IKOS 101: 398–410, 2003

Comparing species richness among assemblages using sample units: why not use extrapolation methods to standardize different sample sizes?

Adriano S. Melo, Rodrigo A. S. Pereira, Adalberto J. Santos, George J. Shepherd, Glauco Machado, Hermes F. Medeiros and Ricardo J. Sawaya

Melo, A. S., Pereira, R. A. S., Santos, A. J., Shepherd, G. J., Machado, G., Medeiros, H. F. and Sawaya, R. J. 2003. Comparing species richness among assemblages using sample units: why not use extrapolation methods to standardize different sample sizes? – Oikos 101: 398–410.

Comparisons of species richness among assemblages using different sample sizes may produce erroneous conclusions due to the strong positive relationship between richness and sample size. A current way of handling the problem is to standardize sample sizes to the size of the smallest sample in the study. A major criticism about this approach is the loss of information contained in the larger samples. A potential way of solving the problem is to apply extrapolation techniques to smaller samples, and produce an estimated species richness expected to occur if sample size were increased to the same size of the largest sample. We evaluated the reliability of 11 potential extrapolation methods over a range of different data sets and magnitudes of extrapolation. The basic approach adopted in the evaluation process was a comparison between the observed richness in a sample and the estimated richness produced by estimators using a sub-sample of the same sample. The Log-Series estimator was the most robust for the range of data sets and sub-sample sizes used, followed closely by Negative Binomial, SO-J1, Logarithmic, Stout and Vandermeer, and Weibull estimators. When applied to a set of independently replicated samples from a species-rich assemblage, 95% confidence intervals of estimates produced by the six best evaluated methods were comparable to those of observed richness in the samples. Performance of estimators tended to be better for species-rich data sets rather than for those which contained few species. Good estimates were found when extrapolating up to 1.8-2.0 times the size of the sample. We suggest that the use of the best evaluated methods within the range of indicated conditions provides a safe solution to the problem of losing information when standardizing different sample sizes to the size of the smallest sample.

A. S. Melo and G. Machado, Museu de História Natural, Inst. de Biologia, Univ. Estadual de Campinas, CP 6109, BR-13083-970 Campinas, SP, Brazil. Present address: Lab. Entomologia Aquática, Dept Biologia, FFCLRP, Univ. de São Paulo, Ribeirão Preto, SP, BR-14040-901, Brazil (grumicha@yahoo.com.br). – R. A. S. Pereira and G. J. Shepherd, Dept Botânica, Inst. de Biologia, Univ. Estadual de Campinas, CP 6109, BR-13083-970 Campinas, SP, Brazil. – A. J. Santos and R. J. Sawaya, Dept Zoologia, Inst. de Biologia, Univ. Estadual de Campinas, CP 6109, BR-13083-970 Campinas, SP, Brazil. – H. F. Medeiros, Dept Genética e Evolução, Inst. de Biologia, Univ. Estadual de Campinas, CP 6109, BR-13083-970 Campinas, SP, Brazil.

Accepted 9 July 2002

ISSN 0030-1299

A straightforward way to compare diversity among assemblages is the use of species richness. Advantages of using this metric are its great intuitive appeal, simple computation, and avoidance of pitfalls in choosing one among several diversity indices available (James and Rathbun 1981, Magurran 1988). It is well known, however, that species richness is strongly dependent on sample size. As more sample units or individuals are collected, more species are found (Walther et al. 1995, Condit et al. 1996, Gotelli and Colwell 2001). Thus, comparing assemblages using different sample sizes may produce erroneous conclusions (Stout and Vandermeer 1975).

Traditionally, there are two forms of standardization of sample sizes when comparing assemblages: the collection/observation of a given number of individuals or of sampling units (e.g. plots, transects or traps, Gotelli and Colwell 2001). Whether individuals or sampling units are better, is a currently disputed question (see discussion in Barbour and Gerritsen 1996, Courtemanch 1996, Vinson and Hawkins 1996) and in most cases researchers opt for the form most frequently used in their research fields. In studies using number of individuals as samples, it is possible to standardize different sample sizes by applying the rarefaction technique to the larger samples to obtain the expected richness in a sample of the same size of the smallest one (Sanders 1968, Hurlbert 1971, Simberloff 1979). Similarly, in studies using sampling units, it is possible to construct a species accumulation curve of the largest sample, remove the additional unit samples, and then record the species richness observed in the standardized sample size. The major criticism of these two approaches is the loss of information represented by the deleted additional individuals or sampling units in the largest samples (Williamson 1973, Magurran 1988, Elphick 1997).

A potential way to circumvent this loss of information is to use a richness estimate for the less sampled assemblages expected to occur if sample size were the same of the largest sample. This approach is not new (Gleason 1922, Evans et al. 1955), but has received little attention from ecologists over the last decades. Solow and Polasky (1999) presented an estimator to be used when sample size is expressed as number of individuals. Tackaberry et al. (1997) suggested a simple extrapolation technique using sampling units, but based on the knowledge of the physical location of each sampling unit. When large and structurally similar data sets are available, a promising extrapolation technique is presented by Plotkin et al. (2000). The technique uses a calibrated parameter obtained from a similar data set to extrapolate species richness in the data set under study.

A second potential situation where extrapolation of species richness to larger sample sizes might be useful is in the comparisons of assemblages differing in the proportion of rare species and heterogeneity. A species-poor assemblage, but with species distributed homogenously in the sampled area, might produce a species richness value higher than one observed in a species-rich assemblage with a large proportion of rare species and with patchy distribution in the area. In these cases, the species accumulation curves intersect and reliable comparisons can only be done using large sample sizes. An example of the problem is presented by Stout and Vandermeer (1975), who compared stream insect assemblages in tropical and temperate areas. Lande et al. (2000) suggest the use of the Simpson diversity index as a solution to ranking assemblages when using small sample sizes. Alternatively, if interest relies specifically in species richness, a second potential solution yet to be evaluated is to compare assemblages using extrapolated sample sizes. If a reliable extrapolation technique is available, one may choose to compare assemblages using for instance richness values estimated to occur if sample sizes were doubled.

Most of the potential methods that can be used to produce richness estimates for a defined sample size when using sampling units are extrapolations of species accumulation curves (Soberón and Llorente 1993). The parameters obtained in the adjusted equation using the species accumulation curve of the less sampled assemblage, are used to extrapolate to a larger sample size. Although simple, only a few models in a small number of papers have evaluated the closeness of estimates in relation to a priori known actual richness (Arrhenius 1923, Palmer 1990, Tackaberry et al. 1997, Keating et al. 1998).

Here we assessed the reliability of 10 currently available estimators plus one here described. We evaluated the accuracy, precision, and bias, and compared their performance in relation to the known actual richness in replicated data sets from two different assemblages. Additionally, in order to investigate the robustness of the evaluated estimators, we applied these 11 methods to estimate the known richness in six data sets using a range of sub-sample sizes. These data sets correspond to different taxa, assemblage structure, and were obtained through disparate sampling methods.

Keating et al. (1998) evaluated several extrapolation techniques to the analogous problem of effectiveness of further sampling in species inventories. They used data sets from beetles, vascular plants, and nine model communities with 10, 100, and 1000 species and high, medium, and low evenness. Here we expand the results of Keating et al. (1998) by (1) focusing on the specific problem of standardization of different sample sizes to compare species richness among assemblages, (2) including data from a large range of real assemblages, (3) including previously unevaluated estimators, and (4) using replicated data sets in order to compare the variability of estimates to the natural variability of observed richness among data sets derived from a same assemblage.

We were mainly concerned with the reliability of using such methods in practical situations. Thus, the approach used in the evaluation process was kept as practical and simple as possible, in order to allow a wide range of potential users to grasp and apply them in their ecological or conservation studies.

Methods

The estimators

The simplest way of estimating species richness in a sample is to count observed species in a sub-sample. Except for the cases in which the sub-sample has already included all species occurring in the sample, this estimate will be a negatively biased estimator of the richness in the sample and its accuracy will depend on the difference between the sample and the sub-sample sizes. We included the observed richness in a sub-sample as an estimate of the richness in the sample from which the sub-sample was drawn in order to examine how much the 11 evaluated estimators are able to improve the reduction in bias and increase in accuracy in relation to this simple estimate (Palmer 1990). From the 11 estimation methods evaluated in our study, seven were extrapolations of functions fitted to species accumulation curves. In these cases, we constructed a species accumulation curve for a sub-sample and fitted one of the models. The fitted parameters obtained using this sub-sample were then used to estimate the species richness in the total sample from which the sub-sample was drawn. Three of these estimators, the Logarithmic (Log), Exponential (Expo), and Clench models, are presented in Soberón and Llorente (1993) and differ from each other in the probabilities of adding new species as more sample units are collected. Stout and Vandermeer (1975) presented a model (hereafter SV) derived from the Island Biogeography theory and, like the Exponential and Clench models, it can be used to estimate total species richness of the species pool. In other words, the number of species expected to be obtained when sample size increases to infinite (Melo and Froehlich 2001a). The Weibull model has been used in several research fields, and was selected because of its good performance on fitting several species accumulation curves of bird data sets recorded from different human land use developments (Flather 1996). Two models traditionally used in the species-area literature were also evaluated, the LogLin (Gleason 1922, Palmer 1990) and the Power model (Arrhenius 1921, Flather 1996). Equations of each of these curve fitting models are given in Table 1.

The remaining four estimators evaluated are based on different rationales and computations. Evans et al. (1955) proposed a simple estimator (called here as ECB) obtained by solving a formula which takes into account only the number of sampling units collected and the respective number of species found,

$$S = \frac{s \log(N+1)}{\log(n+1)}$$

where S is the estimated species richness expected to occur in N unit samples and s is the number of species observed in n unit samples.

We also used the estimators Negative Binomial (NB) and Logarithmic Series (LS) that have recently been well evaluated by Keating et al. (1998) for a slightly different problem, the estimation of the effectiveness of further sampling in species inventories. Estimators NB and LS are given in Efron and Thisted (1976), following Fisher et al. (1943). They were originally based on the information of numbers of species that occurred with 1, 2, 3 etc individuals in a sample. However, in order to standardize all estimators in the study, we opted to use the information of numbers of species which occurred in 1, 2, 3 etc sampling units instead of those occurring with 1, 2, 3 etc individuals. The NB estimator is given by Efron and Thisted (1976) as,

$$\Delta_{\alpha\gamma}(t) = \frac{-\eta_1\{(1+\gamma t)^{-\alpha} - 1\}}{(\gamma\alpha)}$$

Table 1. Curve fitting models used to extrapolate species accumulation curves.

Name	Model	References
Logarithmic (Log)	$S = \frac{1}{z}\ln(1+zax)$	Soberón and Llorente (1993)
Exponential (Expo)	$S = ab(1-e^{-bx})$	Soberón and Llorente (1993)
Clench	$S = \frac{ax}{1+bx}$	Soberón and Llorente (1993)
Stout and Vandermeer (SV)	$S = \frac{a}{x^{-z}+\frac{a}{T_\infty}}$	Stout and Vandermeer (1975)
Weibull	$S = a\{1-e^{-[b(x-c)]^d}\}$	Flather (1996)
LogLinear	$S = a + b\log(x)$	Gleason (1922), Palmer (1990)
Power	$S = ax^b$	Arrhenius (1921), Flather (1996)

where $\Delta_{\alpha\gamma}(t)$ is the estimated number of species expected to be found in the additional sample size t, which is expressed as the proportion of the sample already collected. Thus, for the problem of estimating the number of species in a larger sample size, the estimated value is obtained by summing the number of species observed in the sample and $\Delta_{\alpha\gamma}(t)$. η_1 is the statistical expectation of the number of species occurring in only 1 sampling unit (or 1 word in the work of Efron and Thisted 1976) and is estimated here as the number of species occurring in only 1 unit sample in the sub-sample. The parameters α and γ are obtained by fitting a non-linear regression to the equation,

$$\eta_x = \frac{\eta_1\{\Gamma(x+\alpha)\}\gamma^{x-1}}{\{x!\Gamma(1+\alpha)\}}$$

where η_x is the number of species occurring in exactly x unit samples and γ represents the gamma function.

The LS estimator is obtained when we set $\alpha = 0$ and, following Efron and Thisted (1976), it is given as,

$$\Delta_{0\gamma}(t) = \left(\frac{\eta_1}{\gamma}\right)\log(1+\gamma^t)$$

The last evaluated estimator is an empirical method (A. S. Melo, unpubl.) first evaluated here that relates the number of sampling units necessary to collect a given number of species and the number of sampling units necessary to estimate the same given number of species using a non-parametric richness estimator, such as the first order Jackknife (see a review of non-parametric estimators in Colwell and Coddington 1994). Jackknife estimates of species richness were developed in order to predict the number of species occurring in a given area based on the number of observed species in a sample and the number of these species that were rare, i.e. that occurred in only 1, 2, 3 etc sampling units (Burnham and Overton 1978, Colwell and Coddington 1994). Previous studies have shown, however, that such estimates are dependent on sample size (Schmit et al. 1999, Melo and Froehlich 2001a), and in most cases this dependence is so strong that it can be useful as a predictive tool. Fig. 1 shows the relationship between the number of sampling units necessary to estimate a given species richness by the first order Jackknife and the number of sampling units necessary to observe the same number of species when constructing a species accumulation curve. Note that the points in Fig. 1 do not depict species richness. Species richness was used only to match the corresponding number of sampling units in which the same number of species can be obtained from the cumulative observed list and from estimates of the first order Jackknife. As an example, for the data set

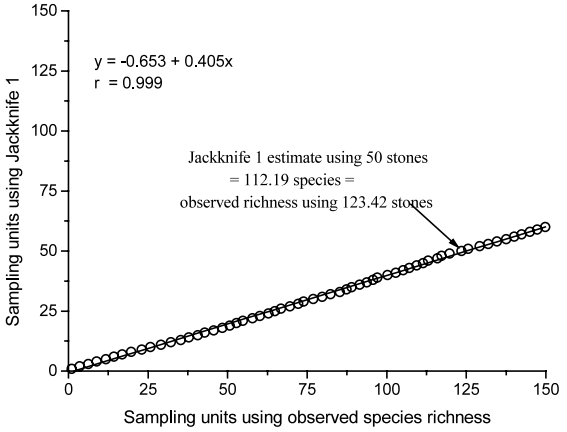

Fig. 1. Relationship between the number of sampling units necessary to observe a given number of species and the number of sampling units necessary to estimate the same species richness using the Jackknife 1 estimator. A precise match was obtained by using Jackknife 1 species richness estimates for each cumulative number of sampling units and the corresponding interpolated value needed to observe the same species richness in a species accumulation curve. Estimator SO-J1 is obtained by extrapolating the linear fitted model to a larger sample size and recording the corresponding species richness estimate produced by Jackknife 1 estimator. Data are from stream macroinvertebrates occurring in the 150 sampling units (stones) data set Pinda used in the replicated study.

used in Fig. 1, the first order Jackknife estimate using 50 sampling units is 112.19 species. When constructing a observed species accumulation curve, such richness value is obtained after 123–124 sampling units are collected. As the Jackknife estimates are continuous values, a precise match between the two axes was achieved by using the species richness estimated by the first order Jackknife for each cumulative number of sampling units (y-axis) and the corresponding interpolated x-value needed to observe each of such first order Jackknife estimates. For the above example, the first order Jackknife estimate 112.19 obtained using 50 sampling units can also be obtained from the observed species accumulation curve using 123.42 sampling units (Fig. 1). Thus, for our proposed method (hereafter SO-J1), (1) we constructed the relationship shown in Fig. 1, (2) fitted a linear model, (3) extrapolated the fitted linear model to the x-value corresponding to the sample size we intended to estimate, (4) and recorded the species richness estimated by the first order Jackknife using the corresponding number of sampling units obtained in (3). A limitation of the method is the inability to produce estimate values when the intended extrapolated sample size is much larger than the sample size of the available sample, usually more than 2-3 times. In these cases, the number of sampling units needed to be used by the first order Jackknife to estimate the extrapolated value (step 3 above) is larger than the number currently available in the sample in use.

Replicated study

In the first part of the study, we assessed accuracy, precision and bias of the 11 evaluated methods by comparing the richness estimate obtained using a sub-sample to the actual observed richness in the sample from which the subsample was obtained. The replicate samples were from two large and homogenous data sets of macroinvertebrates living on stones in stream (Melo and Froehlich 2001a) and fig wasps associated with fig fruits of *Ficus eximia* Schott (Moraceae) (Pereira et al. 2000).

The stream macroinvertebrates data set (here called Pinda) consisted of 10 339 individuals belonging to 117 morphospecies occurring in 150 sampling units (stones) collected from a stream reach in Pindamonhangaba, São Paulo state, Brazil (22° 45′S, 45° 28′W). We divided this data set randomly in 6 samples of 25 stones each. We opted for this sample size as it had been used with success in a previous study of diversity (Melo and Froehlich 2001b). From each 25-stone sample, we randomly selected two distinct subsamples of 12 stones each, which were used to estimate the richness in the 25-stone sample.

The fig wasp data set consisted of 13 582 individuals distributed in 13 species and 300 fig fruits collected in Londrina, Paraná state, Brazil (23° 18′S, 51° 09′W). To determine a meaningful sample size for the fig wasp data set, we constructed a species accumulation curve to obtain the minimum sample size in which we could have a good representation of the fig wasp assemblage occurring in *Ficus eximia* fig fruits. After this visual analysis, we concluded that 25 fig fruits was a good sample size; thus the 300 fig fruit data set gave us 12 distinct samples. Following the macroinvertebrate samples, we opted for using a sub-sample of 12 fig fruits. However, different from the macroinvertebrates samples, only one sub-sample of 12 units was drawn from each sample.

We used the software EstimateS version 5.0.1 (Colwell 1997) to construct species accumulation curves (100 runs) for all 24 sub-samples from the stream macroinvertebrates and fig wasp data sets. Curves were fit using the software Origin version 4.1 (Microcal Software, Northampton, MA, USA). Parameters γ and α used in NB and LS estimators were obtained by writing a specific routine in S-Plus 2000 software (MathSoft, Inc., Cambridge, Ma, USA).

Unreplicated study

In the second part of the study, we assessed the robustness of the different estimators over a range of different sub-sample sizes and assemblage structures. We used data from six different diversity studies on spiders, trees, *Drosophila* spp., stream macroinvertebrates (Melo and Froehlich 2001a), litter harvestmen, and litter frogs. The six assemblages were very different from each other and comprised representatives from different taxa and habitats, and were obtained through distinct collection methods (Table 2).

For each assemblage, we randomly selected four sets of 30 sub-samples, representing sub-samples of sizes 40, 55, 70, and 85 percent of the total number of sample units. Each sub-sample was then used to estimate the species richness in the total sample from which it was derived. Contrary to the replicated study, these sub-samples obviously were not distinct to each other.

Estimates for the seven curve fitting methods were obtained by constructing a species accumulation curve (100 runs) for each sub-sample and fitting each function using non-linear regression. The random draw of the sub-samples, construction of the species accumulation curves, and fit of the seven functions were done by using a routine written in S-Plus 2000. We had difficulties in fitting the Weibull model to some sub-sample sizes in some data sets, because the parameter a which denotes the asymptote of the fitted equation tended to increase indefinitely. For these cases, we fitted each sub-sample individually in Origin software setting the parameter a as 500. Parameters γ and α used in NB and LS estimators and the computation of the SO-J1 method were obtained by writing specific routines in S-Plus 2000 software (all S-Plus routines used in this work are available on request from the first and second authors).

Evaluation of estimation methods

In order to make results comparable among all assemblage data, we used the percentage of error in relation to the actual richness, calculated as the difference between the estimated and actual richness in the total sample, divided by the actual richness in the total sample.

For the replicated study, we explored the results produced by the different estimators in two ways. In the first, we plotted the estimated values using 12 sampling units and the known actual richness in 25-sampling unit samples. This plot provides a simple way of visually assessing accuracy, precision, and bias of the estimates in relation to the actual richness in the respectively sample from which it was drawn. Moreover, the plot shows to what extent the estimated values are correlated to the actual richness values.

We also compared the variability of the estimated values with the variability of the actual richness values observed among the 25-sampling unit samples. This was achieved by calculating percentage of errors for each estimation method and also for the values of richness observed in each 25-sampling unit sample. To compute the percentage of errors we used as actual

Table 2. Summary of the six data sets used in the unreplicated study. All localities in Brazil.

	Spiders	Trees	Stream macroinvertebrates	*Drosophila* spp.	Harvestmen	Frogs
Locality	Linhares, Espirito Santo	Campinas, São Paulo	Jundiai, São Paulo	Barreiro Rico, São Paulo	Ubatuba, São Paulo	Ilha de São Sebastião, São Paulo
Geographical coordinates	19° 10′S, 40° 05′W	22° 49′S, 47° 07′W	23° 14′S, 46° 56′W	22° 40′S, 48° 10′W	23° 26′S, 45° 04′W	23° 47′S, 45° 24′W
Vegetation	Atlantic rain forest	Tropical semi-deciduous forest	Tropical semi-deciduous montane forest	Tropical semi-deciduous forest	Atlantic rain forest	Atlantic rain forest
Sampling units	time intervals	10×10 m contiguous plots	single stones (15–20 cm diameter) in stream riffles	traps using fermented bananas	8×8 m plots on litter	8×8 m plots on litter
Sample size	243	100	75	180	63	92
Species richness	287	101	66	57	40	15
Individuals	1982	1465	3759	8166	764	846

richness the mean richness value observed in 25-sampling unit samples, obtained from a species accumulation curve constructed using the entire data set (150 stones or 300 fig fruits). Accuracy and precision of all estimation methods and also of the observed richness in the 25-sampling unit samples were measured as the mean and the standard deviation of the percentage of error values. Bias was measured as the percentage of values that overestimated actual richness minus 50%. A good estimator method should produce values close to zero for accuracy, precision, and bias.

Results

Replicated study

Estimates for Pinda samples were positively correlated with the actual richness for all evaluated methods, and except for the LogLin estimator, all other methods produced estimates more correlated with the actual richness than the observed species richness in 12 sampling units (Fig. 2). However, correlations between estimates and the actual richness for fig wasps were variable, and even a strongly negative correlation was observed for the NB estimator (Fig. 3). Moreover, estimate values for fig samples were much more variable through all methods than for Pinda samples (Fig. 2 and 3).

Despite the differences in correlation between estimated and observed richness in the two data sets, there were agreements in bias for some estimation methods. Methods that overestimated or underestimated the actual richness in Pinda samples, in general also overestimated or underestimated richness in fig wasp samples (Table 3). As would be expected, the observed richness in 12 sampling units (SO-12) strongly underestimated the actual richness in 25-sampling unit samples. For fig wasps, however, the observed richness in 12 fig fruits (SO-12) produced values more accurate than in Pinda samples. Moreover, in three cases the observed richness in 12 fig fruits were the same as those observed in the 25-fig fruit samples. The Expo model produced estimates very similar to the observed richness in 12 sampling units and in all cases underestimated the actual richness. Following the Expo model, the Clench model also underestimated the actual richness in all but five cases of fig wasps. The Power model produced overestimates for all Pinda samples and 9 of the 12 samples of fig wasps. The ECB estimator also overestimated the actual richness in the two data sets, although with higher accuracy and precision for Pinda than for fig wasp sample. The LogLin estimator tended to underestimate the actual richness in Pinda samples, but was slightly positively biased for fig wasp samples. The estimators Log, SV, Weibull, ECB, SO-J1, NB, and LS produced very similar estimates to each other for Pinda

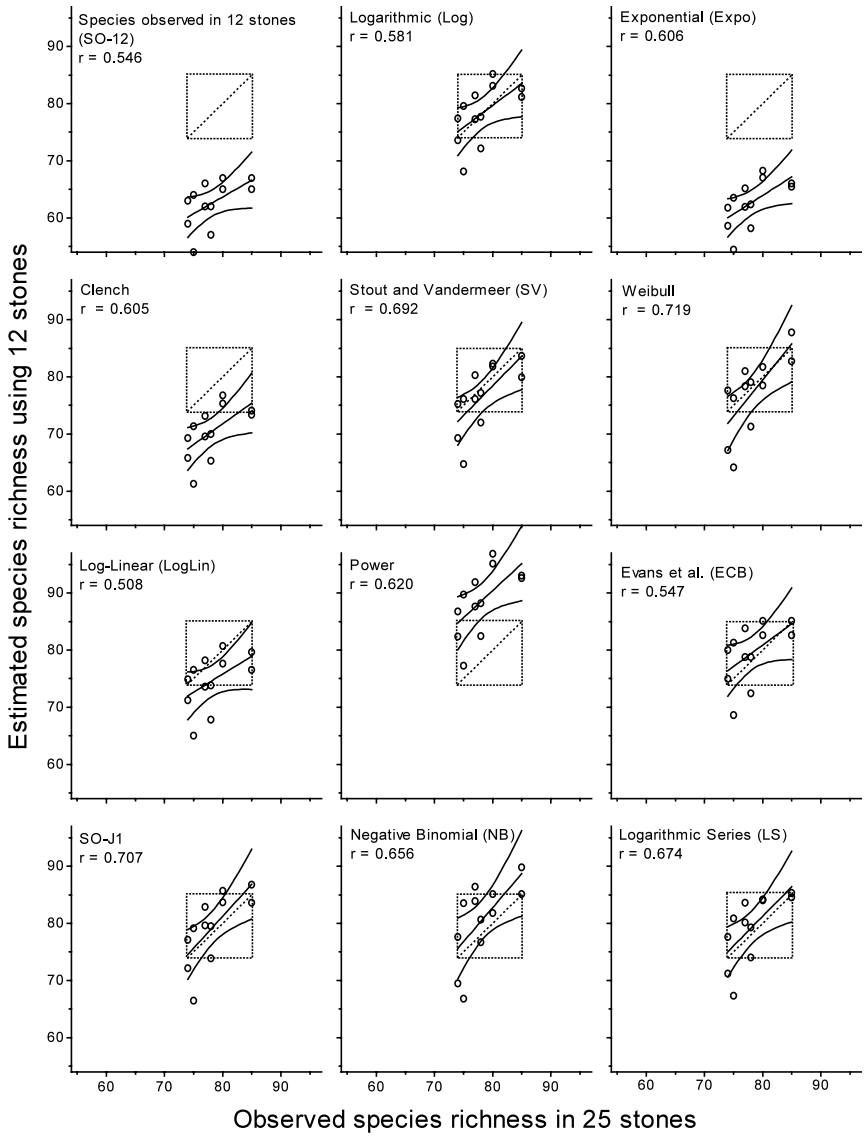

Fig. 2. Relationship between stream macroinvertebrates species richness (Pinda data set) observed in 25 sampling units (stones) and the estimated species richness by 12 methods using subsamples of 12 stones. Two distinct random subsamples were taken from each of six 25-stone samples. The dotted square indicates the range of the six observed richness values in 25-stone samples and the dotted diagonal line the expected estimated richness if estimators produce the same value of observed species richness. Solid lines represent the linear fit to data and 95% confidence limits.

samples and in all cases the 95% confidence limit included the expected actual richness (depicted in Fig. 2 as the diagonal dotted line inside the boxes). For fig wasps, only the estimators Log, LogLin and LS produced 95% confidence limits that included all the expected actual richness range, although the confidence limit lines for methods SV, Weibull, and SO-J1 crossed the expected actual richness (diagonal dotted lines) close to the tails (Fig. 3).

The mean percentage error of each estimation method in relation to the average observed richness in 25 sampling units, obtained from a species accumulation curve using all sampling units in each data set, were consistent with the results described above (Fig. 4). For Pinda samples, the mean percentage of error of estimators Log, SV, Weibull, LogLin, ECB, SO-J1, NB, and LS were close to zero and in all cases the 95% confidence interval (CI) included the zero value (Fig. 4). Moreover, the mean error value and the 95% CI of these estimators were very similar to that produced by the observed richness in the six 25-stone samples (SO-25, Fig. 4). The same estimators that produced mean errors close to zero for Pinda samples, also produced mean errors close to zero for fig wasp samples, except for the methods ECB and NB. The estimator LogLin, that produced a negative mean error for Pinda samples, produced a slightly positive error for fig wasp samples, and the absolute error value was comparable to the other well-evaluated estimators for this data set. However, despite the similarities between the two data sets in rank performance in relation to mean error, the 95% CI of estimates for fig wasp samples were in general

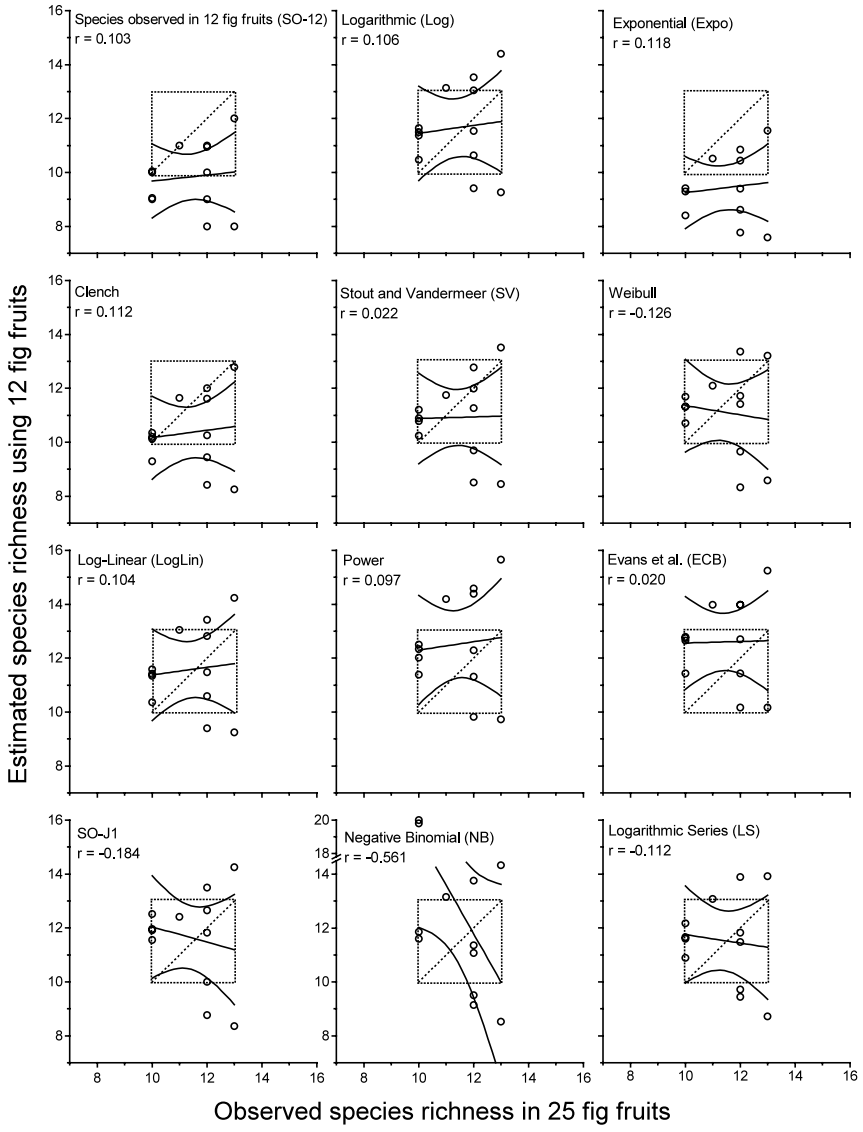

Fig. 3. Relationship between fig wasp species richness observed in 25 sampling units (fig fruits) and the estimated species richness by 12 methods using subsamples of 12 fig fruits. One sub-sample was taken randomly from each of 12 25-fig fruit samples. The dotted square indicates the range of the 12 observed richness values in 25-fig fruit samples and the dotted diagonal line the expected estimated richness if estimators produce the same value of observed species richness. Solid lines represent the linear fit to data and 95% confidence limits.

much wider than in Pinda samples. Furthermore, contrary to the results obtained from Pinda samples, the 95% CI of estimates for fig wasp samples were much wider in relation to that produced by the observed richness in the 12 25-fig fruit samples (Fig. 4).

Unreplicated study

As in the replicate study, the Expo and Clench estimators underestimated the actual richness using all four sub-sample sizes and also in all six data sets (Fig. 5). The Power model tended to overestimate the actual richness, except in the *Drosophila* spp. and frog samples. The LogLin model that yielded underestimates in Pinda samples and slight overestimates in fig wasp samples, underestimated the total richness in all data sets. The ECB estimator presented low accuracy and did not yield a consistent bias for all six data sets. The SV and Weibull estimators tended to produce negative errors, but of low magnitude. The NB and LS estimators, followed by the Log model, were robust for different data sets and presented low negative and positive errors. The estimator SO-J1 produced results comparable to NB and LS estimators, but failed to produce values for the 40% sub-sample size.

Except for the ECB method, which showed no clear trend over the increasing sub-sample sizes, all the remaining estimators increased in accuracy when using larger sub-sample sizes. However, this increase was not conspicuous for NB and LS methods, which produced low errors even when using the 40% sub-sample size (Fig. 5). At least for the overall good estimators NB, LS, SO-J1, Log, SV, and Weibull provided better

Table 3. Performance of estimators in the replicated and in the unreplicated study. Mean and standard deviation (SD) measure respectively accuracy and precision of estimators. Bias is expressed as percentage of overestimates minus 50%. A good estimator should produce mean, standard deviation, and bias close to zero.

Estimators	Replicated study						Unreplicated study		
	Pinda			Fig wasps			Overall accuracy	Robustness to data sets	Robustness to subsample size
	Mean	SD	Bias	Mean	SD	Bias			
SO in subsample	−18.93	5.28	−50	−12.01	11.00	−41.7	Bad	Good	Bad
Logarithmic (Log)	1.40	6.47	25	3.45	14.30	16.7	Excellent	Good	Excellent
Exponential (Expo)	−18.74	5.27	−50	−16.35	10.91	−41.7	Bad	Good	Bad
Clench	−8.76	5.90	−50	−8.03	12.59	−16.7	Bad	Bad	Bad
Stout and Vandermeer (SV)	−0.85	7.34	−8.3	−3.09	13.66	−16.7	Good	Good	Good
Weibull	−0.09	8.65	16.7	−1.37	14.09	16.7	Good	Good	Good
Log-Linear (Log-Lin)	−3.33	6.10	−16.7	2.72	13.88	16.7	Good	Good	Good
Power	14.86	7.50	50	11.07	16.58	33.3	Bad	Bad	Bad
Evans et al. (ECB)	2.97	6.71	25	11.77	13.98	33.3	Bad	Bad	Good
SO-J1	2.60	7.77	16.7	3.30	15.74	25	Excellent	Good	Excellent
Negative Binomial (NB)	3.05	6.88	25	13.97	33.21	16.7	Excellent	Excellent	Excellent
Logarithmic Series (LS)	1.50	5.63	16.7	2.31	14.69	16.7	Excellent	Excellent	Excellent

accuracy and precision when using species-rich than when using the species-poor frog assemblage.

Discussion

Expo, Clench, and LogLin models tended to underestimate the actual richness, while the Power model tended to overestimate it. The ECB estimator produced biased and inaccurate estimates depending on the data set used. In some data sets, the ECB produced consistently negatively biased estimates, while in others the method produced positively biased estimates. The SV and Weibull models performed well in the replicated study, but slightly underestimated the actual richness in the unreplicated study. The NB estimator was in general unbiased and very accurate, except in the species-poor fig wasp samples. The Log, SO-J1 and LS estimators also performed well in the replicated study, but the Log model was less accurate than SO-J1 and LS methods in the unreplicated study.

Our results closely agree with the findings of Keating et al. (1998), who evaluated the bias of estimators Log, Expo, Clench, LogLin, Power, NB, and LS for the analogous problem of effectiveness of further sampling in species inventories. The Expo and Clench models that underestimated actual richness here, were also shown to be negatively biased by Keating et al. (1998). The generally biased LogLin and Power estimators, but that produced good results in some data sets (LogLin with fig wasps and Power with *Drosophila* spp. and frogs), also tended to be biased in Keating et al.'s study, but again produced good unbiased results in some particular data sets. Keating et al. reported biased results for the Log model, although the bias sign was dependent on the data sets evaluated. Also agreeing with our results, LS and NB estimators were well evaluated by Keating et al. and those authors suggested the use of NB estimator as the most robust and generally unbiased estimator. In our study, performance of estimators LS and NB was nearly identical. However, the NB estimator produced some disparate values for the fig wasp samples (Fig. 3), indicating that this esti-

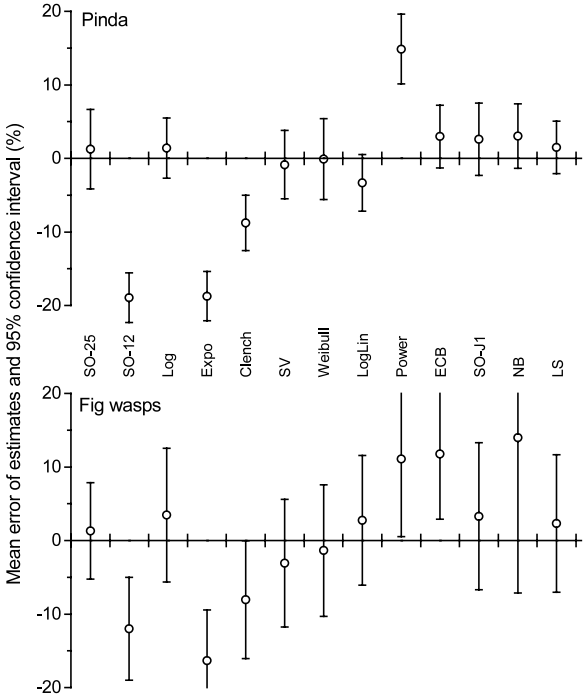

Fig. 4. Error of estimates produced by the 12 methods evaluated and the observed richness in 25 sampling units for Pinda and fig fruit samples. Errors were computed as the difference between estimated and average observed richness, divided by average observed richness. Average observed richness for Pinda samples was 77.2 and for fig wasps 11.04 species and were obtained from species accumulation curves using the entire data sets (150 stones or 300 fig fruits). See Table 3 for abbreviations of estimators.

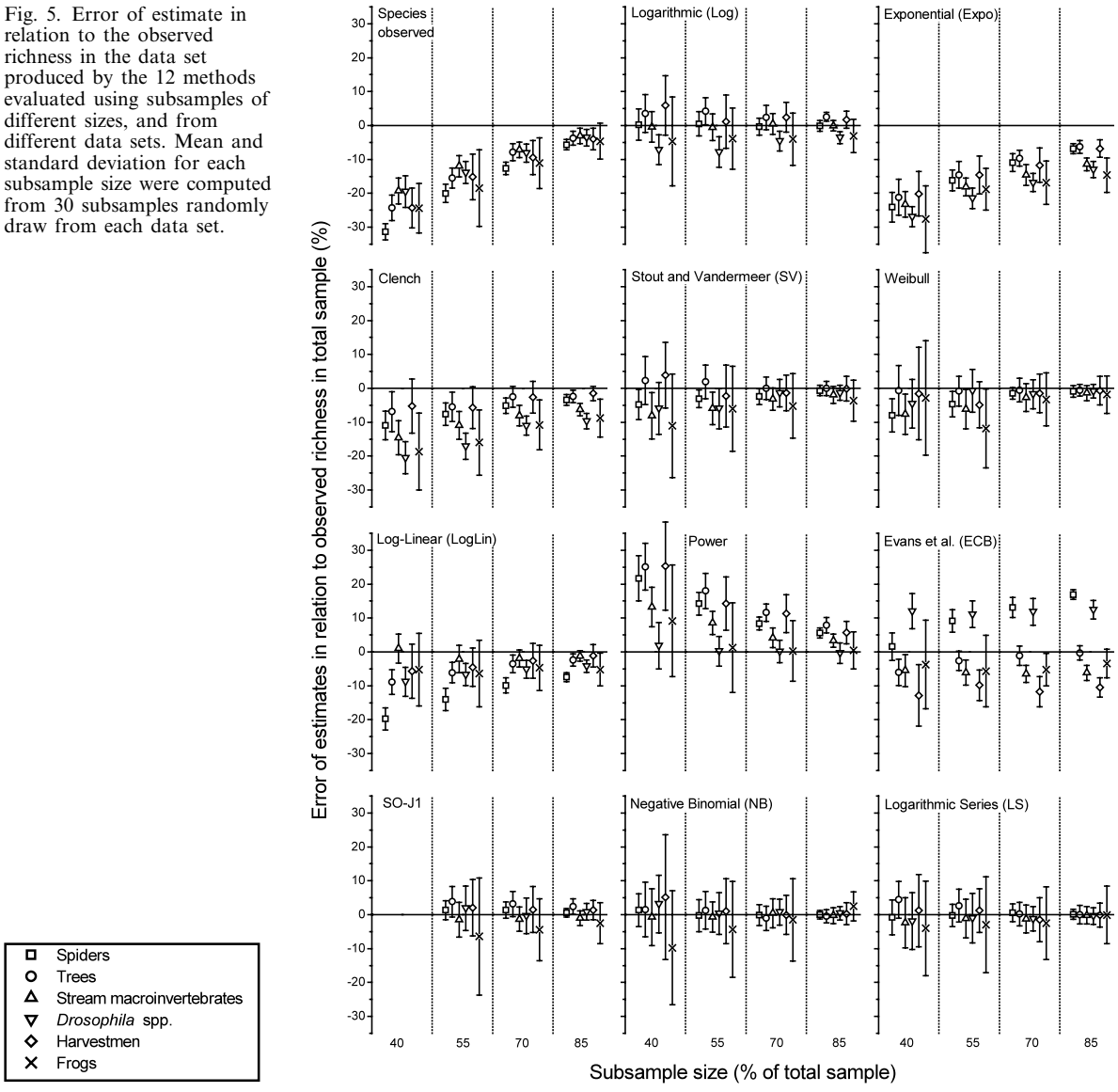

Fig. 5. Error of estimate in relation to the observed richness in the data set produced by the 12 methods evaluated using subsamples of different sizes, and from different data sets. Mean and standard deviation for each subsample size were computed from 30 subsamples randomly draw from each data set.

mator might be less precise than the LS estimator for species-poor samples.

Poor performance of Clench model is further supported by Keating and Quinn (1998). Palmer (1990) and Tackaberry et al. (1997) evaluated the models LogLin and a method analogous to the Power model in predicting plant species richness. They found that the Power model overestimated the actual richness known to occur in study plots. LogLin overestimated actual richness in most samples of Palmer's (1990) study, but tended to underestimate actual richness in the study of Tackaberry et al. (1997). In contrast to the result presented by Palmer (1990) and in accordance to Tackaberry et al. (1997), the LogLin model tended to underestimate actual richness in our study, suggesting that its performance is dependent on data set in study. Models Expo, Clench, LogLin, and Power have been used to fit species accumulation curves in a number of studies (Condit et al. 1996, León-Cortés et al. 1998, Moreno and Halffter 2000). However, despite good fit of these models to species accumulation curves in these studies, and consequently the usefulness of the fitted parameters, we do not recommend the use of these models for predicting extrapolated species richness.

At least for the better-performing estimation methods, NB, LS, SO-J1, Log, SV, and Weibull (Table 3), estimators produced better results in species-rich than in species-poor data sets. Clear evidence of this can be seen comparing results of Pinda (74–85 species) versus fig wasp samples (10–13 species) in the replicated study and the 5 species-rich data sets (40–287 species) versus the species-poor frog data set (15 species) in the un-

replicated study. When calculating estimates for fig wasp samples, we noted that this bad performance is due to the high heterogeneity among sub-samples of species-poor data sets. Given the low number of species in fig wasp sub-samples, the inclusion or exclusion of 1-2 rare species due to chance can produce very different estimates. In the unreplicated study, this is evident from the relatively large error bars of frog sub-samples in Fig. 5. Bad performance of estimators in species-poor assemblages was also found by Keating et al. (1998), who reported biased estimates in their medium-evenness and low-richness (random-fraction, 10 species) model communities.

As would be expected, there was an increase in accuracy and precision of estimates in the unreplicated study with increased sub-sample sizes. An exception was the ECB estimator, that produced even worse estimates on increased sub-sample sizes for the spider data set. For estimators NB and LS, good accuracy were obtained for species-rich data sets even when using sub-samples of only 40% of the total sample size, although with a low precision, shown by the large error bars. Using sub-samples of 55%, both accuracy and precision were considerably improved for these two estimators, except in the frog data set. In the replicated study, sub-samples of 48% (12 from 25 sampling units) produced good results when using the six best estimators (NB, LS, SO-J1 Log, SV, and Weibull), at least for species-rich Pinda samples. For these data, 8–10 of 12 estimated values were included in the range of the observed richness in 25 stones (the dotted squares in Fig. 2). Thus, we suggest that at least for species-rich data sets (>30–40 species), extrapolations are safe up to sample sizes 1.8–2.0 times the size of the sample in study.

Although we found that NB and LS, followed by SO-J1, Log, SV, and Weibull estimators, were in general accurate and not strongly biased (Table 3), this finding was based on averages of several sub-samples. An estimator can be very accurate when averaging several estimates, despite a high variance among estimate values (low precision). In actual use, the researcher commonly will have only one sample, and it is of interest to know how reliable a single estimate produced by an estimation method is. On the other hand, we should recall that an observed richness value in a given sample is only an estimate of the mean actual richness value for that particular sample size in the assemblage under study. Reliability of an estimation method can be assessed comparing the variation of its estimates with the variation of the observed richness values among several independent samples collected from the assemblage in study. From Fig. 4 we can observe that the 95% CI for the 12 estimate errors produced by NB, LS, SO-J1, Log, SV, and Weibull estimators are slightly higher but comparable to those produced by the six observed richness values in Pinda samples. However, for fig wasp samples, estimates of LS, SO-J1, Log, SV, and Weibull methods were reasonably more variable than the observed richness in 12 25-fig fruit samples. Furthermore, the NB method produced a CI three times higher than that produced by the observed richness in the 12 25-fig fruit samples (Fig. 4), due to two outlier values (Fig. 3). As commented earlier, more reliable estimates are likely to be obtained when working with species-rich samples.

Three issues should be considered when choosing and using species richness estimators. The first is the ease of computation. Estimators Log, SV, and Weibull are extrapolations of species accumulation curves. The construction of such curves is easily done using available free software (e.g. EstimateS, Colwell 1997). The nonlinear fitting of curves is available in most of the statistical and graphical softwares available. The methods NB and LS in other hand are based on fitting a model to the number of species occurring in 1, 2, 3 etc sampling units. Also, the model includes in its computations the gamma function, which might not be available in some statistical softwares. A further advantage of using the models Log, SV, and Weibull is the interpretability of the fitted parameter, which might be useful in diversity studies (see respectively, Stout and Vandermeer 1975, Soberón and Llorente 1993, and Flather 1996). The estimator SO-J1 is not usually able to extrapolate to sample sizes larger than twice the sample in a study. For these cases, when extrapolating the linear function in Fig. 1, the obtained y-value is higher than the number of sampling units in the sample. A second issue specific to the estimator SO-J1 is the assumption of the linear relationship depicted in Fig. 1. The relationship is likely to be found in assemblages containing a high proportion of rare species and or when sample size is small. For assemblages well sampled and/or with a few rare species, the Jackknife 1 estimates will attain an asymptote and the consequent linear relationship in Fig. 1 will not be found, thus invalidating the method SO-J1. The last issue to be considered is that as with any statistical estimation method, extrapolation assumes that additional sample sizes come from the same universe from which current samples were collected. If samples were collected at random locations inside a 10 ha plot, extrapolated richness values will be valid only for the same 10 ha area.

Recently, Walther and Martin (2001) suggested the use of methods that estimate species richness in the area as a way of standardizing different sample sizes (see review of these methods in Colwell and Coddington 1994). In this case, comparisons would be made using the estimated richness expected to occur in the total area under study. A practical problem of this approach is the strong dependence of richness estimates produced by these methods on the observed richness. In fact, the well-behaved estimator SO-J1 is based on this dependence (Fig. 1). This dependence is very strong until

sample size is increased enough to collect most of the species in the study area (Colwell and Coddington 1994, Melo and Froehlich 2001a, A. S. Melo unpubl.). While this can be feasible in some species-poor assemblages as the one used by Walther and Martin (2001), it is not usually feasible for species-rich assemblages (Schmit et al. 1999, Gotelli and Colwell 2001, Melo and Froehlich 2001a).

We argue that a good estimator must produce reliable estimates independent of the structure of the data being used. Methods that perform very well with some data structures but not with others should be avoided, as a researcher does not know, a priori, which one is the best for a specific problem. As a first option, we suggest the use of LS estimator, followed by any one of the NB, SO-J1, Log, SV, or Weibull methods. Performance of LS method was consistently good with the replicated and the unreplicated studies and the several data sets used (Table 3). Despite the generally unbiased and accurate average estimates produced by these six estimators throughout the range of situations evaluated, precision was too low in species-poor assemblages (less than 15–20 species) and for extrapolations greater than 1.8–2.0 times the sample in study. The first restriction should not be a major problem in most diversity studies as interest often centers on species-rich assemblages, which can potentially reflect fine environmental differences (Kremen 1992, Brown and Freitas 2000). The restriction on magnitude of extrapolation should be enough for most of the studies where different sample sizes are caused by loss of samples, destruction of traps by animals or bad weather, and shortage of time or money. The great robustness of the six methods suggested to different data structures provides a safe solution to the problem of losing information by standardizing different samples sizes to the size of the smallest sample.

Acknowledgements – We thank all people who assisted us in field work. Part of the analyses and writing of the manuscript was done when the first author was visiting the laboratory of Prof. Colin Townsend, Dep. of Zoology, University of Otago. Robert K. Colwell, Claudio G. Froehlich and Thomas M. Lewinsohn provided useful suggestions that improved the manuscript. ASM received logistical support from Claudio G. Froehlich and a scholarship from Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) (proc. no. 98/00590-0). This work is part of the BIOTA/FAPESP-The Biodiversity Virtual Institute Program (www.biota.org.br).

References

Arrhenius, O. 1921. Species and area. – J. Ecol. 9: 95–99.
Arrhenius, O. 1923. Statistical investigations in the constitution of plant associations. – Ecology 4: 68–73.
Barbour, M. T. and Gerritsen, J. 1996. Subsampling of benthic samples: a defense of the fixed-count method. – J. North Am. Benthol. Soc. 15: 386–391.
Brown, K. S. Jr. and Freitas, A. V. L. 2000. Atlantic forest butterflies: indicators for landscape conservation. – Biotropica 32: 934–956.
Burnham, K. P. and Overton, W. S. 1978. Estimation of the size of a closed population when capture probabilities vary among animals. – Biometrika 65: 625–633.
Colwell, R. K. 1997. EstimateS: statistical estimation of species richness and shared species from samples. Version 5. User's Guide and application published at: http://viceroy.eeb.uconn.edu/estimates.
Colwell, R. K. and Coddington, J. A. 1994. Estimating terrestrial biodiversity through extrapolation. – Philos. Trans. R. Soc. Lond. B 345: 101–118.
Condit, R., Hubbell, S. P., Lafrankie, J. V. et al. 1996. Species-area and species-individual relationships for tropical trees: a comparison of three 50-ha plots. – J. Ecol. 84: 549–562.
Courtemanch, D. L. 1996. Commentary on the subsampling procedures used for rapid bioassessments. – J. North Am. Benthol. Soc. 15: 381–385.
Efron, B. and Thisted, R. 1976. Estimating the number of unseen species: how many words did Shakespeare know? – Biometrika 63: 435–447.
Elphick, C. S. 1997. Correcting avian richness estimates for unequal sample effort in atlas studies. – Ibis 139: 189–190.
Evans, F. C., Clark, P. J. and Brand, R. H. 1955. Estimation of the number of species present on a given area. – Ecology 36: 342–343.
Fisher, R. A., Corbet, A. S. and Williams, C. B. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. – J. Anim. Ecol. 12: 42–58.
Flather, C. H. 1996. Fitting species-accumulation functions and assessing regional land use impacts on avian diversity. – J. Biogeogr. 23: 155–168.
Gleason, H. A. 1922. On the relation between species and area. – Ecology 3: 158–162.
Gotelli, N. J. and Colwell, R. K. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. – Ecol. Lett. 4: 379–391.
Hurlbert, S. H. 1971. The nonconcept of species diversity: a critique and alternative parameters. – Ecology 52: 577–586.
James, F. C. and Rathbun, S. 1981. Rarefaction, relative abundance, and diversity of avian communities. – Auk 98: 785–800.
Keating, K. A. and Quinn, J. F. 1998. Estimating species richness: the Michaelis-Menten model revisited. – Oikos 81: 411–416.
Keating, K. A., Quinn, J. F., Ivie, M. A. and Ivie, L. 1998. Estimating the effectiveness of further sampling in species inventories. – Ecol. Appl. 8: 1239–1249.
Kremen, C. 1992. Assessing the indicator properties of species assemblages for natural areas monitoring. – Ecol. Appl. 2: 203–217.
Lande, R., DeVries, P. J. and Walla, T. R. 2000. When species accumulation curves intersect: implications for ranking diversity using small samples. – Oikos 89: 601–605.
León-Cortés, J. L., Soberón-Mainero, J. and Llorente-Bousquets, J. 1998. Assessing completeness of Mexican sphinx moth inventories through species accumulation functions. – Div. Distrib. 4: 37–44.
Magurran, A. E. 1988. Ecological diversity and its measurement. – Croom Helm.
Melo, A. S. and Froehlich, C. G. 2001a. Evaluation of methods for estimating macroinvertebrates species richness using individual stones in tropical streams. – Freshwat. Biol. 46: 711–721.
Melo, A. S. and Froehlich, C. G. 2001b. Macroinvertebrates in neotropical streams: richness patterns along a catchment and assemblage structure between 2 seasons. – J. North Am. Benthol. Soc. 20: 1–16.
Moreno, C. E. and Halffter, G. 2000. Assessing the completeness of bat biodiversity inventories using species accumulation curves. – J. Appl. Ecol. 37: 149–158.
Palmer, M. W. 1990. The estimation of species richness by extrapolation. – Ecology 71: 1195–1198.

Pereira, R. A. S., Semir, J. and Menezes-Jr., A. O. 2000. Pollination and other biotic interactions in figs of *Ficus eximia* Schott (Moraceae). – Rev. Bras. Bot. 23: 217–224.
Plotkin, J. B., Potts, M. D., Yu, D. W. et al. 2000. Predicting species diversity in tropical forests. – Proc. Natl. Acad. Sci. USA 97: 10850–10854.
Sanders, H. L. 1968. Marine benthic diversity: a comparative study. – Am. Nat. 102: 243–282.
Schmit, J. P., Murphy, J. F. and Mueller, G. M. 1999. Macrofungal diversity of a temperate oak forest: a test of species richness estimators. – Can. J. Bot. 77: 1014–1027.
Simberloff, D. 1979. Rarefaction as a distribution-free method of expressing and estimating diversity. – In: Grassle, J. F., Patil, G.P., Smith, W. and Taillie, C. (eds), Ecological diversity in theory and practice. International Co-operative Publishing House, pp. 159–176.
Soberón, J. and Llorente, J. 1993. The use of species accumulation functions for the prediction of species richness. – Conserv. Biol. 7: 480–488.
Solow, A. R. and Polasky, S. 1999. A quick estimator for taxonomic surveys. – Ecology 80: 2799–2803.
Stout, J. and Vandermeer, J. 1975. Comparison of species richness for stream-inhabiting insects in tropical and mid-latitude streams. – Am. Nat. 109: 263–280.
Tackaberry, R., Brokaw, N., Kellman, M. and Mallory, E. 1997. Estimating species richness in tropical forest: the missing species richness technique. – J. Trop. Ecol. 13: 449–458.
Vinson, M. R. and Hawkins, C. P. 1996. Effects of sampling area and subsampling procedure on comparisons of taxa richness among streams. – J. North Am. Benthol. Soc. 15: 392–399.
Walther, B. A. and Martin, J.-L. 2001. Species richness estimation of bird communities: how to control for sampling effort? – Ibis 143: 413–419.
Walther, B. A., Cotgreave, P., Price, R. D. et al. 1995. Sampling effort and parasite species richness. – Parasitol. Today 11: 306–310.
Williamson, M. 1973. Species diversity in ecological communities. – In: Bartlett, M. S. and Hiorns, R. W. (eds), The mathematical theory of the dynamics of biological populations. Academic Press, pp. 325–335.

Errata to:

Melo, A.S., Pereira, R.A.S., Santos, A.J., Shepherd, G.J., Machado, G., Medeiros, H.F. and Sawaya, R.J. 2003. Comparing species richness among assemblages using sample units: why not use extrapolation methods to standardize different sample sizes? – Oikos 101: 398-410.

1) page 400, Table 1. Correct form for Exponential (expo) function:

$S = a/b(1 - e^{-bx})$

2) page 401, second formula. Correct form:

$\Delta_{0\gamma}(t) = (\eta_1/\gamma) \log(1 + \gamma t)$

Results presented in the article were obtained using the correct formulae.

We apologize for any inconvenience this may cause.

The authors.

April 2003