A nonparametric method for the measurement of size diversity with emphasis on data standardization

Xavier D. Quintana[1]*, Sandra Brucet[1,2], Dani Boix[1], Rocío López-Flores[1], Stéphanie Gascón[1], Anna Badosa[1], Jordi Sala[1], Ramon Moreno-Amich[1], and Juan J. Egozcue[3]

[1]Institute of Aquatic Ecology and Department of Environmental Sciences, Universitat de Girona, Girona, Spain
[2]Department of Freshwater Ecology, National Environmental Research Institute, Vejlsøvej 25, DK-8600 Silkeborg, Denmark
[3]Department of Applied Mathematics III, Universitat Politècnica de Catalunya, Barcelona, Spain

Abstract

The most suitable method for estimation of size diversity is investigated. Size diversity is computed on the basis of the Shannon diversity expression adapted for continuous variables, such as size. It takes the form of an integral involving the probability density function (pdf) of the size of the individuals. Different approaches for the estimation of pdf are compared: parametric methods, assuming that data come from a determinate family of pdfs, and nonparametric methods, where pdf is estimated using some kind of local evaluation. Exponential, generalized Pareto, normal, and log-normal distributions have been used to generate simulated samples using estimated parameters from real samples. Nonparametric methods include discrete computation of data histograms based on size intervals and continuous kernel estimation of pdf. Kernel approach gives accurate estimation of size diversity, whilst parametric methods are only useful when the reference distribution have similar shape to the real one. Special attention is given for data standardization. The division of data by the sample geometric mean is proposed as the most suitable standardization method, which shows additional advantages: the same size diversity value is obtained when using original size or log-transformed data, and size measurements with different dimensionality (longitudes, areas, volumes or biomasses) may be immediately compared with the simple addition of ln k where k is the dimensionality (1, 2, or 3, respectively). Thus, the kernel estimation, after data standardization by division of sample geometric mean, arises as the most reliable and generalizable method of size diversity evaluation.

Several approaches of ataxonomic diversity have been classically used in ecology as an alternative to species diversity and provide additional information about the community structure (Margalef 1991; Han 1998). Ataxonomic diversity can be calculated by classifying organisms by their biomass or size (Oindo et al. 2001), their bio-optical properties (Li 1997), their trophic level (Pianka 1974; Aoki and Mizushima 2001), or their ecological strategies (Weithoff et al. 2001), what has been defined as functional diversity (Hulot et al. 2000; Hooper et al. 2002; Petchey and Gaston 2002). Shannon-Wiener index (Shannon and Weaver 1949; Pielou 1969) is the most widely used diversity parameter for community studies. Body size is one of the most important attributes of an organism from an ecological and evolutionary point of view (Werner and Gilliam 1984). Its importance has been widely stressed (Strayer 1991; Rodríguez and Li 1994; Kerr and Dickie 2001; Marquet et al. 2005), and there are many works that describe patterns in the size structure and its ecological relevance, especially in pelagic marine ecosystems (e.g., Platt and Denman 1977; Platt et al. 1984; Rodríguez and Mullin 1986), but also in marine benthic communities (e.g., Schwinghamer 1981; Duplisea 2000) and in continental waterbodies (e.g., Mittelbach 1981; Gaedke 1992; Rasmussen 1993; Cohen et al. 2003; Basset et al. 2004). In most cases, these works analyze the shape of the biomass size distribution in a log-log plot and its fit to a determinate function, whether a negative slope straight line or other nonlinear function. Adjustments of size distribution to nonlinear functions in the log-log plot usually give better fit to size distribution data (Gasol et al. 1991; Vidondo et al. 1997; Brucet et al. 2005), but the ecological meaning of the parameters of such models of distribution is normally difficult to interpret.

Several investigators have used size diversity to describe the shape of size distributions (Parsons 1969; Piontkovski and van der Spoel 1998; Quiroga et al. 2005; Brucet et al. 2006; Ichinokawa and Takahashi 2006). A size diversity measure gives a unique value per size distribution and has the advantage of a more intuitive interpretation of its ecological meaning, since the concept of diversity is well established. However, as Ruiz (1994) pointed out, a methodological problem emerges when measuring size diversity by clustering the different body sizes into size classes, because the division of a continuum variable (size) into an arbitrarily selected number of size classes is needed. As a result, different size diversity values for the same community can be obtained depending on the number of size intervals chosen. In order to overcome these problems, Lurie and Wagensberg (1983, 1984) proposed a measure of size diversity index suitable for continuous variables, parallel to Shannon entropy used in information theory or species diversity (e.g., Good 1953). This index is based on the probability density function (pdf) of the size of individuals, and it takes an integral form, better than the discrete summation used in traditional diversity computations.

However, the difficulties of describing natural size distributions by means of a simple pdf are evident. Several authors have modeled the relationship between size and abundance as a consequence of allometry in physiological processes and in competitive or predator-prey interactions (Platt and Denman 1977; Dickie et al. 1987; Thiebaux and Dickie 1993; Quiñones 1994; Han and Straskraba 1998), but discontinuities in this parametric relationship are frequent and of great ecological importance (Rodríguez 1994; Havlicek and Carpenter 2001). Vidondo et al. (1997) also proposed the use of the Pareto distribution arguing that most distributions in the nature follow this distribution, and this Pareto distribution has been used for size diversity measurements (Brucet et al. 2006). However, there is no reason to assume that the pdf of an actual size distribution always have to fit to a determinate parametric model. As a consequence of these difficulties with the parametric approach, some authors have proposed the use of a nonparametric kernel estimation of the pdf, especially when they investigate the existence of lumps or gaps in the size spectrum (Havlicek and Carpenter 2001; Ruiz et al. 2002). Nonparametric approaches, such as the kernel estimation, have the advantage that a whole functional expression of pdf is not required, thus, being applicable for most size distributions.

The present goal is two fold. The first one is to propose a suitable way to estimate size diversity adapted to a broad class of pdfs. With the aim to obtain a general method for estimation of size diversity, several parametric and nonparametric approaches were tested using simulated samples. Their fit to different size distributions found in the nature was also analyzed. The second objective is to define a normalization to make the index of different samples and communities as comparable as possible. This is attained by a double standardization: a first one to make the size-data adimensional, and a second one to make size data in different dimensions (length, weight, volume, etc.) comparable.

Materials and procedures

Shannon size diversity index—Let X be a random variable representing the size of individuals, with $p_X(x)$ representing its pdf. A sample from X is available and denoted by x_1, x_2, ..., x_n. The goal is to estimate the Shannon diversity index corresponding to $p_X(x)$ from the available sample. Shannon entropy is to be used as a diversity index of sizes (Good 1953; Lurie and Wagensberg 1983, 1984), and it is defined as

$$\mu_2(X) = -\int_0^{+\infty} p_X(x)\log_2 p_X(x)\,dx \qquad (1)$$

where the limits of the integral are due to the positive character of the size X. From the mathematical point of view, working with logarithms in base 2 is uncomfortable, and natural logarithms will be used instead in Eq. 1 to obtain the expression of diversity using natural logarithms $\mu(X) = \ln 2\, \mu_2(X)$.

There are several alternative methods to compute $\mu(X)$. They can be classified into two groups: (A) Methods assuming data coming from a parametric family of pdfs, $p_X(x|\theta)$, where θ denotes the parameters. These methods proceed to estimate θ and, then, they obtain $\mu(X)$ corresponding to such a family computing (Eq. 1), either numerically or analytically. Some probability distributions have known diversity expressions (e.g., Johnson 2004) and, in practice, some of these probability distributions fit well size data (Vidondo et al. 1997). The resulting expressions of $\mu(X)$ for each distribution are given in Web Appendix A. (B) Methods in which no parametric hypothesis on $p_X(x)$ is stated and the integral (Eq. 1) is estimated using some kind of local evaluation of $p_X(x)$. They are closely related to the density estimation methods, from the simple but effective evaluation of a histogram, to more elaborated kernel-based ones (e.g., Silverman 1986).

Methods in Group A seem appropriate whenever there is some evidence of an adequate parametric family $p_X(x|\theta)$, both fitting the data and modeling the phenomenon. If these circumstances are not present, the statistical uncertainty in the estimation of θ may be increased by the modeling uncertainty. Methods in Group B are more flexible for large or small samples, but they are not able to take advantage of prior information on the adequate model.

Although the allometric relationship between body size and abundance is well-known, there is no strong reason to assume a parametric family of distributions for sizes. Some candidates, such as the Pareto distribution, have been used and seem to fit data suitably (see examples in Vidondo et al. 1997). However, there is no theoretical reason to accept the model. In fact, Pareto distribution has an unbounded support in the positive real line, but the measured sizes are physically bounded both by the measuring device and the nature of the individuals. Because of the lack of evidence in favor of a parametric density model, the present study is centered on nonparametric methods of Group B.

Whenever sizes are assumed positive, the diversity integral in Eq. 1 can be expressed in a logarithmic scale, and this is often convenient both for numeric and analytic computations. Defining $y = \ln x$, the diversity integral is

$$\mu(X) = -\int_{0}^{+\infty} p_X(x) \ln p_X(x)\, dx$$
$$= -\int_{-\infty}^{+\infty} \exp(y)\, p_X(\exp(y)) \ln p_X(\exp(y))\, dy \quad (2)$$

where, in both expressions $p_X(x)$ is the pdf of the size X. From now on, both integrals in Eq. 2 are taken as a definition of $\mu(X)$.

A way to summarize Eq. 2 consists in defining the log-transformed random variable $Y = \ln X$ whose pdf is $p_Y(y) = \exp(y) \cdot p_X(\exp(y))$. Now, the diversity of X can be expressed as

$$\mu(X) = -\int_{-\infty}^{+\infty} p_Y(y) \ln \frac{p_Y(y)}{\exp(y)}\, dy = \mu(Y) + \mathrm{E}\big[Y\big] \quad (3)$$

where $\mathrm{E}\big[\cdot\big]$ denotes the mean or expected value of the argument (see also Eq. 15). An important conclusion is that diversity computed in the original scale and in the logarithmic scale differ only in E[ln X]. This fact will be used to define a proper normalization of data.

Nonparametric approaches: by octaves—The most popular and immediate nonparametric method to estimate size diversity is to compute a histogram of data based on octave intervals, i.e., unitary intervals in the $\log_2$ scale for sizes (Blanco et al. 1994). The estimated histogram is used as an estimation of the pdf, and the diversity computation is then

$$\hat{\mu}_{\mathrm{oct}} = -\sum_{k} \Delta_k \frac{\nu(\eta_k)}{n\Delta_k} \ln \frac{\nu(\eta_k)}{n\Delta_k} \quad (4)$$

where $\nu(\eta_k)$ is the number of sample elements in the k-th octave defined by the interval $\eta_k = (2^{k-1}, 2^k)$, and Δ_k is the length of η_k; n denotes the number of data in the sample, and the sum is extended to the octaves η_k such that $\nu(\eta_k)$ is non-null. Octaves are easily expressed in the logarithmic scale because, taking $\log_2$, the octave is the interval $\log_2 \eta_k = \big[k-1, k\big]$ or, using natural logarithms $\ln \eta_k = \big[(k-1)\ln 2, k\ln 2\big]$. This approach tries to solve a typical question when dealing with nonparametric approaches based on the histogram: how long and how many intervals have to be defined to approach the diversity integral (Eq. 2). Equation 4 establishes a logarithmic scale for sizes and, accordingly, the length of the intervals increase exponentially with the size. Moreover, the number of intervals appearing in Eq. 4 is limited to those intervals with some data within them and is thus determined by the sample.

Modified versions of this approach are found in literature, and differences come from the scale and length of the intervals and the method of interpolation of the histogram. For instance, Eq. 4 corresponds to a constant interpolation of the pdf within each octave but pdf can be interpolated linearly (Ruiz 1994), intervals of equal length can be used, etc.

Nonparametric approaches: kernel estimation—Kernel estimation techniques are used to estimate a pdf from a sample, and they became standard two decades ago (e.g., Silverman

1986). A kernel function is essentially a pdf, usually symmetric, whose dispersion is controlled by a bandwidth parameter. The estimator is a sum of kernel functions centered at the sample points. This gives a continuous estimator of the pdf whose roughness is controlled by a bandwidth parameter. Difficulties may arise when the sample space is constrained to some interval, as in the case of positive sizes. Certainly, the estimator of the pdf should vanish for negative values. A way to deal with positive sizes is to transform the positive real line into the whole real line and then to carry out the kernel estimation in the transformed space (Bowman and Azzalini 1997). A back-transformation into the positive real line provides an estimation in the original support. Following these ideas, a first step is to transform the available sample into the real line by a logarithmic transformation $y_k = \ln x_k$, $k = 1, \ldots, n$.

There are many possible choices of the kernel function, but a Gaussian pdf is a standard one. Since the density estimation is carried out on the log-transformed data, there is no constraint on the support of the pdf and Gaussian kernel does not produce any further problem. The standard Gaussian kernel estimator is

$$\hat{p}_Y(y) = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2} \frac{(y - y_k)^2}{\sigma^2} \right] \quad (5)$$

where σ represents the bandwidth parameter. For positive values of x, the estimate of $p_X(x)$ is readily obtained

$$\hat{p}_X(x) = \hat{p}_Y(\ln x) \frac{1}{x} \quad (6)$$

where the factor $1/x$ is due to the change of scale from X to Y. The value of σ should be determined by some standard method in kernel estimation. In this case, it can be taken as $\sigma = 1.06\, S_Y/n^{1/5}$ (Silverman, 1986, p. 45), where S_Y^2 is an estimator of the variance of the transformed sample y_k, $k = 1, \ldots, n$. This choice of sigma is optimum when estimating a normal density, and it is assumed to be suitable for the present estimation. Another standard choice of the bandwidth may be $\sigma = 0.9\, A_Y/n^{1/5}$ where A_Y is the minimum value of S_Y and the Y-sample interquartile range over 1.34 (Silverman, 1986, p. 48). This approach is sensible in case of heavily skewed or multimodal distributions. A natural estimator of diversity is obtained substituting Eq. 5 in Eq. 3 or, alternatively, Eq. 6 in Eq. 2. Although the mentioned choices of the bandwidth are optimal with respect to some criteria when estimating the probability density of Y, they may be sub-optimal for estimating diversity. Further statistical studies should be performed on this topic.

There are still several options to compute the so-obtained integrals. Two of them are used here. A first one consists in a numerical integration in the log-scale by the trapezoidal rule. The first step is to define a long enough interval, (z_0, z_J), in the y-axis such that $p_Y(y)$ is almost null outside it; an equally spaced sequence of points z_j, $j = 0,1,\ldots,J$, divide (z_0, z_J) into subintervals and then the integral is approximated by

$$\hat{\mu}_{kerI}(X) = \frac{-1}{2h}\sum_{j=1}^{J}\left\{\left[\hat{p}_Y(z_{j-1})\ln\hat{p}_Y(z_{j-1})\right] + \left[\hat{p}_Y(z_j)\ln\hat{p}_Y(z_j)\right]\right\} + \bar{y} \quad (7)$$

where h is the length of the subintervals and should be small enough to guarantee a reliable result. The right hand term in Eq. 7, $\bar{y} = n^{-1}\sum y_k$, is the average of the log-transformed sample and accounts for the term $E[Y]$ in Eq. 3. Apparently, the selection of the integration interval (z_0, z_J) and the number of points J makes this approach quite unpractical. However, the use of the log-scale allows selecting a very long interval and a large number of points valid for most of the cases. Further criticism can be stated on this kind of estimate because it relays on the assumed accuracy of $\hat{p}_Y$.

An interesting alternative is to approximate the diversity integral Eq. 3 using the Monte Carlo approach (e.g., Robert and Casella 2000). The basic principle of Monte Carlo methods is that an expected value of a random variable, for instance $E[Y]$, can be estimated by an average of a simple sample from Y. This fact is frequently used in statistics to estimate mean values. Moreover, the accuracy of the approach depends on the sample size n, more precisely, the standard deviation of the estimate decrease with $n^{1/2}$ as derived from the weak law of the large numbers. In the case of diversity, the log-transformed sample, $y_1, y_2, \ldots, y_n$, is available and is assumed to come from the random variable Y. The diversity of Y is then approached by

$$\mu(Y) = -\int_{-\infty}^{+\infty} p_Y(y)\ln p_Y(y)dy = -E\left[\ln p_Y(Y)\right] \cong -\frac{1}{n}\sum_{i=1}^{n}\ln p_Y(y_i) \quad (8)$$

Accepting the kernel estimator $\hat{p}_Y(y)$ in Eq. 5 as an adequate estimator of $p_Y(y)$, and substituting $\hat{p}_Y(y)$ into Eq. 8, a new estimator of $\mu(Y)$ is obtained. The expression of the estimator of $\mu(Y)$ is

$$\hat{\mu}_{kerMC}(X) = \bar{y} - \frac{1}{n}\sum_{k=1}^{n}\ln\left[\frac{1}{n\sqrt{2\pi}\,\sigma}\sum_{j=1}^{n}\exp\left(-\frac{1}{2}\frac{(y_k - y_j)^2}{\sigma^2}\right)\right] \quad (9)$$

where again $\bar{y}$ estimates $E[Y]$.

The most important difference between the integrated estimator $\hat{\mu}_{kerI}(X)$ in Eq. 7 and the Monte Carlo approach $\hat{\mu}_{kerMC}(X)$ in Eq. 9 is that the second one does not try to exploit the shape of the estimated pdf (Eq. 5) using more integration points than sample points so that the relevance of the pdf interpolation is minimized.

Size diversity measurements in simulated samples—To compare the size diversity values obtained by the different approaches described before and to analyze the accuracy in each case, a set of simulations has been carried out. Some standard probability distributions with known diversity (exponential, normal, log-normal and generalized Pareto; see Web Appendix A) have been simulated for fixed values of their parameters. For each distribution and value of parameters, 1000 simulated samples, 500 cases each, have been obtained using the inverse cumulated probability distribution method (Robert and Casella 2000). For each sample, eight size diversity values have been computed, and they can be grouped into three categories:

1. Theoretical size diversity, μ_{exact}, obtained from expressions given in Appendix A and the value of the parameters used to generate the sample,

2. Parametric estimations corresponding to the four distribution types tested (exponential, $\hat{\mu}_E$; generalized Pareto, $\hat{\mu}_{GPD}$; normal, $\hat{\mu}_N$; log-normal, $\hat{\mu}_{LN}$). They are obtained estimating parameters and, then, substituting in the theoretical size diversity (Web Appendix A),

3. Nonparametric approaches consisting of three size diversity values obtained from the corresponding nonparametric procedure: Gaussian-Kernel estimation and Monte Carlo integration $\hat{\mu}_{kerMC}$ (9), Gaussian-Kernel estimation and trapezoidal integration $\hat{\mu}_{kerI}$ (7), and simple integration by octaves $\hat{\mu}_{oct}$ (4).

The comparison (Fig. 1) between μ_{exact} and the mean of the parametric size diversity estimations coming from the related 1000 samples shows that, as expected, the mean does not differ from μ_{exact} (standard deviations range from 2.5×10^{-2} in Fig. 1C to 4.54×10^{-1} in Fig. 1F), when the estimated distribution and that used to generate the samples are of the same type. Size diversity values are also similar to μ_{exact} when the estimated distributions have similar shape to the theoretical one, but may differ substantially in other cases. Concerning nonparametric approaches, mean values of $\hat{\mu}_{kerMC}$ are very similar to μ_{exact} in all cases, being the differences between the mean of the $\hat{\mu}_{kerMC}$ estimations and μ_{exact} always less than 15%, except for the log-normal simulations (Fig. 1E, 1F), whose differences may reach 70%. Diversity $\hat{\mu}_{kerI}$ always gives values slightly higher than $\hat{\mu}_{kerMC}$. This fact may be caused by several reasons, such as the biasing of the kernel-based estimator of diversity. Moreover, $\hat{\mu}_{kerMC}$, theoretically, is integrating over the whole real line and in practice the integral is based only in the values present in the sample. Finally, the integration by octaves gives irregular results, being $\hat{\mu}_{oct}$ very similar to μ_{exact} in some cases (Fig. 1A–F) but very different in other cases (Fig. 1E,G,H). Standard deviations of estimated size diversities are similar to the corresponding standard deviation when using the simulated distribution in the corresponding parametric method. This suggests that deviations are mainly inherent to the process of simulation and that the source of variability is not due to method of estimation.

Standardization of size diversity—Changes in size diversity values caused by transformation of the variable are described in Appendix B including shifting, scaling, truncation, and log or power tranformations.

In order to compare diversities of populations with different scales, a standardization of data is necessary. A first approach to standardize data is to multiply data times a factor. Let X be the reference size in the original scale and x a vector containing a sample of X. A standardized variable Y may be obtained scaling by a constant c, i.e., $Y = X/c$. The question is to decide which is an appropriate value for c. Some approaches have been used to this end. Standard ones are $c = \min[x]$, the sample minimum; $c = \bar{x}$, the sample mean; or $c = \sigma_X$, the
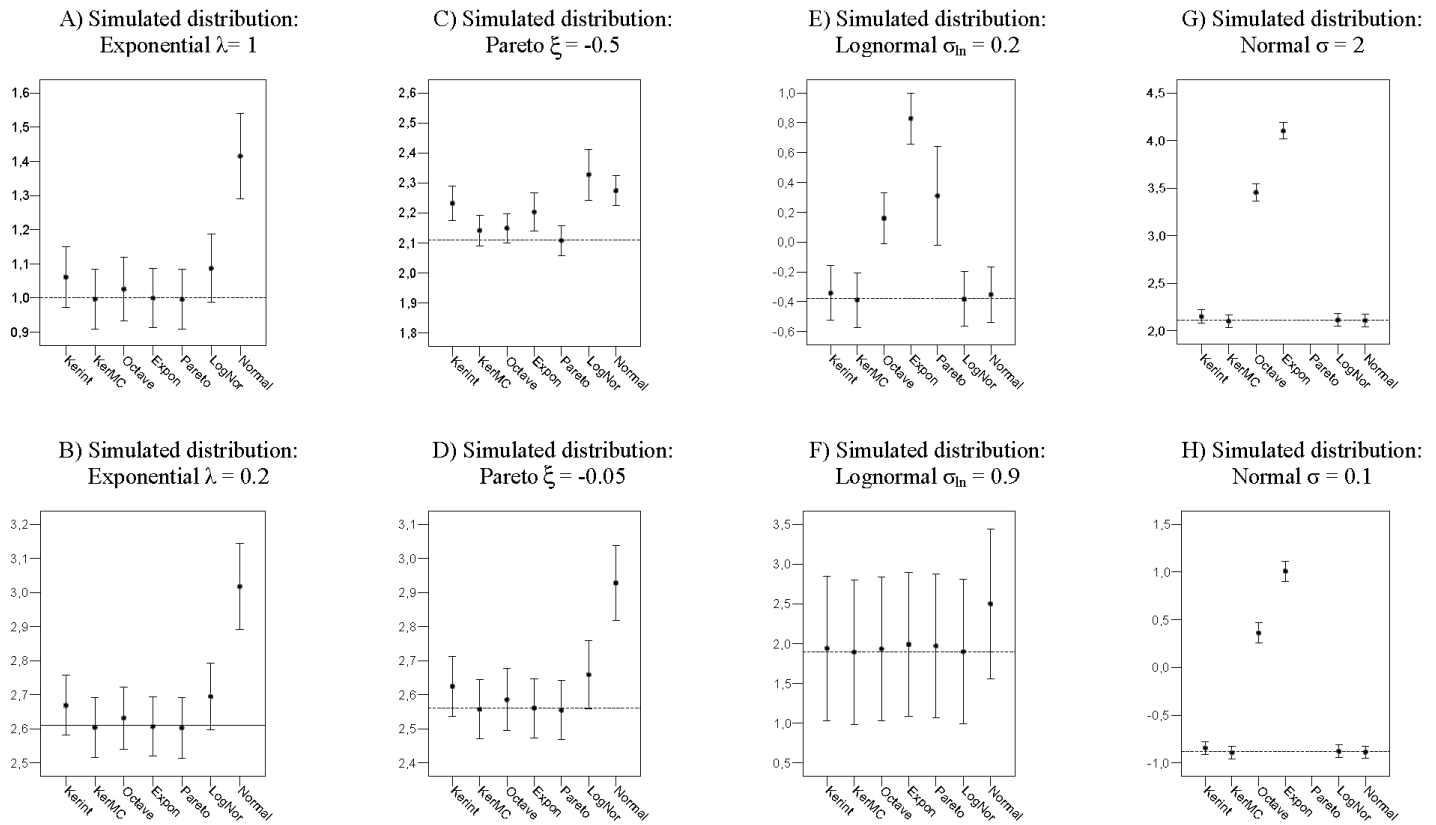
Fig. 1. Comparison of the size diversity measurements, obtained using different approaches, for simulated distributions. Approaches are nonparametric kernel, integrated by the trapezoidal rule (Kerint = $\hat{\mu}_{kerI}$) (Eq. 7) and using Monte Carlo (KerMC = $\hat{\mu}_{kerMC}$) (Eq. 9); discrete octaves model (Octave = $\hat{\mu}_{oct}$) (Eq. 4); parametric: exponential (Expon = $\hat{\mu}_E$), generalized Pareto (Pareto = $\hat{\mu}_{GPD}$), log-normal (LogNor = $\hat{\mu}_{LN}$), normal (Normal = $\hat{\mu}_N$). For each pdf, and for the 1000 simulated samples (500 cases), the mean value of the estimated diversity and a ± 2 standard deviations interval are represented. In each panel, simulated distribution and the parameters used are shown. Horizontal line indicates the theoretical value of diversity ($\hat{\mu}_{exact}$). Size-diversity values computed using Pareto are less than 2 in G and less than –1 in H and are not represented.

standard deviation. Platt and Denman (1977) proposed the first one, $c = \min[x]$ for data whose abundance decrease with size. However, the estimation of this minimum has an important problem: the large variability of the minimum of a sample, especially in normal-shaped distributions or other ones where small sizes are scarce. This may lead to very different results even for samples from the same population. Alternatively, the minimum may be determined by the characteristics of the measuring device, in this case, comparisons of diversities coming from different devices may be affected by artifacts.

The option $c = \sigma_X$ seems to be inadequate because σ_X participate directly in the computation of the diversity and part of this information may be lost in the standardization. For instance, after this kind of standardization all normal pdfs would correspond to size diversities equal to $(1 + \ln[2\pi])/2$ (see Appendix A). The option $c = \bar{x}$, or even the division by the sample median, can be accepted as candidates.

A typical statistical standardization includes a translation as in $Y = (X - m_x)/ \sigma_X$, where m_X denotes the mean of X. This option may be inconvenient when dealing with positive sizes,

because the standardized sample is not positive and the intrinsic lower limit, 0 in positive data, is then lost. Furthermore, this standardization includes σ_X which seems again inappropriate.

The proposal is to standardize the data taking $c = g(x)$, the sample geometric mean of the positive data. This approach is based on three criteria: (a) the natural scale of most positive data; (b) the easy treatment and comparison of diversities calculated on equivalent samples with different dimensionality given in length, area, or volume; and (c) a technical property that makes the calculation of diversity in ordinary and logarithmic scales equivalent.

The first criterion (a) is that positive measurements are frequently given in a relative scale. For instance, lengths of 1 μm and 2 μm may be considered as one being the half of the other one, whereas lengths of 100 μm and 101 μm are considered as very similar. This claims for treating the data in logarithmic scale. In these conditions, a center of the distribution of X is $\exp(E[\ln X])$ (Pawlowsky-Glahn and Egozcue 2001), i.e., transform the variable as $Y = \ln X$, compute the mean of Y, $E[Y]$, and then back-transform it taking exponential. When evalu-

ating the center exp(E[ln X]) from a n-sample, x, the sample geometric mean

$$g(\mathrm{x}) = (x_1 \cdot x_2 \cdots x_n)^{1/n} = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\ln x_i\right) \qquad (10)$$

is an appropriated estimator (Aitchison 1986).

Criterion (b) tries to homogenize the results of diversity when computed from equivalent samples with different dimensionality. For instance, let X be a random length, e.g., a radius of a sphere, and let V be a scaled power of X, defined by $V = aX^k$, e.g., the volumen of a sphere $a = 4\pi/3$, $k = 3$. The relationship of the corresponding diversities is given in Eq. 16,

$$\mu(V) = \mu(X) + \ln(ka) + (k-1)\mathrm{E}\left[\ln X\right].$$

Assume that both variables, V and X, are divided by the respective centers,

$$Y = \frac{X}{\exp\left(\mathrm{E}\left[\ln X\right]\right)} ; W = \frac{V}{\exp\left(\mathrm{E}\left[\ln V\right]\right)}$$

as corresponds to the proposed standardization. Taking into account the scaling property of Eq. 14, the power transformation of variables (Eq. 16) and that

$$\mathrm{E}\left[\ln V\right] = \ln a + k\mathrm{E}\left[\ln X\right]$$

the relationship between the size diversities of the standardized variables is obtained:

$$\mu(W) = \ln k + \mu(Y) \qquad (11)$$

Equation 11 reveals that dimension of the measurements only appears as a known additive constant, ln 2, ln 3 in case of areas and volumes respectively, whereas the constant a does not play any role. As a consequence of the standardization, no reference to the moments of the variable, e.g., E[ln X], E[X], appear in Eq. 11, and thus it is independent of the particular scale or units in which the sample is given.

Criterion (c) is mainly of technical character. The standardization by the geometric mean implies that the logarithm of $Y = X/\exp(\mathrm{E}(\ln X))$ has null mean, E[ln Y] = 0. Due to the property Eq. 15, the diversity of Y and its log-diversity are equal: $\mu(Y) = \mu(\ln Y)$. Again, this suggests all computation should be carried out in log-scale.

From a theoretical point of view, the proposed standardization can be carried out just taking logs of the sample and computing the size diversity of the log-transformed sample; if desired, subtraction (addition) of ln k reduces to one-dimensional (k-dimensional) measurements. In fact, subtraction of the logarithmic mean is not necessary because a shifting does not alter the diversity result (Eq. 13). However, the pdf of the log-transformed data changes accordingly, and characteristics of the estimation of these new pdfs may change, especially in the case of parametric methods. In practice, the standardization is adequately carried out following the steps: (1) purge null data and compute the geometric mean of the sample using Eq. 10; (2) divide the whole sample by the geometric mean; (3) com-

pute the size diversity using a suitable approach, e.g., Eq. 9; (4) if necessary, add or subtract ln k to account for dimensionality of data in agreement with Eq. 11.

Assessment

The assessment of size diversity measurements has been carried out by measuring it in available size distributions of natural samples. The size distributions of these natural samples have been analyzed in order to decide whether or not they adjust to some parametric distribution models. Moreover, the values of size diversity obtained for these natural samples using the different parametric and nonparametric estimations have been compared. For this comparison, the size distributions of five different communities of aquatic organisms have been analyzed. Among them, there are remarkable differences in size, sampling procedure, counting, and size measurement. Samples of bacterioplankton, phytoplankton, zooplankton, hypobenthic, and epibenthic aquatic invertebrates have been selected, and the resulting values of size diversity have been compared. Sample sites, all located in Girona (NE Spain), include planktonic and hypobenthic organisms of two coastal Mediterranean salt marshes (Empordà salt marshes and Baix Ter salt marshes) and epibenthic organisms of one temporary and groundwater flooded karstic pond (Espolla Pond).

Sampling procedure in natural samples—Bacterioplankton and phytoplankton samples were taken monthly in a group of brackish coastal lagoons in the Baix Ter salt marshes. Bacterio- and phytoplankton samples were counted using a Facscalibur flow cytometer. Samples were filtered through 50 μm mesh, fixed with 1% parafomaldehyde and 0.05% glutaraldehyde (final concentration), deep frozen in liquid nitrogen, and stored frozen at –20°C. The transformation of the forward scatter (FSC) signals of the flow cytometer to cell volume was carried out through a calibration curve as described in López-Flores et al. (2006). Bacterial biomass estimations were performed following Gasol et al. (2000). Zooplankton samples were randomly selected from a weekly frequency study in several brackish water basins (permanent and temporary) of the Empordà salt marshes. Each zooplankton sample was taken from 4 L filtered 53 μm water, at a depth of 15–20 cm, and was preserved in situ in 4% formalin. Hypobenthic samples were also taken in the same basins using 5.2 cm diameter cores (5 cm depth) and sieved through 300 and 100 μm mesh-size sieves. Epibenthic organisms were sampled weekly during 1 hydroperiod, which occurred between December 1996 and March 1997 in a temporary fishless karstic pond (Espolla Pond). Seven 20 m transects for each sample were performed using a modified Elster beam trawl (opening of 50 × 30 cm^2 and a mesh size of 250 μm) that was dragged the length of the transect. Samples were preserved in situ with 4% formalin. Zooplankton, hypobenthic, and epibenthic invertebrate biomass (dry mass) were estimated from length or biovolume using existing equations for microinvertebrates and macroinvertebrates. More details on sampling procedure, references
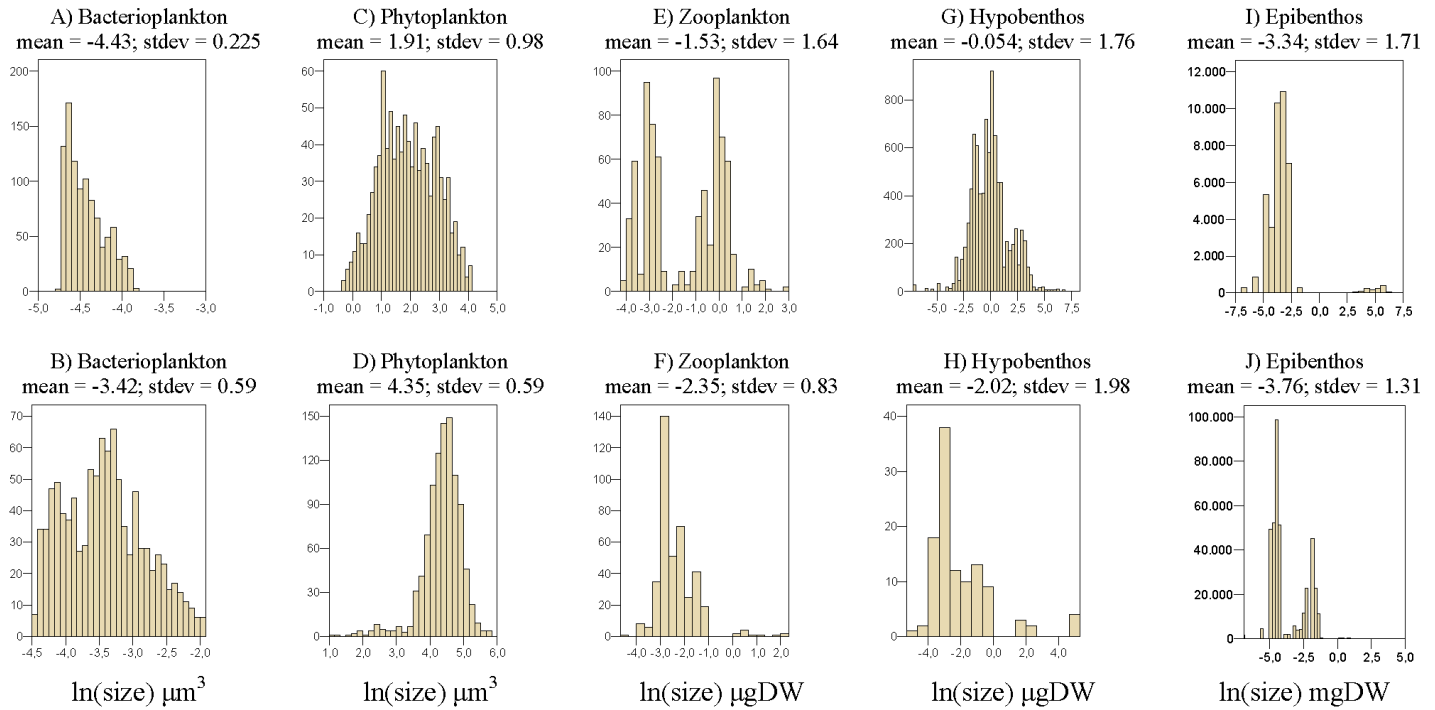
Fig. 2. Examples of size distributions in natural samples. Histograms are given in absolute frequency (vertical axes).

used for biomass estimations, and ecological characteristics of waterbodies sampled are found in López-Flores et al. (2006) for bacterio- and phytoplankton, Brucet et al. (2005) for zooplankton, Gascón et al. (2005) for hypobenthic invertebrates, and Boix et al. (2004) for epibenthic invertebrates.

Size distributions in natural samples—Size distributions found in natural samples analyzed were very variable as is shown in examples of Fig. 2. Most found histograms were unimodal and had more or less symmetric shape (see examples of these normal-shaped distributions in Figs. 2B, 2C, and 2G). Some histograms had the mode displaced to the small sizes, resembling exponential-shaped distributions (Figs. 2A and 2C), whilst others had the mode displaced to great sizes, and closer to a generalized Pareto distribution (Fig. 2D). Bimodal histograms were also present. In some cases, such as in the epibenthic distributions, the modes found coincided with the presence of organsims of two different functional groups (Fig. 2J), but also bimodal distributions were found in intraguild size distributions, as is the case of zooplankton distributions of Fig. 2E. Irregular histograms were also frequent (Figs. 2F, 2H, and 2I). Ecological relevance of these irregularities has been described elsewhere (Boix et al. 2004; Brucet et al. 2006; López-Flores et al. 2006).

Goodness-of-fit of the parametric distributions considered (normal, N; exponential, E; log-normal, LN; and generalized Pareto, GPD) to the samples have been tested using the Kolmogorov-Smirnov (K-S) procedure (e.g., Rohatgi 1976). All epibenthic and hypobenthic size samples were rejected ($p < 0.01$) to be distributed as any hypothetical distribution. Adjustment

to a log-normal distribution could not be discarded in 7.1% of zooplankton samples ($p > 0.05$). The K-S results depended on the extension of size-sample considered. For example, the percentage of phytoplankton samples that could not be discarded to fit a log-normal distribution increases from 25%, when 1000 individuals were sized, to 66.7%, when individuals sized were only 100 (Table 1). Consequently, there is no reason, neither theoretical nor empirical, to affirm that size distributions follow a determinate parametric pdf.

Size diversity measurements of natural samples—Standardized size diversity values obtained for the different size distribution types are compared in Table 2. Bacterioplankton size distributions showed significantly lower values of $\hat{\mu}_{kerMC}$ when compared with those of the other size distribution types analyzed, since size ranges of bacterioplankton distributions were very small. On the other hand, $\hat{\mu}_{kerMC}$ values of hypobenthic invertebrate distributions, whose size range was the largest, were significantly higher. No significant differences were found in size diversity values of phytoplankton, zooplankton, and epibenthos distributions, although their size ranges differed strongly (Table 2). Thus, differences in size diversity were not only attributable to differences in size range, but also to the relative abundance of sizes, as it is inherent to the Shannon-Wiener expression. Further discussions of the ecological meaning of the variation of size diversity in the zooplankton community (computed using a parametric approach) are found in Brucet et al. (2006) and Badosa et al. (2007).

Table 1. Kolmogorov-Smirnov goodness-of-fit test results for natural samples

	Samples	*n*	K-S	E	N	LN	GPD	None
		2000	**	100	100	100	100	
Epibenthos	15	to	*	0	0	0	0	100
		10^5	+	0	0	0	0	
		112	**	100	100	100	100	
Hypobenthos	14	to	*	0	0	0	0	100
		13776	+	0	0	0	0	
		63	**	100	95.2	85.7	100	
Zooplankton	42	to	*	0	2.4	7.1	0	81.0
		531	+	0	2.4	7.1	0	
			**	100	100	66.7	100	
Phytoplankton	24	1000	*	0	0	8.3	0	66.7
			+	0	0	25.0	0	
			**	83.3	75.0	25.0	79.2	
Phytoplankton	24	100	*	8.3	12.5	8.3	16.7	16.7
			+	8.3	12.5	66.7	4.2	
			**	100	100	83.3	100	
Bacterioplankton	24	1000	*	0	0	16.7	0	83.3
			+	0	0	0	0	
			**	100	50.0	20.8	100	
Bacterioplankton	24	100	*	0	33.3	8.3	0	20.8
			+	0	16.7	70.8	0	

Parametric distributions tested: exponential, E; normal, N; log-normal, LN; and generalized Pareto, GPD. *P* values are classified into three groups: ** $p < 0.01$; * $0.01 < p < 0.05$; + $p > 0.05$. Number of samples is given in % over the total number specified in *Samples*. Under *None*, the number (%) of samples for which all tested distribution were rejected at $p > 0.05$ significance.

Figure 3 shows the comparison, in all these natural samples, of the size diversity values obtained by the kernel Monte Carlo approach ($\hat{\mu}_{kerMC}$) with those size diversity values obtained with other approaches tested: the log-normal, the exponential, and the octave approaches ($\hat{\mu}_{LN}$, $\hat{\mu}_{E}$, and $\hat{\mu}_{oct}$). The discrete approach of octaves gave very similar size diversity values to the kernel-MC approach (Fig. 3A), except when size diversities were close to 0. In these cases, where sizes strongly accumulate close to a determinate size, a discrete approach will always overestimate the size diversity. In most of cases, the obtained values of $\hat{\mu}_{LN}$ were also similar, but slightly higher, than $\hat{\mu}_{kerMC}$ (Fig. 3B). These lower values of $\hat{\mu}_{kerMC}$ may be a consequence of the integration limited to the

values in sample. Although K-S values suggest that most of the size distributions of natural samples do not fit to any of the probability density functions tested (see Table 1), differences in $\hat{\mu}_{LN}$ and $\hat{\mu}_{kerMC}$ were appreciable only when size distributions strongly differed from a log-normal distribution (Fig. 3B). The exponential approach gave more unsatisfactory results. $\hat{\mu}_{E}$ values generally were higher than $\hat{\mu}_{kerMC}$, and there are a lot of cases where $\hat{\mu}_{E}$ strongly differed from $\hat{\mu}_{kerMC}$ (Fig. 3C). These differences can be attributed to the fact that the number of size distributions that strongly differ from an exponential function were especially high (see Fig. 1 and Table 1). It might be concluded that the different approaches give different results of size diversity only when the shape of

Table 2. Ranges (original and standardized after division by the geometric mean) and standardized diversity index (mean and standard deviation over a number of samples)

	Samples	Size range (min, max)	Units	Std. size range (min, max)	μ_{kerMC} (mean, SD*)
Bacterioplankton	24	8.1×10^{-3}, 6.5×10^{-1}	μm^3, biovolume	0.3, 32	0.61 (a), 0.47
Phytoplankton	24	1.6×10^{-1}, 4.9×10^3	μm^3, biovolume	2.1×10^{-2}, 6.1×10^2	1.58 (b), 0.35
Zooplankton	42	3.1×10^{-4}, 2.1×10^2	μg, dry weight	7.1×10^{-3}, 4.0×10^3	1.49 (b), 0.52
Hypobenthos	14	2.6×10^{-4}, 1.3×10^5	μg, dry weight	2.5×10^{-4}, $1.2 \cdot 10^5$	2.53 (c), 0.38
Epibenthos	15	1.0×10^{-3}, 9.0×10^2	mg, dry weight	4.3×10^{-2}, 9.8×10^4	1.35 (b), 0.63

Results of μ_{kerMC} followed by the same letter are not significantly different ($p < 0.05$), but differ if they are followed by different letter.
*SD, standard deviation
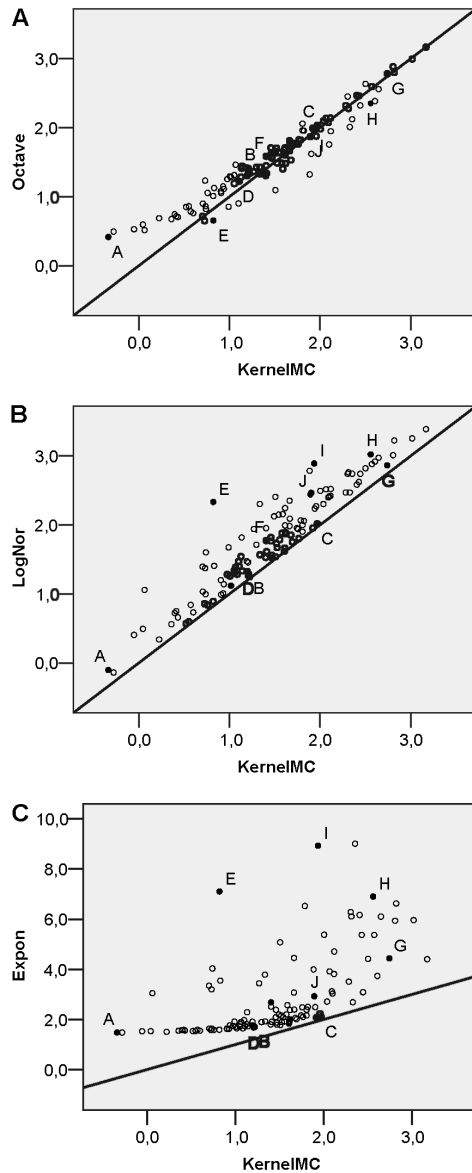
Fig. 3. Comparison of the size diversity values in natural samples. The kernel-Monte Carlo approach (KernelMC = $\hat{\mu}_{kerMC}$) is compared with the discrete octave (Octave = $\hat{\mu}_{oct}$), parametric log-normal (LogNor = $\hat{\mu}_{LN}$), and exponential (Expon = $\hat{\mu}_E$) approaches. The straight line represents equal values of the compared estimations. Filled points with capital letters indicate size samples shown in Fig. 2.

the natural size distribution is strongly different than that of the parametric pdf used.

Discussion

A nonparametric approach appears to be more suitable than parametric approaches for size diversity estimations in most types of natural samples. This is due to the fact that most size distributions observed in nature do not fit available parametric families. Although some processes or some generalities

in size distributions can be well modeled (Platt and Denman 1977; Dickie et al. 1987; Thiebaux and Dickie 1993; Quiñones 1994; Han and Straskraba 1998), there is no evidence that any parametric family appropriately models all phenomena affecting size distributions, such as the combined effects of size based allometric growth with predatory and competitive interactions. Irregularities in density, such as gaps or dumps, are also relevant in the study of size distributions (Rodríguez 1994) but neglected when a parametric approach is used. The kernel approach provides a universal method for the estimation of the size diversity of size distributions found in natural samples, since it gives accurate estimations of size diversity for all the probability distributions tested. Furthermore, it is applicable to all distribution types, even if they do not fit to any parametric distribution, such as multimodal distributions (Havlicek and Carpenter 2001; Ruiz et al. 2002).

The standardization by means of the division by the geometric mean have several advantages when compared with other standardization methods: the estimation of the geometric mean is always more robust than the estimation of the minimum, the same size diversity values were obtained when using original size data or data log transformed and, finally and importantly, samples where size were measured with different dimensionality (longitudes, areas, volumes, or biomasses) may be immediately compared, by the simple addition of ln k, where k is the dimension used. This makes the results independent from the measuring method used, with the consequent advantages for data comparison and for further analysis of regularities in different ecosystems.

The most frequently used methodology in studies related to size distributions is the normalized biomass size spectrum (NBSS), where organism abundances or biomass are plotted in front of their size (Blanco et al. 1994). The resulting plot, in double logarithmic scale, is a straight line with a negative slope b with a value close to –1. Many works analyzed the variation of this slope and its ecological relevance (Rodríguez and Mullin 1986; Rodríguez et al. 1987; Gaedke 1992; Quintana et al. 2002). Since a straight line in a log-log plot may be modeled as an exponential-shaped pdf, the negative slope b of the NBSS may be easily related to the size diversity. Thus, the size diversity gives a more general theoretical context, within which the discussion related to the slope of the NBSS and its variations is a particular case. Furthermore, some authors pointed out that some nonlinear parametric models give better fits to size distributions in natural samples than the NBSS straight line (Gasol et al. 1991; Vidondo et al. 1997; Brucet et al. 2005). However, these nonlinear models have the disadvantage that the ecological meaning of the parameters that define the model is difficult to interpret. The size diversity gives a unique value per size distribution, which integrates the amplitude of the size range and the evenness, that is, the relative distribution of sizes along the size range, in the same way that Shannon diversity integrates species composition and species relative abundances.

The Shannon species diversity is a universal tool in studies in community structure based on the analysis of species composition and its variations. However, this traditional species-based approach may fail when the correct determination of all species found is not easy, such as in the case of studies dealing with phytoplankton, especially if they include nanoplankters and picoplankters. Modern techniques in phytoplankton counting, such as flow cytometry, provide accurate information of cells counting and its size, but not on species composition. It is also difficult to obtain an accurate species diversity measure in the case of some invertebrate communities dominated by hard to determine larval stages, such as naupliar stages of copepods, which in most cases have been operationally ignored (see Hopcroft et al. 1998, and references therein), although they dominate most aquatic communities (Hopcroft et al. 2001). These problems might be solved by using a functional approach which classifies organisms either by trophic levels (McQueen et al. 1986) or by trophic guilds (e.g., Cohen et al. 1993; Yodzis 1993). This functional approach provides important information about the food web structure (Gaedke 1995) and is especially useful for macroecology studies (Brown 1995) or studies dealing with the whole aquatic organisms (e.g., from bacteria to fishes).

Size diversity provides a useful tool for the study of size-structured communities (Werner and Gilliam 1984), that is, those communities where size plays an important role in their ecological interactions and where the traditional taxonomic or functional approaches may be not suitable. Two cases, very commonly found in aquatic ecology fit with this: 1) when several species with a similar trophic role may coexist and their competitive interactions are size dependent. This is the case of many phytoplankton communities, where competition for nutrient uptake is strongly related to size (Malone 1980; Harrison and Wood 1988; Armstrong 2003; Irwin et al. 2006), and 2) when one species undergo ontogenetic shifts in food or habitat use (Werner and Gilliam 1984; Armsby and Tisch 2006). Gradual changes on trophic niche along ontogeny have been described for many copepod species (Mullin and Brooks 1967; Paffenhöfer 1971; Gophen 1977; Poulet 1977; Hansen et al. 1994). Furthermore, studies of zooplankton communities show that size diversity and species diversity follow different patterns along succession (Brucet et al. 2006). Further analysis on the ecological relevance of these differences appears to be an interesting goal in theoretical ecology.

References

Aitchison, J. (1986). The statistical analysis of compositional data. Chapman & Hall.

Aoki, I., and T. Mizushima. (2001). Biomass diversity and stability of food webs in aquatic ecosystems. Ecol. Res. 16:65–71.

Armsby, M., and N. Tisch. (2006). Intraguild predation and cannibalism in a size-structured community of marine amphipods. J. Exp. Mar. Biol. Ecol. 333(2):286–295.

Armstrong, R. A. (2003). A hybrid spectral representation of phytoplankton growth and zooplankton response: The "control rod" model of plankton interaction. Deep Sea Res. II 50:2895–2916.

Badosa, A., D. Boix, S. Brucet, R. López-Flores, S. Gascón and X. D. Quintana. (2007). Zooplankton taxonomic and size diversity in Mediterranean coastal lagoons (NE Iberian Peninsula): Influence of hydrology, nutrient composition, food resource availability and predation. Estuar. Coast. Shelf Sci. 71:335–346.

Basset, A., F. Sangiorgio, and M. Pinna. (2004). Monitoring with benthic macroinvertebrates: advantages and disadvantages of body size descriptors. Aquat. Conserv. Mar. Freshw. Ecosys. 14:43–58.

Blanco, J. M., F. Echevarría, and C. M. García. (1994). Dealing with size spectra: some conceptual and mathematical problems. Sci. Mar. 58(1-2):17–29.

Boix, D., J. Sala, X. D. Quintana, and R. Moreno-Amich. (2004). Succession of the animal community in a Mediterranean temporary pond. J. North Am. Benthol. Soc. 23(1): 29–49.

Bowman, A. W., and A. Azzalini. (1997). Applied smoothing techniques for data analysis: the kernel approach with S-plus illustration. Oxford Univ. Press.

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

Brucet, S., D. Boix, R. López-Flores, A. Badosa, R. Moreno-Amich, and X. D. Quintana. (2005). Zooplankton structure and dynamics in permanent and temporary Mediterranean salt marshes: taxon-based and size-based approaches. Arch. Hydrobiol. 162:535–555.

——, ——, ——, ——, and X. D. Quintana. (2006). Size and species diversity of zooplankton communities in fluctuating Mediterranean salt marshes. Estuar. Coast. Shelf Sci. 67(3):424–432.

Cohen, J. E., and others. (1993). Improving food webs. Ecology 74:252–258.

——., T. Jonsson, and S. R. Carpenter (2003). Ecological community description using the food web, species abundance, and body size. Proc. Nat. Acad. Sci. 100(4): 1781–1786.

Dickie, L. M., S. R. Kerr, and P. R. Boudreau. (1987). Size-dependent processes underlying regularities in ecosystem structure. Ecol. Monogr. 57(3):233–250.

Duplisea, D. E. (2000). Benthic organism biomass size-spectra in the Baltic Sea in relation to the sediment environment. Limnol. Oceanogr. 45(3):558–568.

Gaedke, U. (1992). The size distribution of plankton biomass in a large lake and its seasonal variability. Limnol. Oceanogr. 37(6):1202–1220.

——. (1995). A comparison of whole-community and ecosystem approaches (biomass size distributions, food web analysis, network analysis, simulation models) to study the structure, function and regulation of pelagic food webs. J. Plank. Res. 17:1273–1305.

Gascón, S., D. Boix, J. Sala, and X. D. Quintana. (2005). Variability of benthic assemblages in relation to the hydrological pattern in Mediterranean salt marshes (Empordà wetlands, NE Iberian Peninsula). Arch. Hydrobiol. 163: 163–181.

Gasol, J. M., and P. A. Del Giorgio. (2000). Using flow cytometry for counting natural planktonic bacteria and understanding the structure of planktonic bacterial communities. Sci. Mar. 64:197–224.

———, R. Guerrero, and C. Pedrós-Alió. (1991). Seasonal variations in size structure and procaryotic dominance in sulfurous lake Cisó. Limnol. Oceanogr. 36(5):860–872.

Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. Biometrika 40: 237–264.

Gophen, M. (1977). Food and feeding habits of *Mesocyclops leuckarti* (Claus) in Lake Kinneret (Israel). Freshw. Biol. 7: 513–518.

Han, B. P. (1998). On the diversity of trophic structures and processes in ecosystems. Ecol. Model. 107:51–62.

——— and M. Straskraba. (1998). Size dependence of biomass spectra and population density. I. The effects of size scales and size intervals. J. Theor. Biol. 191:259–265.

Hansen, B., P. K. Bjornsen, and P. J. Hansen. (1994). The size ratio between planktonic predators and their prey. Limnol. Oceanogr. 39(2):395–403.

Harrison, W. G. and L. J. E. Wood. (1988). Inorganic nitrogen uptake by marine picoplankton: Evidence for size partitioning. Limnol. Oceanogr. 33(3):468–475.

Havlicek, T. D., and S. R. Carpenter. (2001). Pelagic species size distributions in lakes: are they discontinuous? Limnol. Oceanogr. 46(5):1021–1033.

Hooper, D. U., and others. (2002). Species diversity, functional diversity, and ecosystem functioning, p. 195–181. *In* M. Loreau, S. Naeem, and P. Inchausti [eds.], Biodiversity and ecosystem functioning: synthesis and perspectives. Oxford Univ. Press.

Hopcroft, R. R., J. C. Roff and F. P. Chavez. (2001). Size paradigms in copepod communities: a re-examination. Hydrobiologia 453/454:133–141.

———, ———, and D. Lombard. (1998). Production of tropical copepods in Kingston Harbour, Jamaica: the importance of small species. Mar. Biol. 130(4):593–604.

Hulot, F. D., G. Lacroix, F. Lescher-Moutoué, and M. Loreau. (2000). Functional diversity governs ecosystem response to nutrient enrichment. Nature 405:340–344.

Ichinokawa, M., and M. M. Takahashi. (2006). Size-dependent carbon flow in the epipelagic food web of the Western Equatorial Pacific. Mar. Ecol. Progr. Ser. 313:13–26.

Irwin, A. J., Z. V. Finkel, O. M. E. Schofield, and P. G. Falkowski. (2006). Scaling-up from nutrient physiology to the size-structure of phytoplankton communities. J. Plank. Res. 28(5): 459–471.

Johnson, O. (2004). Information theory and the central limit theorem. Imperial College Press.

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

Li, W. K. W. (1997). Cytometric diversity in marine ultraphytoplankton. Limnol. Oceanogr. 42(5):874–880.

López-Flores, R., D. Boix, A. Badosa, S. Brucet, and X. D. Quintana. (2006). Pigment composition and size distribution of phytoplankton in a confined Mediterranean salt marsh ecosystem. Mar. Biol. 149:1313–1324.

Lurie, D., and J. Wagensberg. (1983). On biomass diversity in ecology. Bull. Math. Biol. 45(2):287–293.

——— and ———. (1984). An extremal principle for biomass diversity in ecology, p. 257–271. *In* I. Lamprecht and A. Zotin, [eds.], Thermodynamics and regulation of biological processes. Walter de Gruyter.

Malone, T. C. (1980). Algal size and phytoplankton ecology, p. 433–464. *In* I. Morris, [ed.], Physiological ecology of phytoplankton. Blackwell.

Margalef, R. (1991). Teoría de los sistemas ecológicos. Publicacions Universitat de Barcelona.

Marquet, P. A., R. A. Quiñones, S. Abades, F. Labra, M. Tognelli, M. Arim, and M. Rivadeneira. (2005). Scaling and power-laws in ecological systems. J. Exp. Biol. 208:1749–1769.

McQueen, D. J., J. R. Post, and E. L. Mills. (1986). Trophic relationships in freshwater pelagic ecosystems. Can. J. Fish. Aquat. Sci. 43(8):1571–1581.

Mittelbach, G. G. (1981). Patterns of invertebrate size and abundance in aquatic habitats. Can. J. Fish. Aquat. Sci. 38: 896–904.

Mullin, M. M., and E. R. Brooks. (1967). Laboratory culture, growth rate, and feeding behaviour of a planktonic marine copepod. Limnol. Oceanogr. 12(4):657–666.

Oindo, B. O., A. K. Skidmore, and H. T. Prins. (2001). Body size and abundance relationship: an index of diversity for herbivores. Biodivers. Conserv. 10:1923–1931.

Paffenhöfer, G. A. (1971). Grazing and ingestion rates of nauplii, copepodids and adults of the marine planktonic copepod *Calanus helgolandicus*. Mar. Biol. 11(3):286–298.

Parsons, T. R. (1969). The use of particle size spectra in determining the structure of plankton community. J. Oceanogr. Soc. Japan 25:172–181.

Pawlowsky-Glahn, V., and J. J. Egozcue. (2001). Geometric approach to statistical analysis on the simplex. Stoch. Environ. Res. Risk Assess. 15(5):384–398.

Petchey, O. L., and K. J. Gaston. (2002). Functional diversity (FD), species richness and community composition. Ecol. Lett. 5(3):402–411.

Pianka, E. R. (1974). Evolutionary ecology. Harper and Row.

Pielou, E. C. (1969). An introduction to mathematical ecology. Wiley-Interscience.

Piontkovski, S., and S. van der Spoel. (1998). Similarities in size, in taxonomic, and in ecological mesoscale structure of tropical zooplankton communities, p. 297–306. *In* A. C. Pierrot-Bults and S. van der Spoel [eds.], Pelagic biogeogra-

phy ICoPB II, Proceedings of the 2nd international confer-
ence IOC/UNESCO. UNESCO, Paris.

Platt, T., and K. Denman. (1977). Organisation in the pelagic
ecosystem. Helgoländer Wissenschaftliche Meeresunter-
suchungen 30:575–581.

———, M. Lewis, and R. Geider. (1984). Thermodynamics of
the pelagic ecosystem: Elementary closure condition for
biological production in the open ocean, p. 49–84. *In*
M. J. R. Fashham [ed.], Flows for energy and materials in
marine ecosystems. NATO Conference Series. IV Marine
Sciences 13. Plenum Press.

Poulet, S. A. (1977). Grazing of marine copepod development
stages on naturally occurring particles. J. Fish. Res. Board
Can. 34:2381–2387.

Quiñones, R. A. (1994). A comment on the use of allometry in
the study of pelagic ecosystem processes. Sci. Mar. 58(1-2):
11–16.

Quintana, X. D., F. A. Comín, and R. Moreno-Amich. (2002).
Biomass-size spectra in aquatic communities in shallow
fluctuating Mediterranean salt marshes (Empordà wet-
lands, NE Spain). J. Plank. Res. 24(11):1149–1161.

Quiroga, E., R. Quiñones, M. Palma, J. Sellanes, V. A. Gallardo,
D. Gerdes, and G. Rowe. (2005). Biomass size-spectra of
macrobenthic communities in the oxygen minimum zone
of Chile. Estuar. Coast. Shelf Sci. 62:217–231.

Rasmussen, J. B. (1993). Patterns in the size structure of littoral
zone macroinvertebrate communities. Can. J. Fish. Aquat.
Sci. 50(10):2192–2207.

Robert, C. P., and G. Casella. (2000). Monte Carlo statistical
methods. Springer.

Rodríguez, J. (1994). Some comments on the size-based struc-
tural analysis of the pelagic ecosystem. Sci. Mar. 58(1-2):
1–10.

———, F. Jiménez, B. Bautista, and V. Rodríguez. (1987). Plank-
tonic biomass spectra dynamics during a winter production
pulse in Mediterranean coastal waters. J. Plank. Res. 9(6):
1183–1194.

——— and W. K. W. Li. (1994). The size structure and metab-
olism of the pelagic ecosystem. Consejo Superior de Inves-
tigaciones Científicas and Institut de Ciències del Mar.

——— and M. M. Mullin (1986). Relation between biomass
and body weight of plankton in a steady state oceanic
ecosystem. Limnol. Oceanogr. 31(2):361–370.

Rohatgi, V. K. (1976). An introduction to probability theory
and mathematical statistics. Wiley.

Ruiz, J. (1994). The measurement of size diversity in the
pelagic ecosystem. Sci. Mar. 58(1-2):103–107.

———, L. Prieto, and F. Ortegón. (2002). Diatom aggregate for-
mation and fluxes: a modeling analysis under different size-
resolution schemes and with empirically determined aggre-
gation kernels. Deep-Sea Res. I 49(3):495–515.

Schwinghamer, P. (1981). Characteristic size distributions of
integral benthic communities. Can. J. Fish. Aquat. Sci. 38:
1255–1263.

Shannon, C. E., and W. Weaver. (1949). The mathematical the-
ory of communication. Univ. of Illinois Press.

Silverman, B. W. (1986). Density estimation for statistics and
data analysis. Chapman & Hall.

Strayer, D. L. (1991). Perspectives on the size structure of lacus-
trine zoobenthos, its causes, and its consequences. J. North
Am. Benthol. Soc. 10(2):210–221.

Thiebaux, M. L., and L. M. Dickie. (1993). Structure of the
body-size spectrum of the biomass in aquatic ecosystems: a
consequence of allometry in predator-prey interactions.
Can. J. Fish. Aquat. Sci. 50(6):1308–1317.

Vidondo, B., Y. T. Prairie, J. M. Blanco, and C. M. Duarte.
(1997). Some aspects of the analysis of size spectra in
aquatic ecology. Limnol. Oceanogr. 42(1):184–192.

Weithoff, G., N. Walz, and U. Gaedke. (2001). The intermedi-
ate disturbance hypothesis-species diversity or functional
diversity? J. Plank. Res. 23(10):1147–1155.

Werner, E. E., and J. F. Gilliam. (1984). The ontogenetic niche
and species interactions in size-structured populations.
Ann. Rev. Ecol. System. 15:393–425.

Yodzis, P. (1993). Environment and trophodiversity, p. 26–38.
In R. E. Ricklefs and D. Schluter [eds.], Species diversity in
ecological communities. Univ. of Chicago Press.