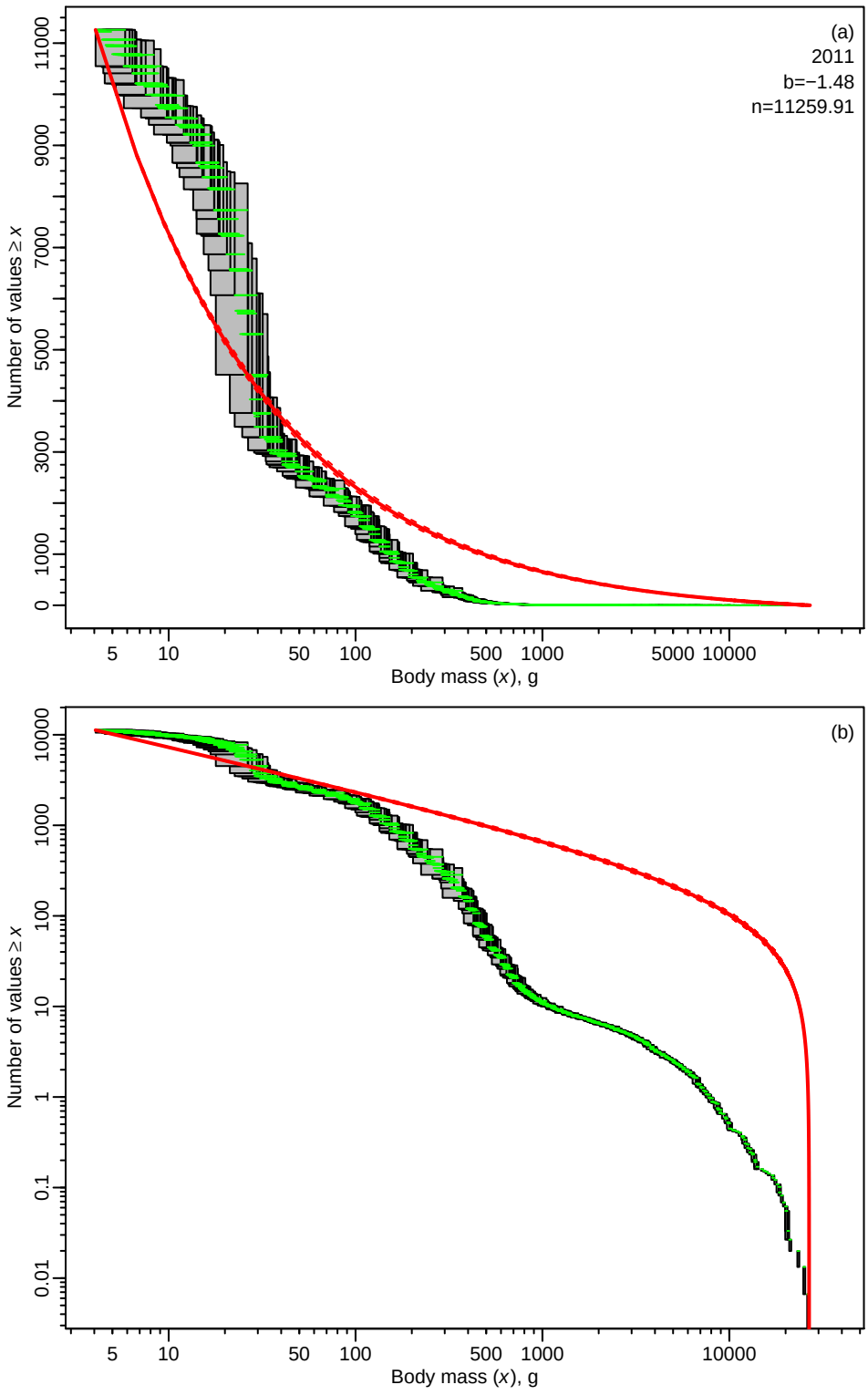

Figure S.30: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 2011. Details as in Figure S.5.

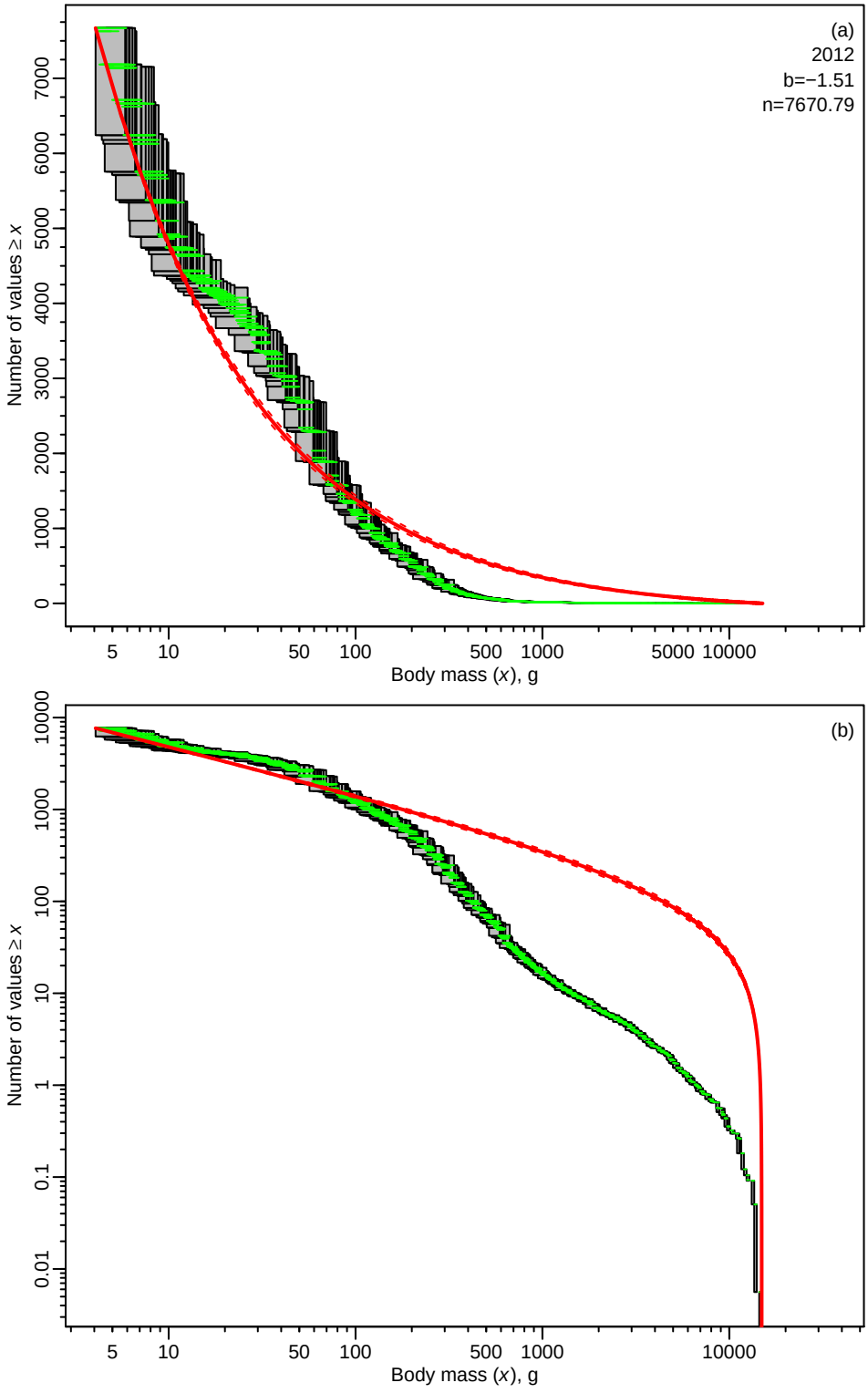

Figure S.31: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 2012. Details as in Figure S.5.

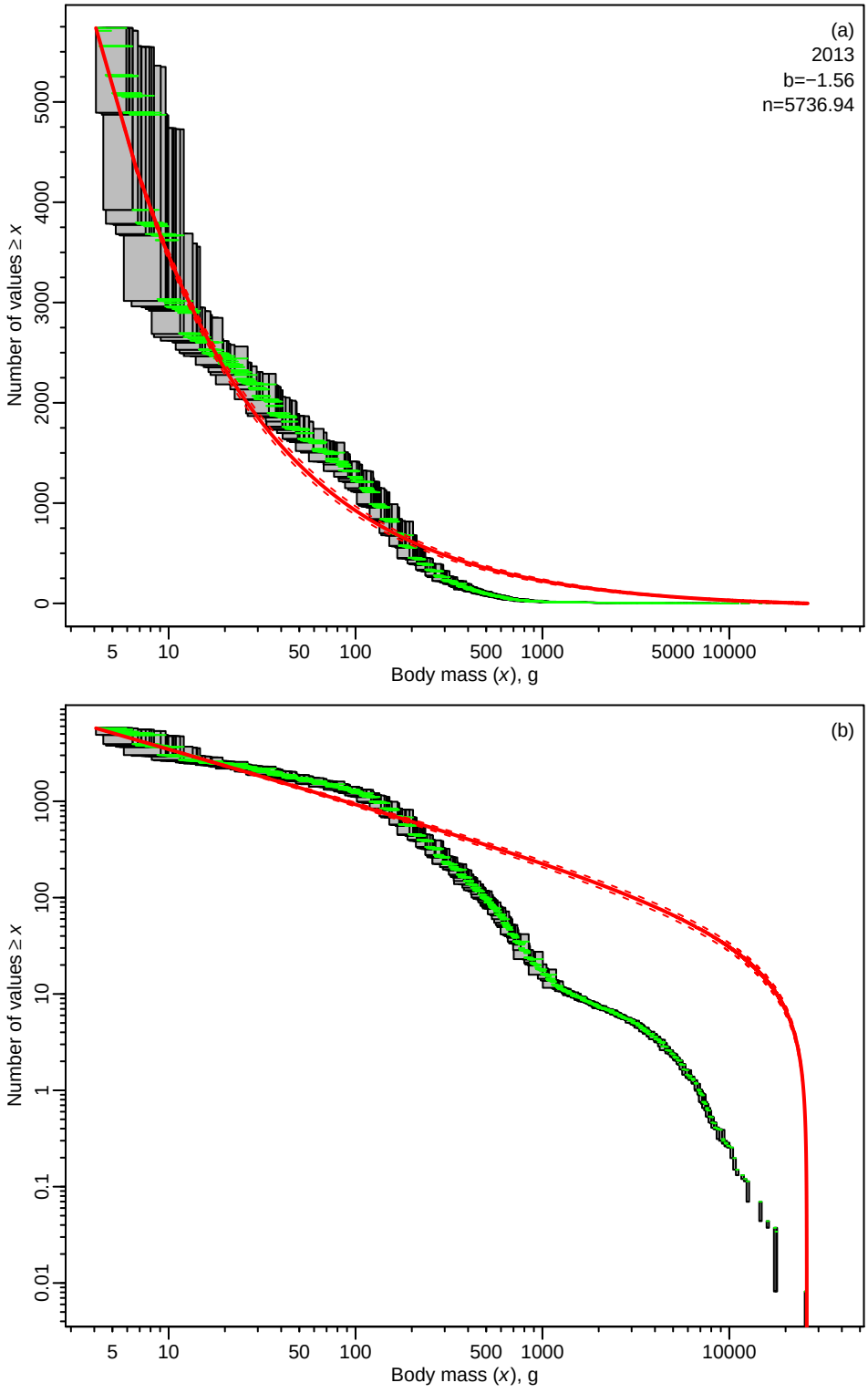

Figure S.32: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 2013. Details as in Figure S.5.

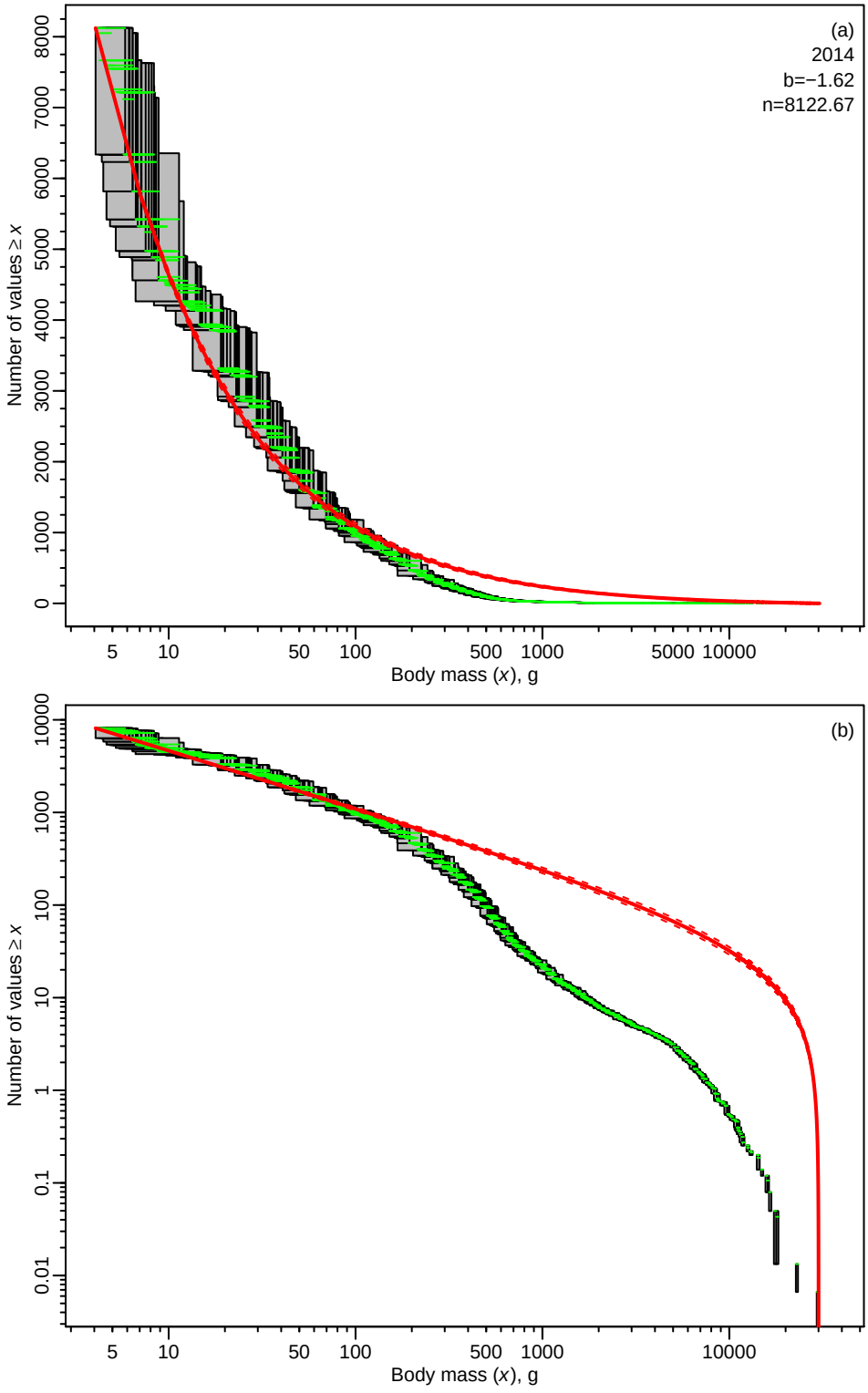

Figure S.33: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 2014. Details as in Figure S.5.

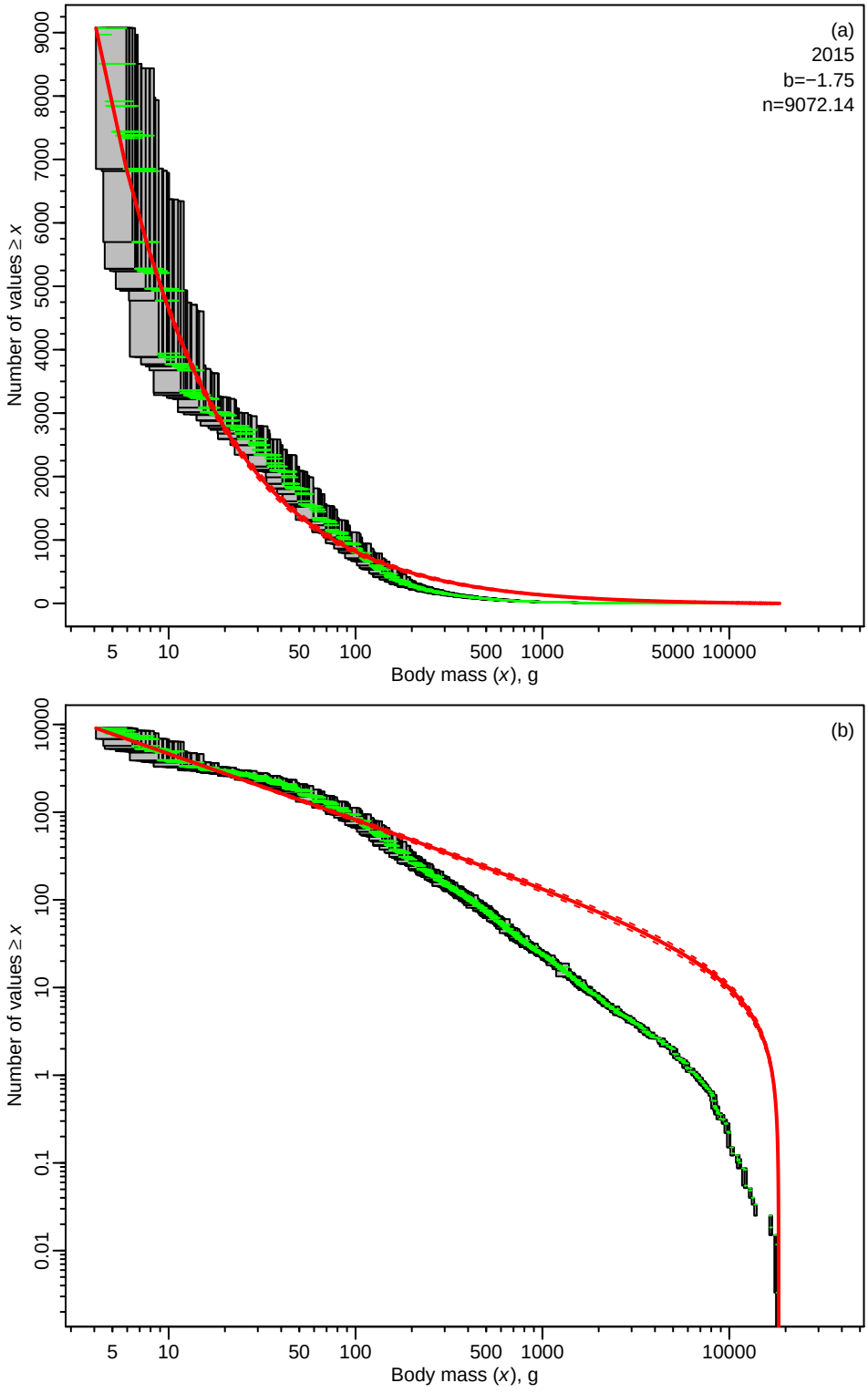

Figure S.34: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 2015. Details as in Figure S.5.

Table S.2: Results for each year of the IBTS data from using the MLEbins method. Both $x_{\min}$ and $x_{\max}$ are calculated as described earlier, n is the total number of individuals per hour, 'Low' and 'High' give the 95% confidence interval for the maximum likelihood estimate (MLE) of b, and C is the normalisation constant from (3) based on the MLE of b.

Year	$x_{\min}$	$x_{\max}$	n	Low	MLE of b	High	C
1986	4.05	28722.51	4799.29	-1.50	-1.49	-1.47	0.97
1987	4.16	25974.17	6202.85	-1.54	-1.53	-1.51	1.12
1988	4.06	29439.75	7711.72	-1.62	-1.60	-1.59	1.41
1989	4.06	35210.99	7667.87	-1.56	-1.55	-1.54	1.20
1990	4.05	34811.19	6399.75	-1.53	-1.52	-1.50	1.07
1991	4.06	26412.52	6639.58	-1.50	-1.49	-1.48	0.99
1992	4.05	25316.66	7973.32	-1.54	-1.53	-1.52	1.12
1993	4.16	29439.75	8628.05	-1.50	-1.49	-1.48	1.00
1994	4.16	22200.90	6282.84	-1.53	-1.52	-1.50	1.10
1995	4.16	31818.64	9401.90	-1.64	-1.62	-1.61	1.52
1996	4.05	36462.12	5607.19	-1.44	-1.42	-1.41	0.78
1997	4.16	19360.99	9113.14	-1.65	-1.63	-1.62	1.58
1998	4.06	22801.53	6927.12	-1.57	-1.55	-1.54	1.22
1999	4.06	36462.12	7199.60	-1.60	-1.58	-1.57	1.33
2000	4.06	22801.53	10816.80	-1.61	-1.60	-1.59	1.41
2001	4.05	18824.87	9058.45	-1.46	-1.45	-1.44	0.85
2002	4.06	20464.76	6766.12	-1.44	-1.42	-1.41	0.79
2003	4.05	21611.31	5614.19	-1.45	-1.44	-1.42	0.82
2004	4.05	30911.23	5445.26	-1.52	-1.50	-1.49	1.02
2005	4.05	19360.99	4189.79	-1.39	-1.38	-1.36	0.67
2006	4.05	21032.63	7864.87	-1.63	-1.62	-1.60	1.48
2007	4.06	18299.10	5278.55	-1.52	-1.50	-1.49	1.03
2008	4.06	28722.51	6516.37	-1.57	-1.56	-1.54	1.23
2009	4.05	24036.35	5628.83	-1.59	-1.58	-1.56	1.30
2010	4.06	23293.68	8049.43	-1.54	-1.53	-1.52	1.13
2011	4.06	26723.46	11259.91	-1.49	-1.48	-1.47	0.94
2012	4.06	14899.37	7670.79	-1.53	-1.51	-1.50	1.07
2013	4.08	25974.17	5736.94	-1.57	-1.56	-1.54	1.23
2014	4.06	30025.12	8122.67	-1.63	-1.62	-1.61	1.48
2015	4.08	18362.51	9072.14	-1.76	-1.75	-1.73	2.14

S.1.9 Further results from fitting simulated data

Table S.3 gives the summary statistics for the simulation results shown in Figure 4. The mean and median estimates of b (that has a true value of -2) for the MLEbin method are all -1.99 or -2.00. But for the MLEmid method the estimates are less accurate and depend upon the bin width.

We repeat those simulations but with $x_{\min} = 16$ rather than $x_{\min} = 1$ to test the sensitivity of the methods. Figures S.35 and S.36 show the equivalent results to Figures 4 and 5, and Table S.4 gives the summary statistics. The MLEmid method behaves better than it did for $x_{\min} = 1$, but not as well as the MLEbin methods that still gives consistently good results.

We also simulate a community with $x_{\min} = 1$, but with the observed data consisting of all body masses $\geq c$, where c is a cutoff value. In practice, c could be known from the sampling protocol or assigned a value based on previous experience. Sufficient random body masses were sampled to obtain a sample size of $n = 1,000$ for the observed data, and we set $c = 16$. The results (Table S.5 and Figures S.37 and S.38) are essentially identical to those for $x_{\min} = 16$. This emphasises the scale-free nature of the power-law distribution, whereby the cutoff value does not affect the estimate of b. Whereas for a scale-dependent distribution, such as the lognormal, the estimated parameters would depend on the cutoff value.

Table S.3: Summary statistics for the 10,000 simulations of 1,000 samples from (2), corresponding to the histograms in Figure 4. Each sample is binned using each of four binning types: 'Linear 1', 'Linear 5' and 'Linear 10' for bins of constant bin widths 1, 5, and 10, respectively, and '2k' for bin widths that double in size. Each binned sample is then fit using the MLEmid and MLEbin methods, where the MLEmid method uses the midpoint of each bin and the MLEbin method (shaded rows) more explicitly accounts for the binning in the likelihood function. Statistics relate to the resulting 10,000 estimates of b, with the final column giving the percentage of simulations for which the estimate is below the true value of $b = -2$.

Binning type	Method	5% quantile	Median	Mean	95% quantile	Percentage below true
Linear 1	MLEmid	-1.98	-1.94	-1.94	-1.89	1
	MLEbin	-2.05	-1.99	-1.99	-1.94	43
Linear 5	MLEmid	-1.63	-1.61	-1.60	-1.57	0
	MLEbin	-2.06	-1.99	-1.99	-1.94	42
Linear 10	MLEmid	-1.43	-1.41	-1.40	-1.35	0
	MLEbin	-2.06	-1.99	-1.99	-1.93	42
2k	MLEmid	-1.95	-1.90	-1.90	-1.86	0
	MLEbin	-2.05	-2.00	-2.00	-1.94	47

Table S.4: As for Table S.3 but with $x_{\min} = 16$, corresponding to the histograms in Figure S.35.

Binning type	Method	5% quantile	Median	Mean	95% quantile	Percentage below true
Linear 1	MLEmid	-2.06	-1.99	-2.00	-1.94	44
	MLEbin	-2.06	-1.99	-2.00	-1.94	45
Linear 5	MLEmid	-2.05	-1.99	-1.99	-1.93	36
	MLEbin	-2.06	-1.99	-2.00	-1.94	45
Linear 10	MLEmid	-2.02	-1.96	-1.96	-1.90	14
	MLEbin	-2.06	-2.00	-2.00	-1.94	45
2k	MLEmid	-1.93	-1.87	-1.87	-1.82	0
	MLEbin	-2.07	-2.00	-2.00	-1.94	51

Table S.5: As for Table S.3 with $x_{\min} = 1$, but only sampling data above the cutoff value (i.e. $\geq c = 16$), corresponding to the histograms in Figure S.37. Each simulated data set had a sample size of 1,000 above the cutoff value. Results are identical to those in Table S.4 except for a few differences of 0.01 in the statistical values and 1% in the final column.

Binning type	Method	5% quantile	Median	Mean	95% quantile	Percentage below true
Linear 1	MLEmid	-2.06	-1.99	-2.00	-1.93	45
	MLEbin	-2.06	-1.99	-2.00	-1.94	45
Linear 5	MLEmid	-2.05	-1.99	-1.99	-1.93	36
	MLEbin	-2.06	-2.00	-2.00	-1.94	45
Linear 10	MLEmid	-2.02	-1.96	-1.96	-1.90	13
	MLEbin	-2.06	-2.00	-2.00	-1.93	46
2k	MLEmid	-1.93	-1.87	-1.87	-1.82	0
	MLEbin	-2.07	-2.00	-2.00	-1.94	51

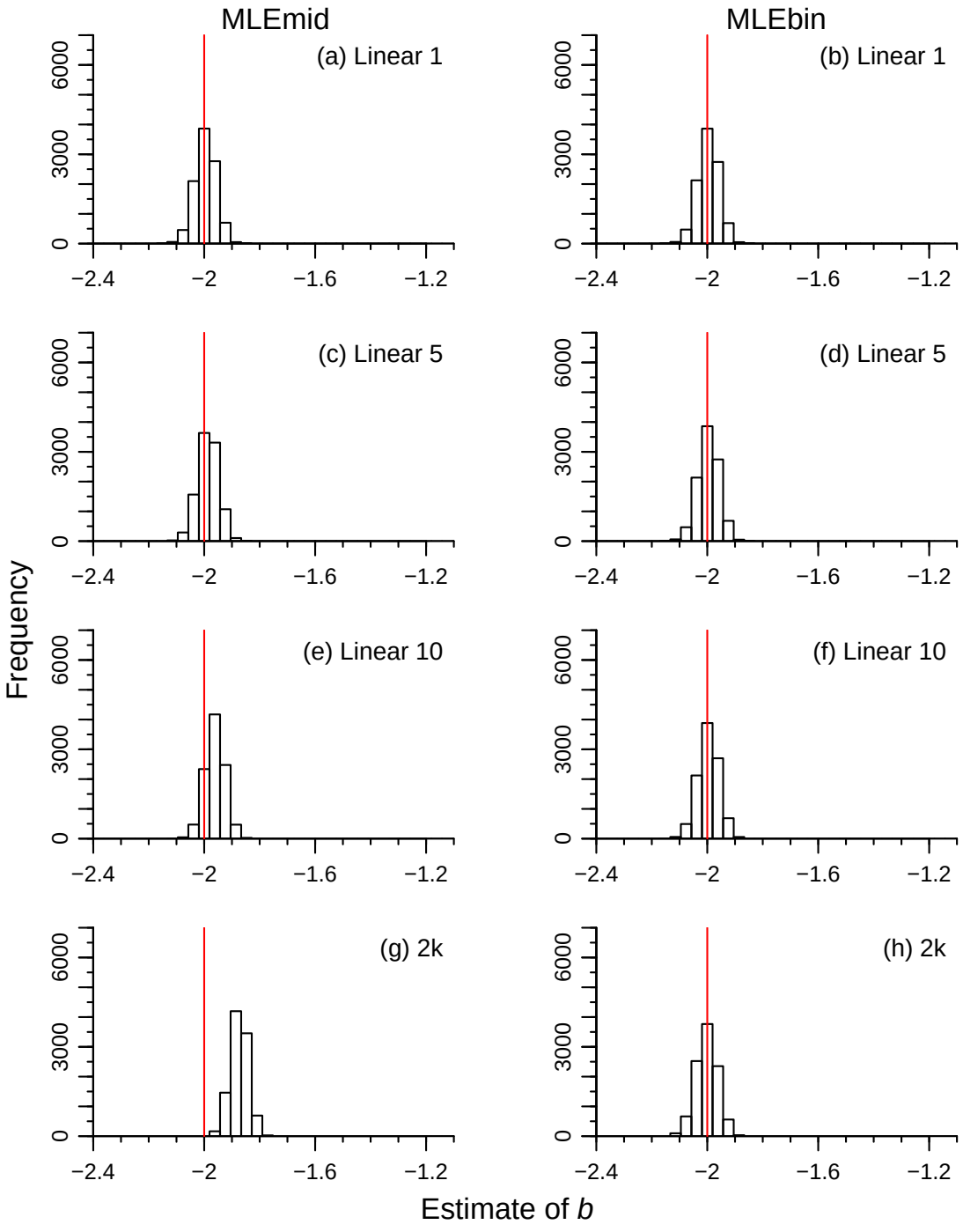

Figure S.35: As for Figure 4 but with $x_{\min} = 16$.

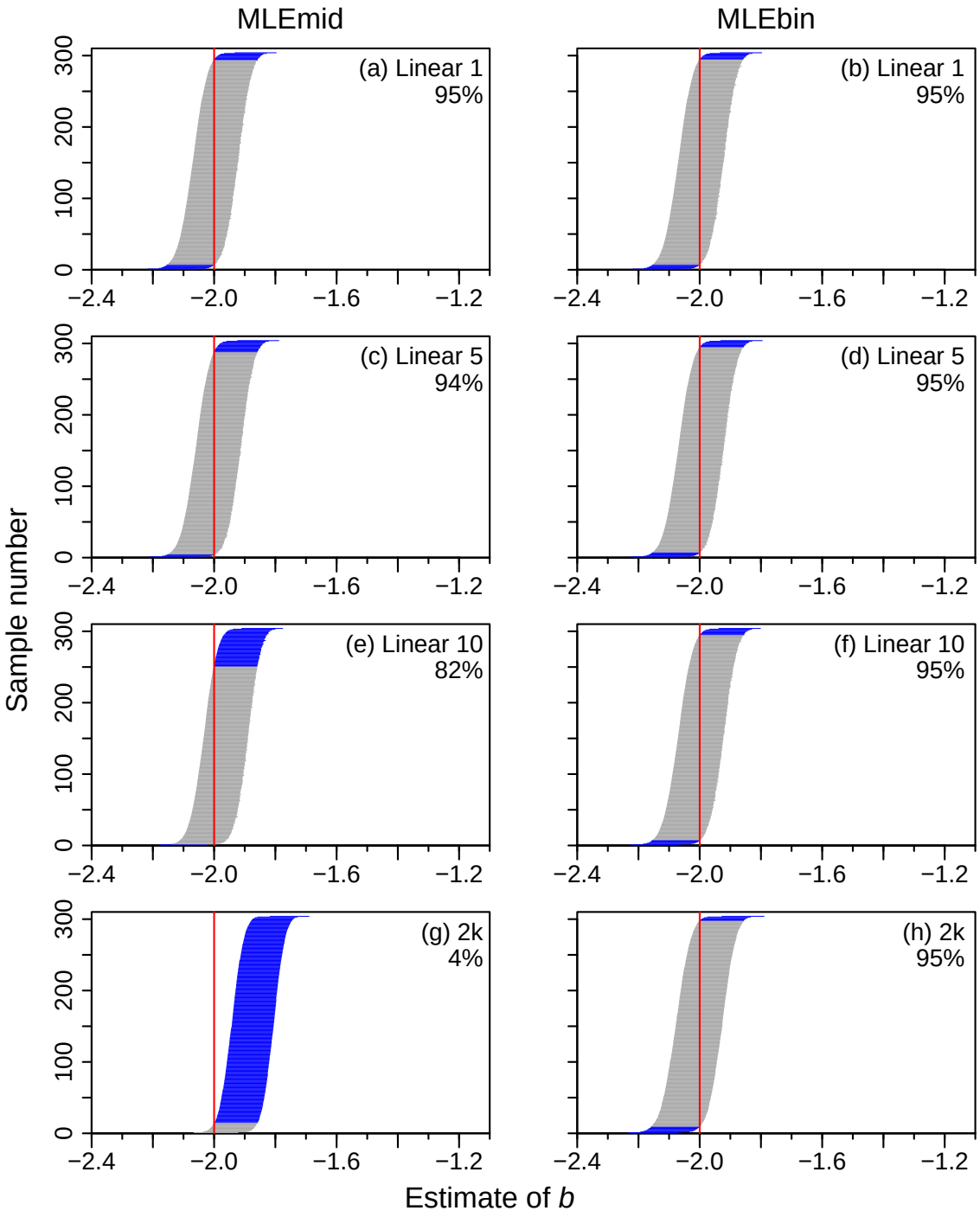

Figure S.36: As for Figure 5 but with $x_{\min} = 16$.

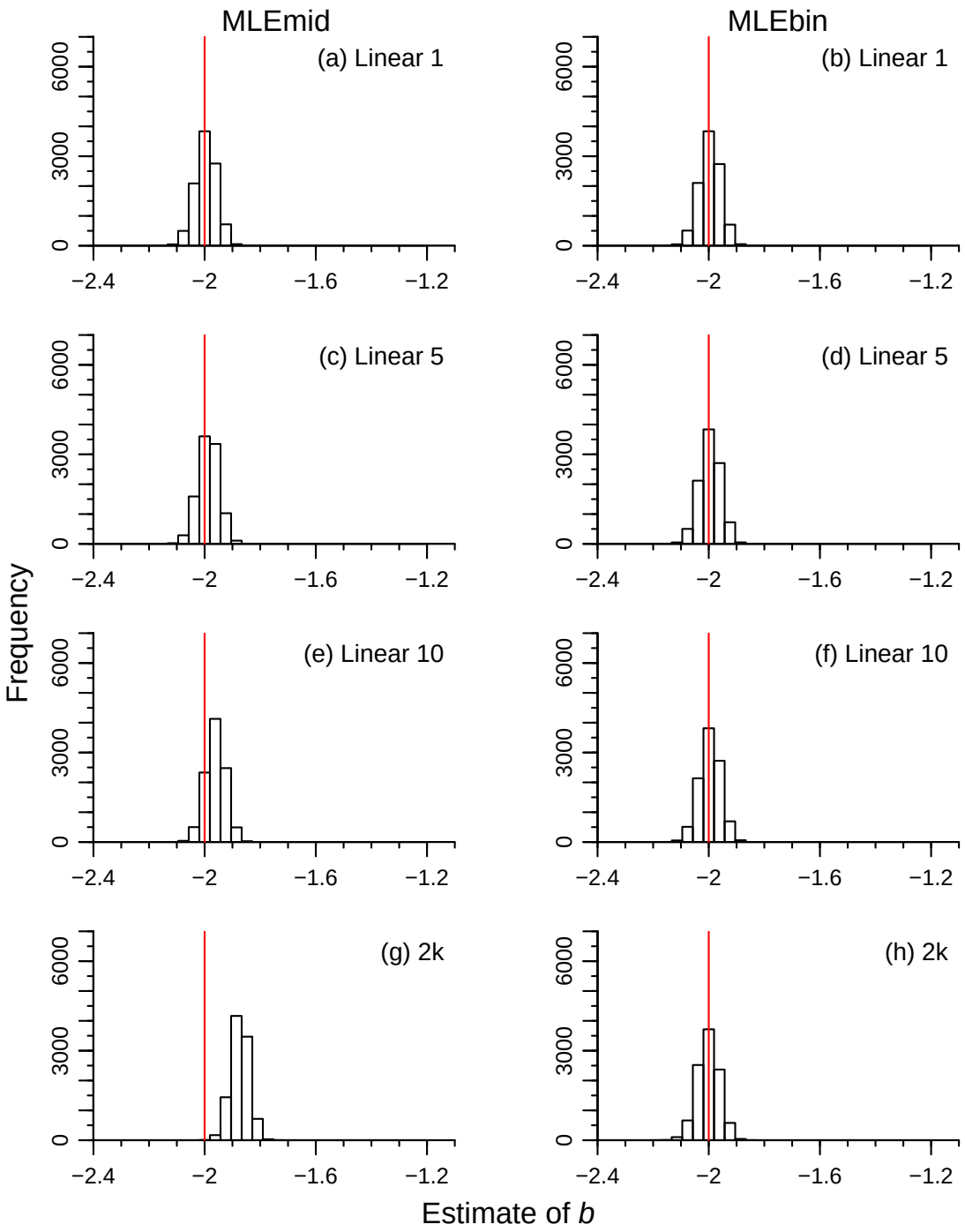

Figure S.37: As for Figure 4 but only observing data above a cutoff of 16. There are only very minor differences to Figure S.35, as confirmed in Table S.5.

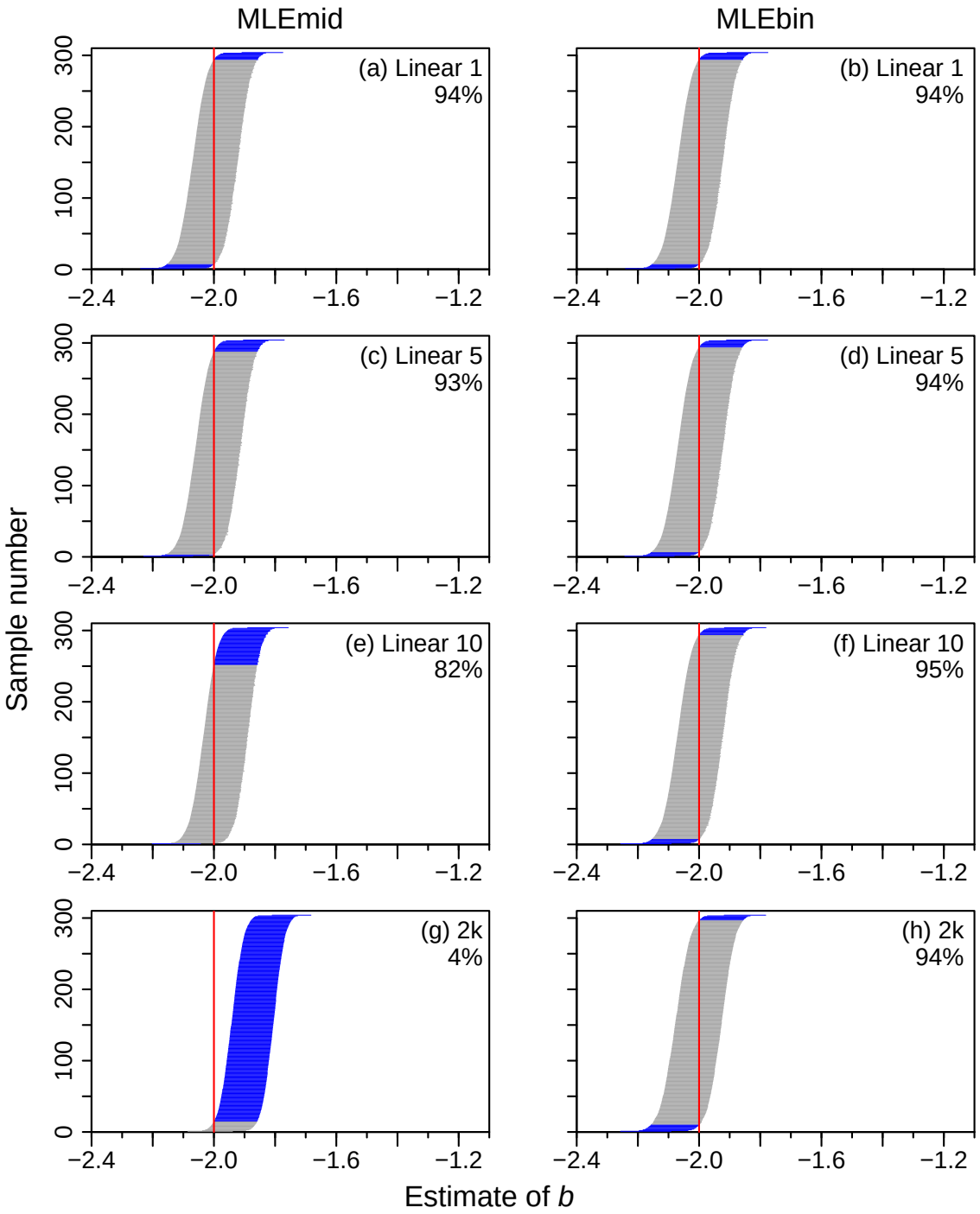

Figure S.38: As for Figure 5 but only observing data $\geq$ the cutoff value of 16. There are only very minor differences to Figure S.36, with five of the observed coverage values changing by 1%.

S.1.10 Further results from changing the minimum cutoff for the IBTS data

Following Piet and Jennings (2005) and Blanchard *et al.* (2005) we originally removed all body masses < 4 g from the IBTS data before analysing the data. Figure 7 for 1999 showed a good fit of the PLB model to small fish (that contribute most of the counts) but not to larger fish above $\sim$100 g. We suggested that such an effect could be a consequence of fishing pressure. An alternative hypothesis (suggested by a reviewer) is that the small fish are poorly sampled due to characteristics of the fishing gear, yet still dominate estimation of b because they are so numerous. Fitting just the larger fish may give a good fit of the PLB distribution that may not imply an effect on community structure due to fishing. Therefore, here we re-run the analyses using only fish ≥ 100 g. This is much larger than the original 4 g to help reveal any consequences of the choice of cutoff value.

In Figure S.39 we show the resulting fitted ISD for 1999, only considering fish ≥ 100 g. Compared to Figure 7 for the full data set: (i) the sample size is reduced from 7,200 to 892, due to the elimination of the numerous small fish, (ii) the fitted value of b changes from -1.58 to -2.51, (iii) the confidence intervals of b appear larger (due to the smaller sample size), (iv) the distribution appears to fit fairly well across most of the data, except in roughly the middle range 400-3,000 g (and a few larger fish). The large change in b emphasises the need to use a consistent cutoff value between years (else b can change simply due to changing the cutoff value). So for 1999 it appears that the few large fish are indeed conforming to the PLB distribution. Figures (and code) for all years are shown as a movie in the `sizeSpectra` package's vignette `MEPS_IBTS_all_min_100`. The pattern shown in Figure S.39 is not universal – for some years the fitted curve underestimates the numbers of larger fish (above 2,000 g).

Figure S.40 shows the temporal trend in estimated b, as in Figures 1 and 8 but with the cutoff of 100 g. Compared to the MLEbins method for the full data (Figure 8), Figure S.40 suggests an increase in b over time and thus a shallowing of the size spectrum (rather than no change). Also,

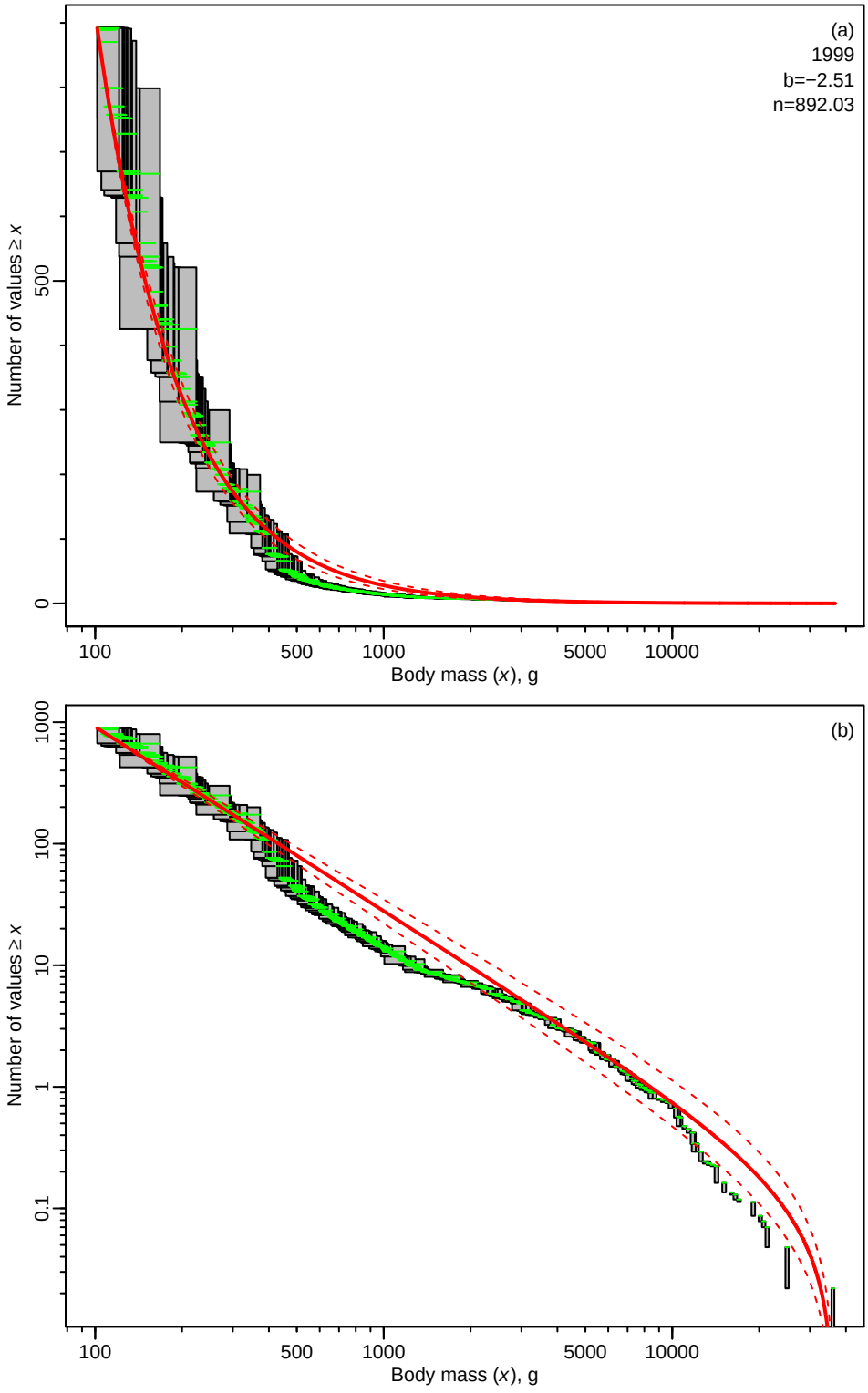

Figure S.39: Individual size distribution and MLEbins fit with 95% confidence intervals for IBTS data in 1999 considering only body masses > 100 g. Details as in Figure 7.

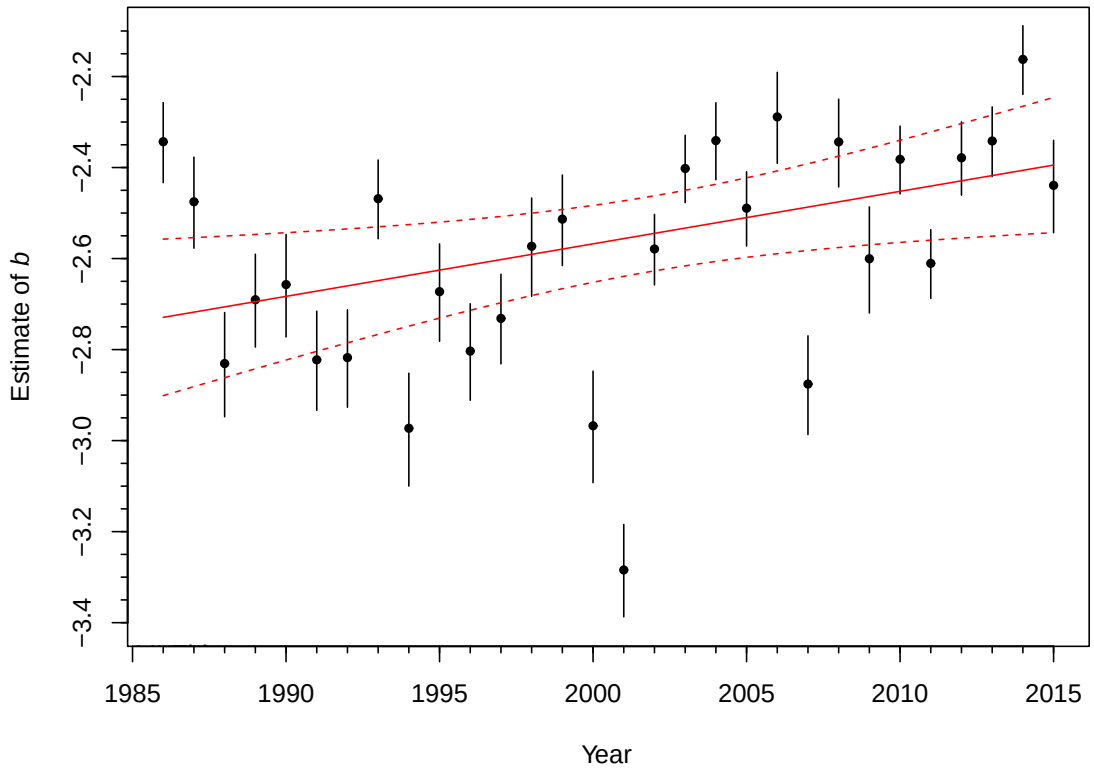

Figure S.40: Annual estimates of b for the IBTS data considering only body masses > 100 g. Details as in Figure 1, with the red lines showing a statistically significant increase in the estimate of b (the equivalent statistics to the Table S.1 results are: Low$= 0.0021$, Trend$= 0.0115$, High$= 0.0210$, $p = 0.02$, $R^2 = 0.18$).

for example, 2001 had the fourth highest estimate of b in Figure 8, but has the lowest estimate (by some way) in Figure S.40.

The important qualitative differences (no change in b compared to an increase) that occur when using alternative cutoff values shows the importance of carefully selecting the cutoff value when using size spectra to understand ecosystems and providing advice to managers. However, the cutoff value should be chosen based on careful consideration of the size-selectivity characteristics of the fishing gear rather than examination of results, would likely lie somewhere in the range 4-100 g for these data, and needs to be the same for all years.

S.1.11 Histograms using different binning types

Figures S.41-S.45 show histograms of the same set of 1,000 random numbers sampled from the PLB distribution (with $b = -2$, $x_{\min} = 1$ and $x_{\max} = 1,000$), plotted using the different binning types described in the main text. These illustrate the highly skewed nature of the PLB distribution, and how the first bin can contain most of the counts.

S.1.12 Literature cited only in Supplementary Material

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd ed., Springer, New York.

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

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

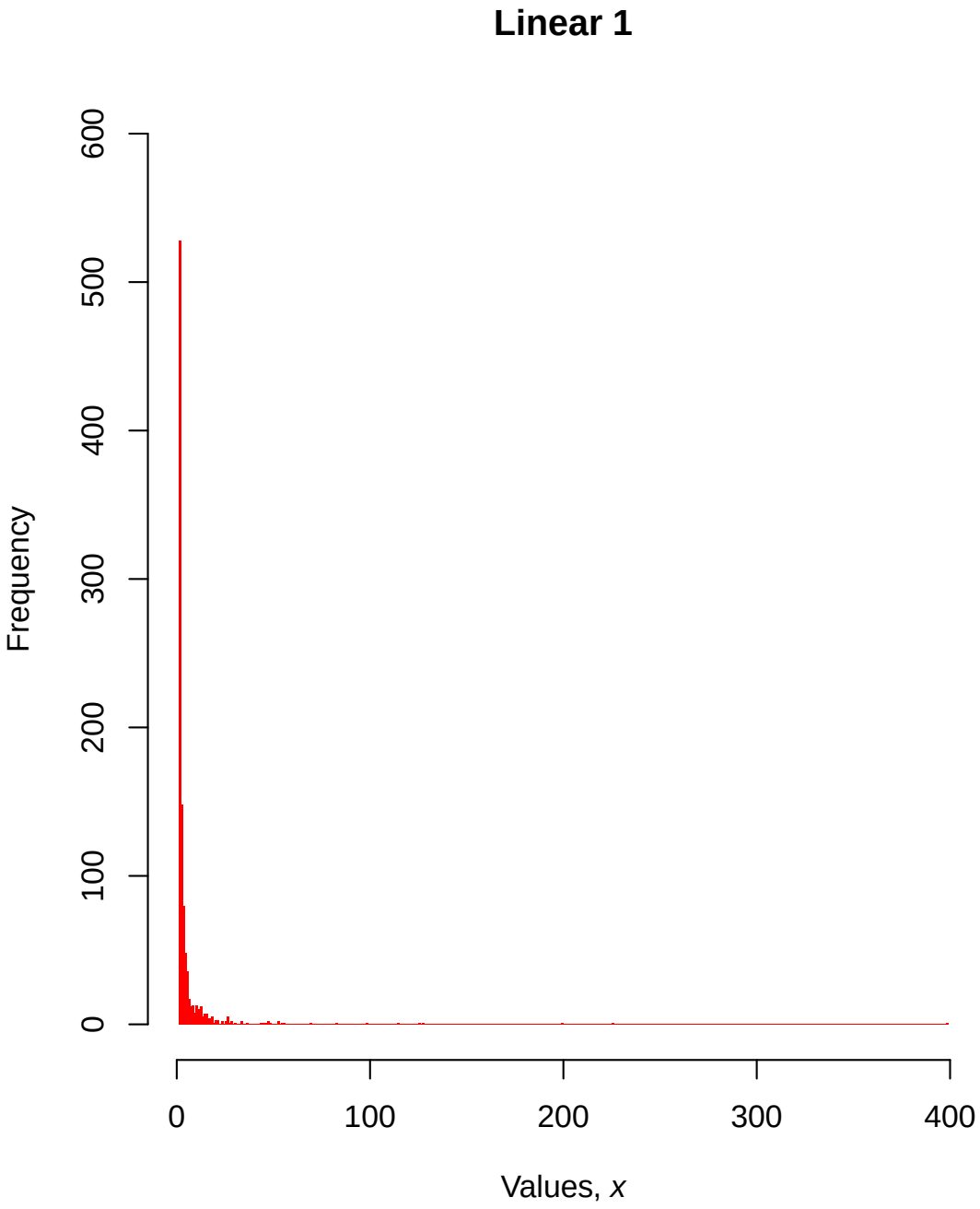

Figure S.41: Histogram of the 1,000 random numbers with bin widths of 1.

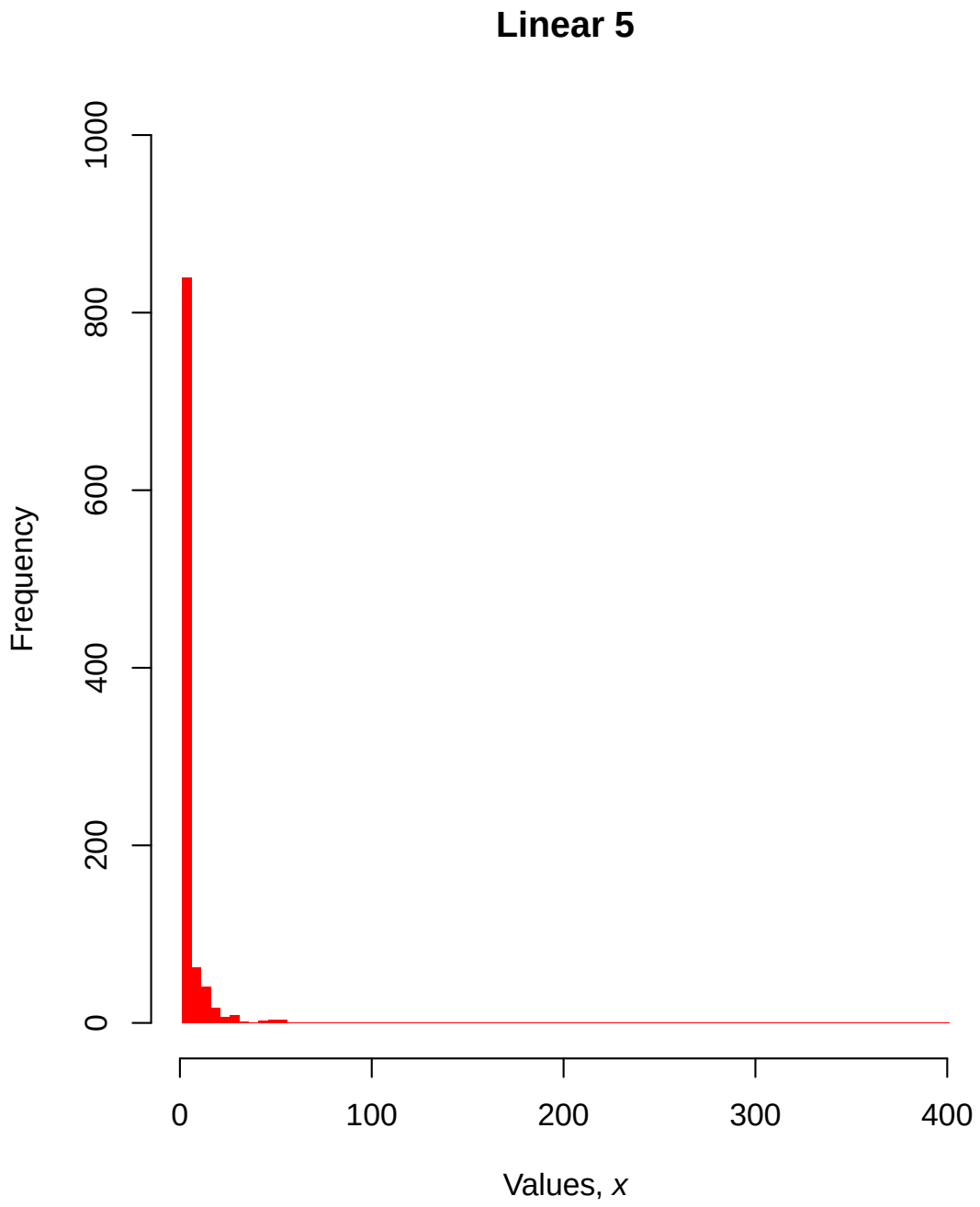

Figure S.42: Histogram of the 1,000 random numbers with bin widths of 5.

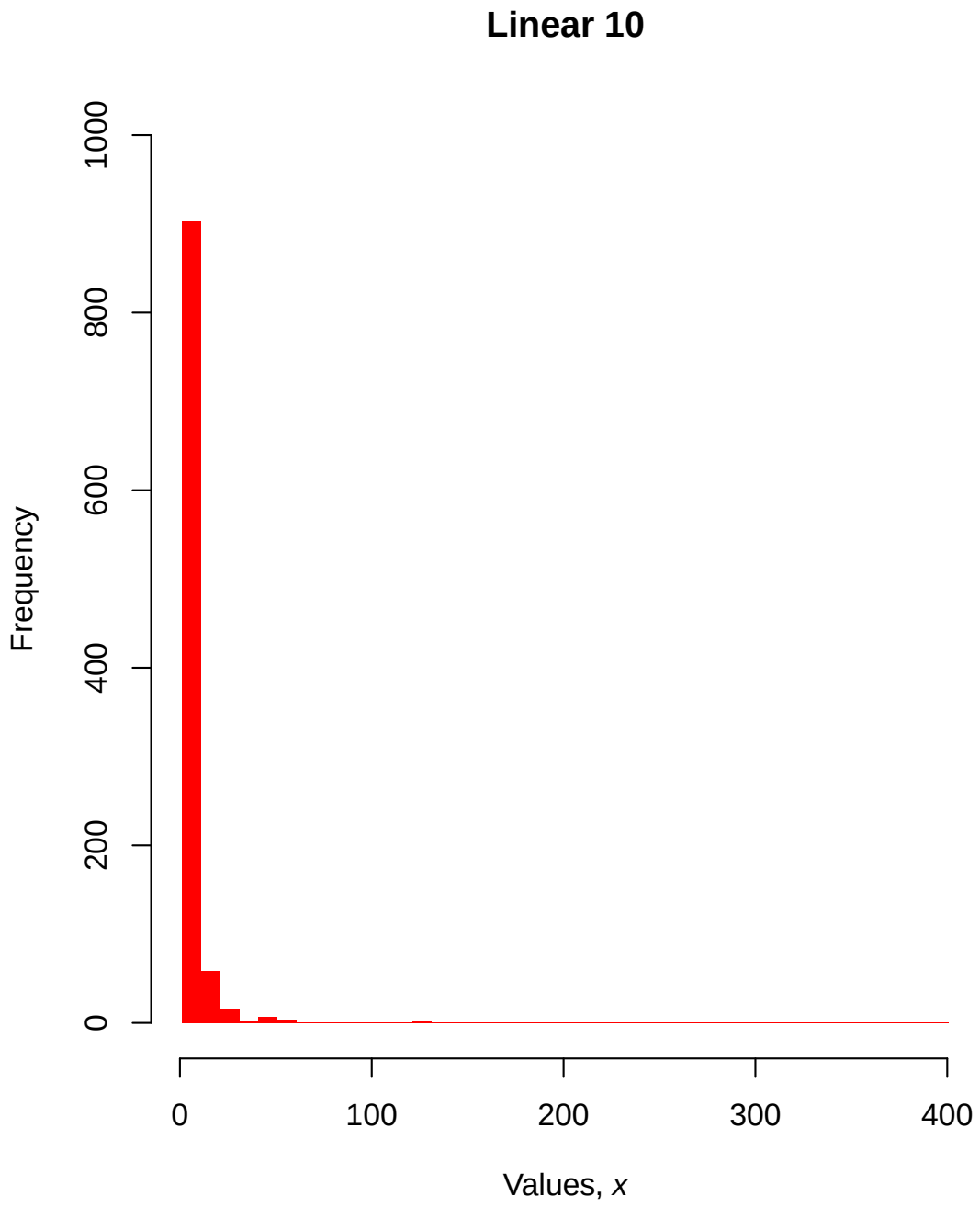

Figure S.43: Histogram of the 1,000 random numbers with bin widths of 10.

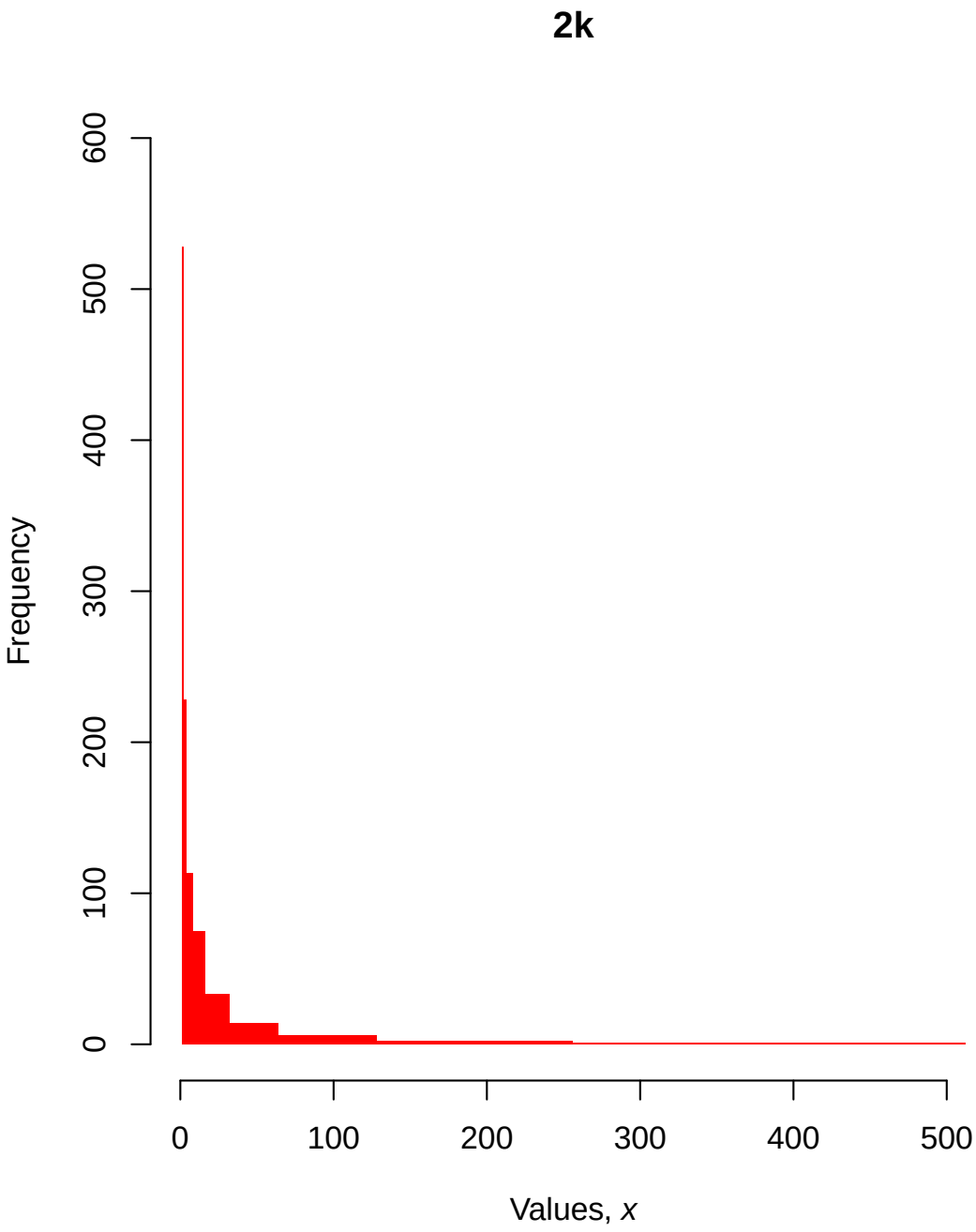

Figure S.44: Histogram of the 1,000 random numbers with bin widths that double in size. Note that heights of bars represent the counts in each bin, and that because of the non-constant bin widths the areas are not proportional to the counts – see Figure S.45.

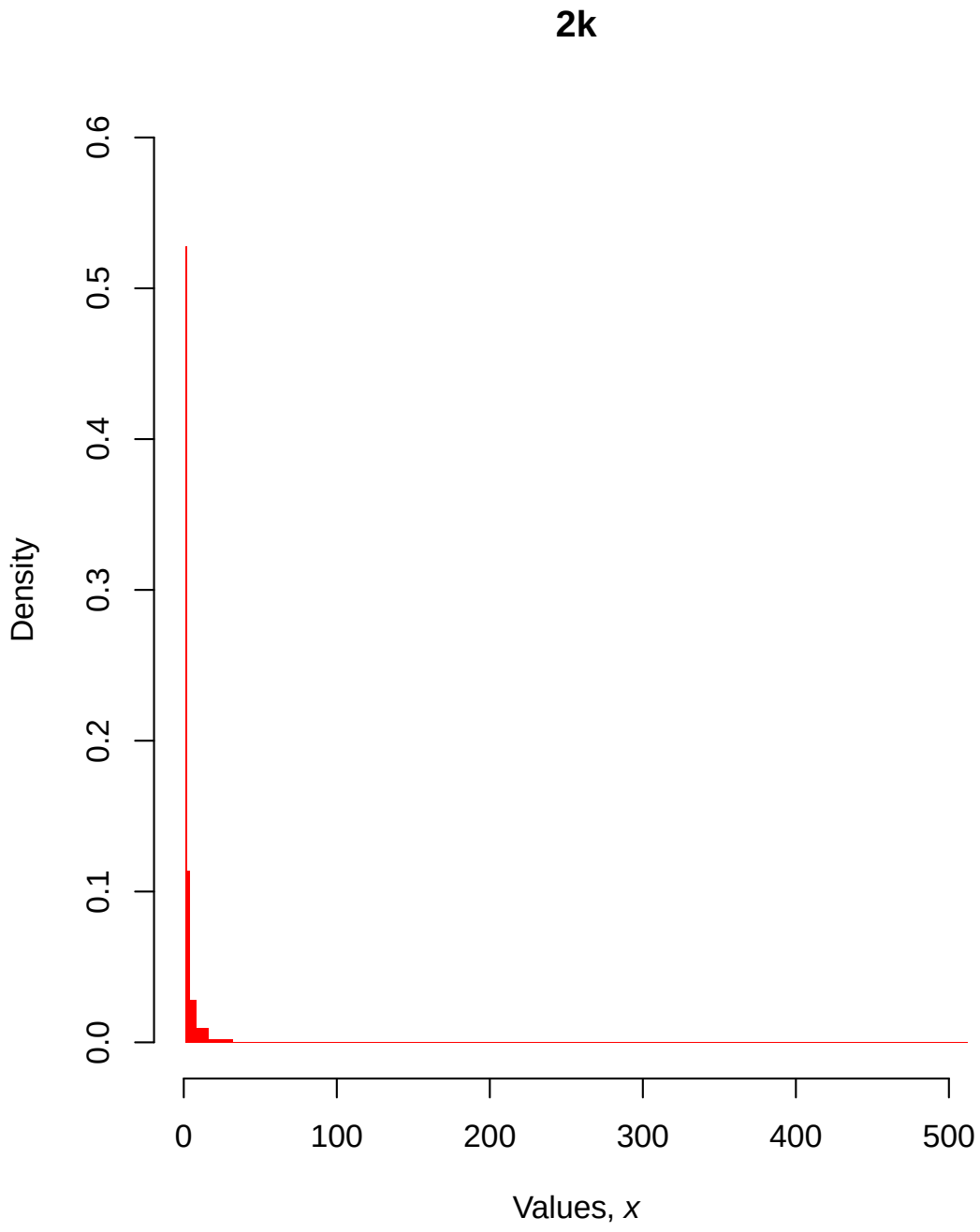

Figure S.45: Density plot of the 1,000 random numbers with bin widths that double in size, showing density rather than frequency.

S.2 Summary of R package `sizeSpectra`

For full reproducibility, all R code has been functionalised and documented in our new R package `sizeSpectra`, which is is freely available at `https://github.com/andrew-edwards/sizeSpectra`. This enables reproduction of all our results, tables and figures for the current work plus those in Edwards *et al.* (2017).

Version 1.0.0.0 of the package is also archived here in the submitted file `sizeSpectra_1.0.0.0.tar.gz`, and can be installed in R by running

```
install.packages("sizeSpectra_1.0.0.0.tar.gz", repos=NULL)
```

To get started, see the vignettes at `https://github.com/andrew-edwards/sizeSpectra` or directly in R with the commands:

```
library(sizeSpectra)
browseVignettes("sizeSpectra")
```

Please post any issues or questions at `https://github.com/andrew-edwards/sizeSpectra`. The code will updated if necessary on the GitHub site, so check there for any updates.