Freshwater Biology (2001) **46**, 711–721

Evaluation of methods for estimating macroinvertebrate species richness using individual stones in tropical streams

ADRIANO S. MELO* and CLAUDIO G. FROEHLICH†
**Programa de Pós-Graduação em Ecologia, Museu de História Natural, IB, Universidade Estadual de Campinas, Campinas, SP 13083-970, Brazil*
†Departamento de Biologia, FFCLRP, Universidade de São Paulo, Ribeirão Preto, SP 14040-901, Brazil

SUMMARY

1. The most straightforward way to assess diversity in a site is the species count. However, a relatively large sample is needed for a reliable result because of the presence of many rare species in rich assemblages. The use of richness estimation methods is suggested by many authors as a solution for this problem in many cases.
2. We examined the performance of 13 methods for estimating richness of stream macroinvertebrates inhabiting riffles both at local (stream) and regional (catchment) scales. The evaluation was based on (1) the smallest sub-sample size needed to estimate total richness in the sample, (2) constancy of this size, (3) lack of erratic behaviour in curve shape and (4) similarity in curve shape through different data sets. Samples were from three single stream sites (local) and three from several streams within the same catchment basin (regional). All collections were made from protected forest areas in south-east Brazil.
3. All estimation methods were dependent on sub-sample size, producing higher estimates when using larger sub-sample sizes. The Stout and Vandermeer method estimated total richness in the samples with the smallest sub-sample size, but showed some erratic behaviour at small sub-sample sizes, and the estimated curves were not similar among the six samples. The Bootstrap method was the best estimator in relation to constancy of sub-sample sizes, but needed an unacceptably large sub-sample to estimate total richness in the samples. The second order Jackknife method was the second best estimator both for minimum sub-sample size and constancy of this size and we suggest its use in future studies of diversity in tropical streams. Despite the inferior performance of several other methods, some produced acceptable results. Comments are made on the utility of using these estimators for predicting species richness in an area and for comparative purposes in diversity studies.

Keywords: accumulation curves, jackknife, macroinvertebrates, species richness, tropical streams

Introduction

An old and common problem when studying species assemblages is to know how many different kinds of organisms there are in the study area. Two situations where this information is very important are (a) setting priorities in choosing conservation areas and, in a more general sense and (b) when comparing species assemblages from contrasting situations (Magurran, 1988; Colwell & Coddington, 1994). In the first case, knowledge of biological diversity in different areas can optimize the use of scarce resources in choosing areas of high diversity and/or with many

---

Correspondence: Adriano Sanches Melo, Programa de Pós-Graduação em Ecologia, Museu de História Natural, IB, Universidade Estadual de Campinas, Campinas, SP 13083-970, Brazil. E-mail: adrimelo@unicamp.br

rare and endemic species (Prendergast *et al.*, 1993; Pressey *et al.*, 1993). In the second case, one can assess the impact of human disturbance on species loss through comparison with undisturbed areas (Flather, 1996; Keddy & Drummond, 1996).

To achieve a good estimate of species richness in an area, one must take sufficient samples to include most of the rare species (Cao, Williams & Williams, 1998). For the same sample size a species poor assemblage, but with equally abundant species, might produce a greater richness than another truly rich assemblage, but composed of a large number of rare species. A good example of this problem was presented by Stout & Vandermeer (1975) who found greater richness in temperate streams than in tropical counterparts when using few samples. With an increased sample size, tropical streams were shown clearly to be richer than temperate ones.

The assessment of richness in a given area requires a count of observed species, but the larger the sampling effort, the larger will be the number of observed species (Walther *et al.*, 1995). To circumvent this problem, it is necessary either to collect all the individuals in the area, something that is rarely possible, or to use an estimation method, as suggested by many authors (Burnham & Overton, 1979; Smith & van Belle, 1984).

Several mathematical models have been developed, mainly in the past 20 years, to allow for such estimation (see reviews in Bunge & Fitzpatrick, 1993; Colwell & Coddington, 1994). These models are based on different mathematical approaches and can be grouped in three classes: parametric, non-parametric (NP) and extrapolations of species accumulation curves (ESAC). Parametric methods need information on the abundance of each observed species, which are then fitted to a theoretical model of expected frequencies for each class of abundance. Because of the requirement for data on the abundance of each species, laborious computation and bad performance in some previous evaluations (Palmer, 1990; Walther & Morand, 1998), they have been little used in recent years and are not considered in the present study. Non-parametric methods are easily computed and most of them do not need information about the abundance of each species (Burnham & Overton, 1979; Colwell & Coddington, 1994). Most are composed of the number of observed species and a second term related to the number of rare species. The third class of estimators is composed of models of ESAC. Included in this class are the Stout & Vandermeer (1975) method used mainly in stream ecology, and several modifications and transformations of a hyperbole of two parameters known as the Michaelis–Menten equation, first used by biochemists on enzyme kinetics (Clench, 1979; Raaijmakers, 1987; Lamas, Robbins & Harvey, 1991; Soberón & Llorente, 1993).

The choice of a class of methods or a single method is not an easy task. Few evaluations have used either simulated or real data, while a good model for one kind of data might be poor for another. For example, the Bootstrap NP method of resampling was one of the best evaluated in a parasite richness study (Walther & Morand, 1998), but was one of the worst in another evaluation using plants in a tropical forest (Chazdon *et al.*, 1998). This discrepancy occurs because of differing characteristics of the data sets, such as the patchy distribution of species in the area (Chazdon *et al.*, 1998), size of sampling units and the relative number of the sampled species to the available pool (Walther & Morand, 1998).

The best way to evaluate different methods of estimating richness is to compare the estimates to the true richness, something that is usually not possible. In species rich assemblages such as macroinvertebrates in tropical streams, species accumulation curves rarely reach an asymptote, reflecting the continuous collection of rare, previously unseen species. Many of these rare species are from different habitats and hence are collected accidentally. A good way to solve this problem is to remove from analyses those species considered accidental, based on knowledge of their natural history. However, at the moment, this is practically impossible for stream macroinvertebrates in most tropical regions. Thus, our evaluation approach was to estimate total richness in unusually large sample sizes by using different sub-sample sizes.

We evaluated 13 available methods for richness estimation of macroinvertebrates inhabiting riffles at local (single stream) and regional (several streams within the same catchment basin) scales. Evaluation was mainly carried out observing the behaviour of each method and its ability to estimate the richness in a large sample using different sub-sample sizes.

Methods

Study sites

Evaluation was carried out using six macroinvertebrate data sets from quantitative collections in streams located in protected forest areas in São Paulo State, Brazil. Three data sets were from single streams and are here called local samples. The other three, termed regional samples, were from several stream sites within the same catchment basin.

The first local sample was collected in the Ermida Stream (23°14′S, 46°56′W), a third order stream at an elevation of 860 m located in the Serra do Japi, municipality of Jundiai. The mean annual precipitation in the area is about 1400 mm and the vegetation is Tropical Semi-Deciduous Montane Forest. The area was impacted recently by fire and at the collection site the riparian vegetation was an old secondary growth that partially shaded the stream. At the time of the collection, the stream bed was not covered with sediment and the water was clear. Collections were made from September to mid-November 1996, comprising the end of the dry and beginning of the wet seasons. Further information about physical and biological aspects of the area can be found in Morellato (1992). The second collecting site was the Carmo River (24°18′S, 48°25′W), a fourth order stream at an elevation of 520 m, that drains a well conserved and protected area, the Parque Estadual Intervales. The vegetation is Tropical Ombrophilous Submontane–Montane Forest, commonly known as Tropical Rain Forest (Mueller-Dombois & Ellenberg, 1974). The mean annual rainfall is 1696 mm (8 year record) and the collection was carried out in July 1997, in the middle of the dry season. The third local sample was collected in the Cedro Stream (22°45′S, 45°28′W) at an elevation of 950 m in the Serra da Mantiqueira, municipality of Pindamonhangaba. The vegetation is Tropical Evergreen Seasonal Submontane Forest and at the collecting site we did not observe any major human disturbance. The stream has many waterfalls, and in the stream bed large boulders (>0.5 m) are common. Sampling was carried out in December 1998 and January 1999 and although these months were in the middle of the rainy season, we did not observe any spate during the collecting period.

Regional samples were from the Carmo River Basin located in the Parque Estadual Intervales, the same area as the second local sample. The first regional sample was collected in nine stream sites in summer and the second one in the same previous nine stream sites plus one in winter. The third one was composed by the sum of the previous summer and winter samples. This last sample does not present any new information, but was used because it represents a very large sample and so allowed us to evaluate the performance of the estimators on unusually large samples. The streams were first to fifth order and ranged from 1 to 21 m in width. Summer collections were made in the rainy period of February and March and the winter collections in the dry months of July and August, both in 1997.

Sampling and sorting of macroinvertebrates

The sampling and sorting procedures were the same for all samples. They consisted of sampling individual stones 15–20 cm maximum diameter in riffles using a U-net with a 250 μm mesh, designed to avoid the loss of active swimmers (Scrimgeour, Culp & Glozier, 1993). After taking the stones from the stream bed, they were immediately examined for attached individuals. The contents of the U-net were transferred to a white tray and all visible invertebrates fixed in 80% ethanol.

Seventy-five stones were collected in each of the two first local samples, hereafter called Japi and Carmo samples. Because of the non-stabilization of the species accumulation curves in Japi and Carmo samples and given our intention to explore the effects of sample size on the performance of the estimators, we collected 150 stones in the third local sample, termed here Pinda. For the regional samples, 25 stones were collected in each stream totalling 225, 250 and 475 stones, respectively, for the Summer, Winter and the pooled third sample, here called Intervales. These sample sizes are remarkably higher than that used in other studies. As examples, we have used with success, 25 stones to access richness in single streams (Melo & Froehlich, 2000) while Lake *et al.* (1994) sampled 28 stones in a tropical stream in Australia. The only diversity study using stones as sampling units along a catchment we are aware is that of Minshall, Petersen & Nimz (1985), who used 60–110 stones per season. Thus, despite none of our samples reaching an asymptote, we reason that the six samples represent unusually large efforts and the observed richness found in each one should be close to the true richness in the studied assemblages in the sampling period.

Because of the poor knowledge of the macroinvertebrate fauna in south-east Brazil, individuals were identified to the lowest possible taxonomic level and then sorted as morphospecies. Because of difficulties in separation, even as morphospecies, for mites and chironomid larvae, these were not included in the analysis.

Evaluation of estimators

Thirteen estimators were evaluated, comprising seven NP and six ESAC methods (Table 1). Computation of the NP methods, MMMean and MMRuns were carried out using the software EstimateS version 5.0.1 (Colwell, 1997). The estimators MMClench, Expo and SV were computed by non-linear regression (see Keating & Quinn, 1998) and the estimator MMLamas by solving the function using spreadsheet software. The fitting of the ESAC methods was carried out using a species accumulation curve produced by 200 randomizations of the order of sampling units appearances computed in EstimateS. For the estimators calculated by EstimateS, we present estimations for most of the sub-sample sizes while only a sufficient number of points to construct a curve were calculated for the others, because of the necessity of independent calculation for each sub-sample size.

Common approaches used to evaluate estimation methods includes some measure of bias and accuracy of the estimated richness in relation to the true richness using an *a priori* chosen sub-sample size. However, in rich assemblages as used here, the estimated richness is strongly dependent on sample size. Thus, for a given estimation method, different sub-samples sizes will produce different bias and accuracy values (Hellmann & Fowler, 1999). As we do not have any reason to choose a specific sub-sample size *a priori*, we opted for not using such bias and accuracy statistics. Instead, we used four criteria we argue are more practical and realistic. The four criteria were: (1) the smallest sub-sample size required to estimate the observed richness in the total sample, (2) constancy of the sub-sample size needed to estimate the observed richness in the total sample, measured as 1 standard deviation (SD) of the previous criterion, (3) lack of erratic behaviour in curve shape, specifically large variations of estimates for closely similar sub-sample sizes and (4) similarity in curve shape through the six sample data sets. In order to apply the first two criteria and to allow future practical uses, the results are presented as percentages instead of the absolute number of stones.

Table 1 Summary of the 13 evaluated estimators comprising seven non-parametric (NP) methods and six estimators based on the extrapolation of species accumulation curves (ESAC)

Class	Abbreviation	Estimator	References
NP	ACE	Abundance-based coverage estimator	Chao, Ma & Yang (1993); Colwell (1997)
NP	ICE	Incidence-based coverage estimator	Lee & Chao (1994); Colwell (1997)
NP	Chao 1	Chao 1	Chao (1984); Colwell (1997)
NP	Chao 2	Chao 2	Chao (1987); Colwell (1997)
NP	Jack 1	First order Jackknife	Burnham & Overton (1978, 1979)
NP	Jack 2	Second order Jackknife	Burnham & Overton (1978, 1979)
NP	Boot	Bootstrap	Smith & van Belle (1984)
ESAC	MMRuns	Transformation of Michaelis–Menten hyperbole by Raaijmakers. Estimate curves averaged over randomizations (runs)	Raaijmakers (1987); Colwell (1997)
ESAC	MMMean	Transformation of Michaelis–Menten hyperbole by Raaijmakers. Estimate curve computed once for mean species accumulation curve	Raaijmakers (1987); Colwell (1997)
ESAC	MMClench	Michaelis–Menten hyperbole	Clench (1979)
ESAC	MMLamas	Transformation of Michaelis–Menten hyperbole. The curve is adjusted in order to pass through the last point of the species accumulation curve	Lamas, Robbins & Harvey (1991)
ESAC	Expo	Negative exponential function	Soberón & Llorente (1993); Colwell & Coddington (1994)
ESAC	SV	Derivation from the equations of Island Biogeography Theory	Stout & Vandermeer (1975)

Results

The six data sets used were very different from each other (Table 2). The Pinda sample was richer than the two other local samples with 101 observed species on 75 stones, contrasting to 66 for Japi and 71 in Carmo, for the same sample size. Also, the number of species and individuals per stone in Pinda was about twice the number found in Carmo. Finally, the Pinda sample was composed of twice the number of stones sampled in the other two local samples. Regional samples also presented obvious differences to each other, either in observed richness or in sample size (Table 2). Despite these structural differences in the sampled assemblages, results presented by the different estimators were very similar in all six used data sets (Figs 1 & 2).

The estimators MMRuns, MMMean and Expo estimated values only slightly larger than the observed richness at small sub-sample sizes. Increasing the sub-sample sizes, they tended to produce values smaller than the observed richness (Figs 1 & 2). Additionally, MMRuns produced erratic behaviour at small sub-samples sizes.

The estimator Boot produced values only slightly larger than the observed richness but in contrast to the three previous estimators, they were consistent for all sub-sample sizes. MMLamas and MMClench showed similar behaviours to Boot except that they estimated larger values at small sub-sample sizes. The mean sub-sample sizes required to estimate the total richness in the samples were 64, 65.3 and 70.58%, respectively, for MMLamas, Boot and MMClench (Table 3).

The SV method estimated the total richness in the sample with the smallest sub-sample size (Mean 15.58%) and was rather constant through the six samples (SD 4.43%) (Table 3). However, it produced erratic behaviour in Summer and Intervales samples at small sub-sample sizes. Also, the curve shape was not constant along all samples. The SV curves were similar to NP methods for Japi, Winter and Intervales samples. However, in Carmo and Pinda samples the estimates were too high at large sub-sample sizes (Figs 1 & 2).

Unlike the ESAC methods, NP estimators (except Boot) presented more similar behaviours to each other. Incidence-based coverage estimator (ICE) produced erratic behaviour at small sub-sample sizes. In Japi, Carmo and Pinda samples, Jack 2 was similar to Chao 1 and Chao 2 methods, but produced larger estimated values for the other three samples (Figs 1 & 2). Jack 2 estimated total richness with the smallest sub-sample size among the NP methods and the constancy of this estimate was inferior only to Boot method (Table 3).

The best estimator with respect to minimum sub-sample size was the SV method, but it did not present good results in relation to two other criteria. Boot estimator had the best performance in relation to the constancy of minimum sub-sample size, but needed an unacceptably large sub-sample size to estimate total richness in the sample (Table 3). On the other hand, Jack 2 was the second best both in minimum sub-sample size and in constancy and did not present problems in relation to the other two criteria (Table 4). Although Chao 1, Chao 2 and Jack 1 scored lower when compared with Jack 2, they presented good results in relation to all four criteria. Abundance-based coverage estimator (ACE) and ICE methods followed in an inferior level. Boot, MMLamas and MMClench estimators performed poorly with regard to observed richness. The worst of the evaluated methods were MMRuns, MMMean and Expo, which underestimated the observed richness at many sub-sample sizes, including when using the total sample (Table 3).

Table 2 Summary of observed richness and sample size of the three local samples (Japi, Carmo and Pinda) and the three regional samples (Summer, Winter and Intervales) used in the evaluations

Sample	Observed richness	Sample size (stones)	Total individuals	Mean species per stone	Mean individuals per stone
Japi	66	75	3759	12.67	50.12
Carmo	71	75	2673	9.91	35.64
Pinda	117	150	10 339	18.46	68.93
Summer	119	225	6831	9.68	30.36
Winter	142	250	10 767	10.99	43.07
Intervales	162	475	17 598	10.62	37.05

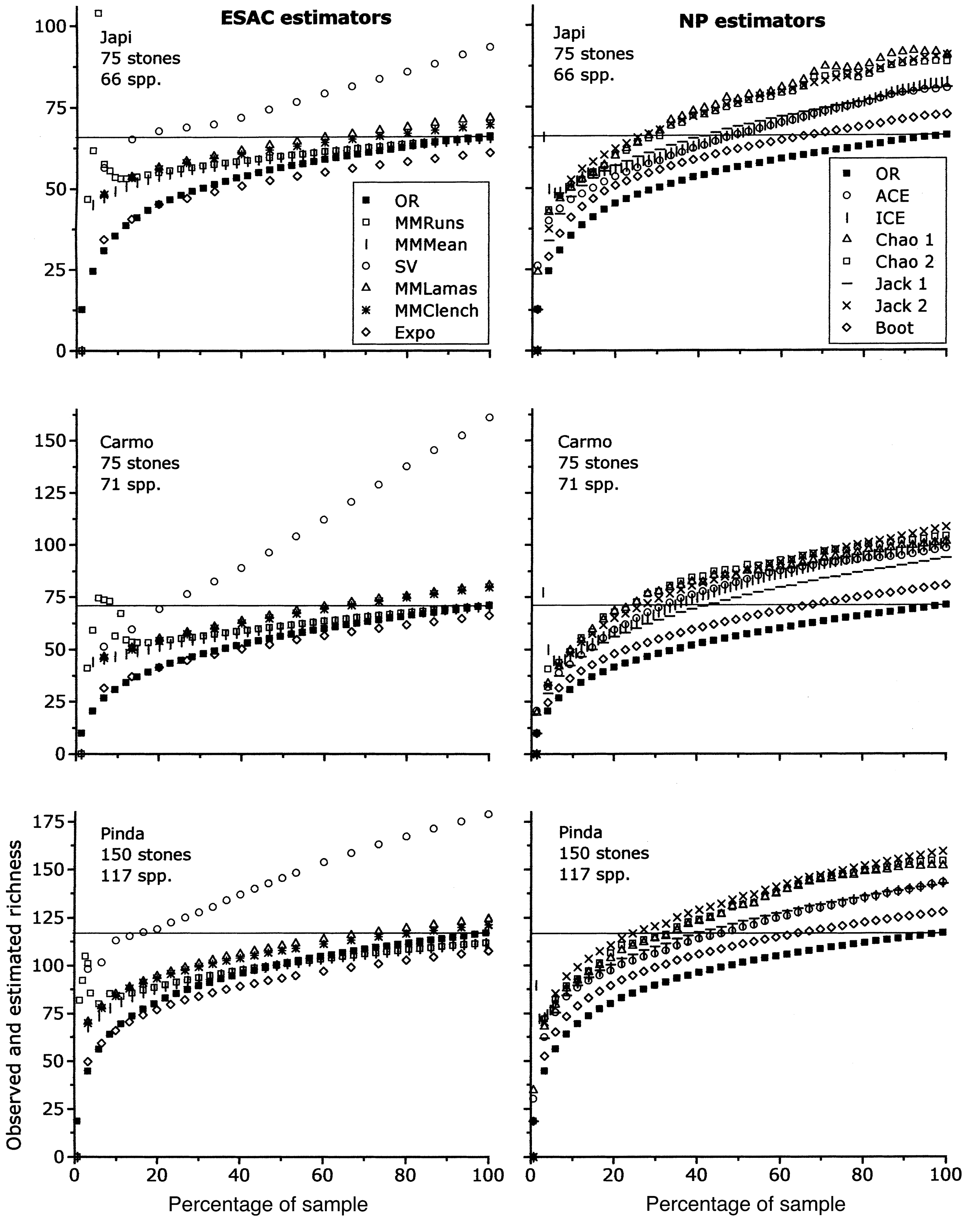

Fig. 1 Curves of accumulation of observed species and estimates for local samples. The horizontal line indicates the total number of species observed in the sample. OR = Observed richness. See Table 1 for definitions of the estimators.

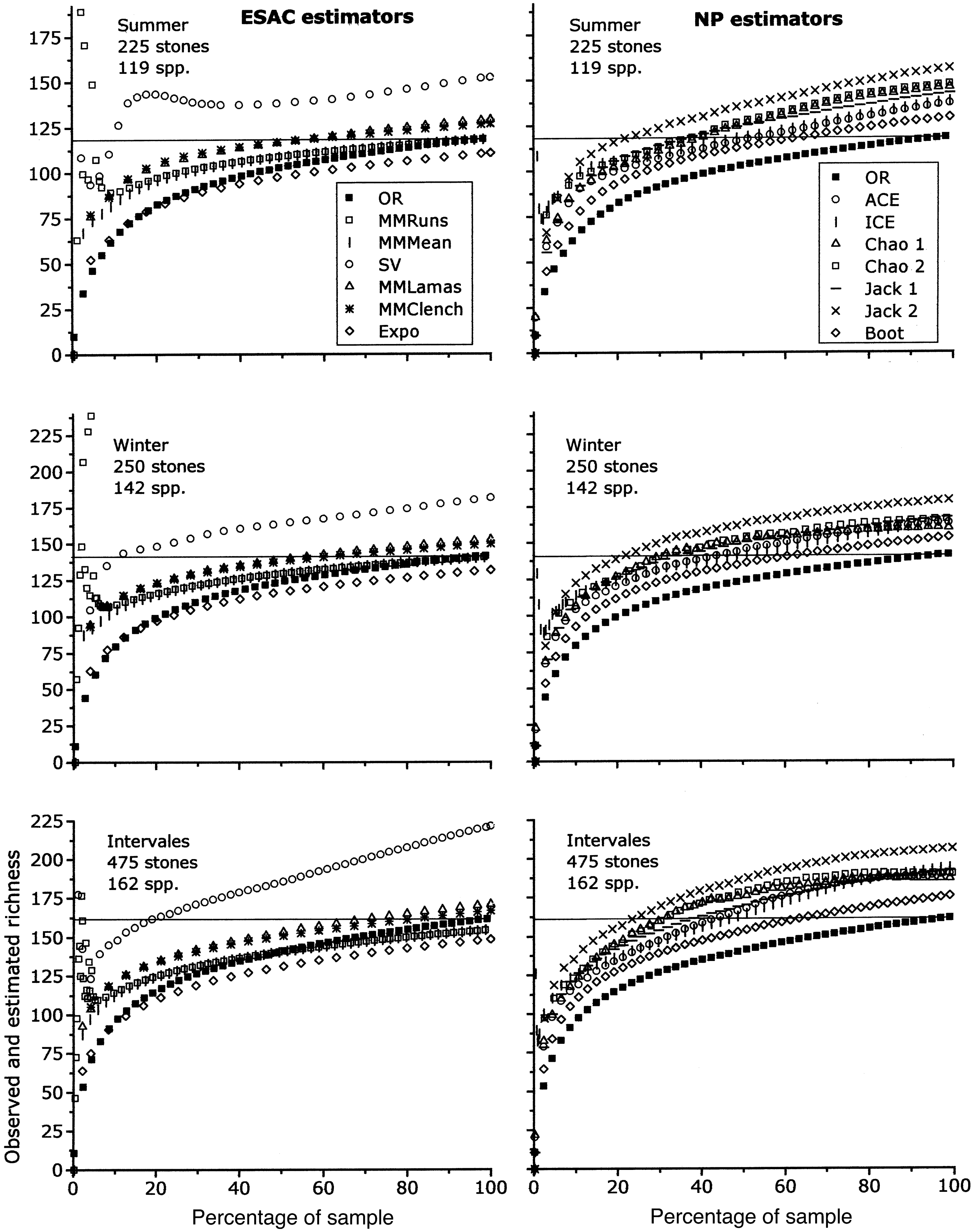

Fig. 2 Curves of accumulation of observed species and estimates for regional samples. The horizontal line indicates the total number of species observed in the sample. OR = Observed richness. See Table 1 for definitions of the estimators.

Table 3 Minimum sub-sample size expressed as percentage required by each method to estimate the observed richness in the three local and three regional samples. MMRuns, MMMean and Expo estimators underestimated observed richness in the total sample even when using all sampling units

Estimator	Japi	Carmo	Pinda	Summer	Winter	Intervales	$\bar{X}$	SD
ACE	48.0	32.0	47.3	54.7	45.2	44.0	45.20	7.46
ICE	46.7	34.7	44.7	50.7	48.8	47.6	45.53	5.67
Chao 1	28.0	24.0	34.0	40.9	31.6	31.4	31.65	5.71
Chao 2	30.7	24.0	33.3	38.7	32.8	31.8	31.88	4.75
Jack 1	41.3	41.3	40.0	37.3	35.6	38.5	39.00	2.29
Jack 2	26.7	26.7	25.3	23.6	22.4	24.2	24.81	1.73
Boot	66.7	66.7	65.3	64.9	63.6	64.6	65.30	1.22
MMRuns	–	–	–	–	–	–	–	–
MMMean	–	–	–	–	–	–	–	–
MMClench	72.0	66.7	80.7	58.2	64.0	81.9	70.58	9.42
MMLamas	62.7	62.7	70.7	57.8	59.2	70.9	64.00	5.61
Expo	–	–	–	–	–	–	–	–
SV	16.0	21.3	16.0	10.2	10.8	19.2	15.58	4.43

Table 4 Performance of the estimators scored by four criteria. Minimum sample size (MSS) required to estimate richness in total sample; CSS, constancy of required sub-sample size (1 SD of MSS); EB, erratic behaviour; SCS, similarity in curve shape. MSS and CSS ordered by performance

Estimator	MSS	CSS	EB	SCS
ACE	6	9	Good	Good
ICE	7	7	Bad	Good
Chao 1	3	8	Good	Good
Chao 2	4	5	Good	Good
Jack 1	5	3	Good	Good
Jack 2	2	2	Good	Good
Boot	9	1	Good	Good
MMRuns	11	11	Bad	Good
MMMean	11	11	Good	Good
MMClench	10	10	Good	Good
MMLamas	8	6	Good	Good
Expo	11	11	Good	Good
SV	1	4	Bad	Bad

Discussion

Non-parametric methods were, overall, better than ESAC. Expo, Boot and all four transformations or modifications of the Michael–Menten hyperbole presented the poorest results, and should not be used for species rich data with many rare species. The ACE and ICE methods performed at an intermediate level. Jack 2 was the best of the evaluated methods based on the four previously chosen criteria and we recommend its use in future studies of macroinvertebrate diversity in tropical streams. Jack 1, Chao 1 and Chao 2 presented good results and were inferior to Jack 2 only because of the requirement for larger sub-sample sizes.

In spite of the use of different approaches and different data sets by different authors to evaluate estimation methods, there is some congruence in results. The bad performance of Expo estimator was also found by Peterson & Slade (1998). The Boot estimator, which needed an unacceptably large sub-sample to estimate sample richness (65.3%), was also a poor estimator in the studies of Colwell & Coddington (1994) and Chazdon *et al.* (1998). On the other hand, Boot was considered a good estimator when utilized in species poor samples with few rare species, such as parasite species richness (Walther & Morand, 1998), and this is also in accord with Smith & van Belle (1984) who suggested the Boot method for well sampled assemblages. MMClench produced poor estimates of richness in our samples and was also considered inadequate by Keating & Quinn (1998) on simulated and real data sets. The SV method, mainly used in stream ecology (Stout & Vandermeer, 1975; Minshall *et al.*, 1985; Haynes, 1987), which has only been evaluated once on a simulated data set, produced a poor result when compared with the other two methods tested, Jack 1 and the parametric Cohen method (Baltanás, 1992). The SV estimator was the best in relation to minimum sub-sample size required (Tables 3 & 4), but the curve shape was not constant among the six data sets and hence cannot be used with confidence (Figs 1 & 2). We observed that when fitting SV to observed richness at different sub-sample sizes, the fitted curve was very similar to the observed points. However, the extrapolated part of the fitted curves was too sensitive to small differences in the shape of the observed richness curve, resulting in very different values of the extrapolated asymptote.

The recently developed ICE and ACE methods received intermediate scores in our analyses and have only been evaluated twice before, producing conflicting results (Chazdon *et al*., 1998; Walther & Morand, 1998). In the species poor assemblage of parasites per host they were incapable of producing any estimate for many sub-sample sizes and were considered inadequate for this kind of data set (Walther & Morand, 1998). On the other hand, Chazdon *et al*. (1998) evaluated eight methods for predicting species richness of seedlings and saplings in tropical forests in Costa Rica and concluded that the ICE method was the best. They found the method to be robust to sample size and patchiness and suggest its use in sites with high species richness. One problem stated by Walther & Morand (1998) is that ACE and ICE methods need definition, *a priori*, of a parameter related to the definition of rare species. We used the default value 10 proposed by Colwell (1997) in EstimateS and also used by Chazdon *et al*. (1998). Increasing this value gave higher estimated values, reaching values close (ICE) or superior (ACE) to those estimated by Jack 2, when the parameter was set to the number of stones in the sample (results not shown here).

The best of the evaluated methods, Jack 2, Jack 1, Chao 1 and Chao 2 were also considered the best or among the best estimators in other evaluations and their use has been suggested by many authors. Colwell & Coddington (1994) evaluated the performance of eight methods on a seed-bank data set and suggested the use of Chao 2 and Jack 2. Chazdon *et al*. (1998) in a study of seedling and sapling diversity found that ICE and Chao 2 were robust to sample size and patchiness. Although Palmer (1990, 1991) used a different approach to evaluate estimation methods, he proposed the use of Jack 1 and Jack 2, while Walther & Morand's (1998) evaluation of parasites per host data sets recommended the use of Chao 2 and Jack 1. Peterson & Slade (1998) tested seven methods for their ability to estimate the number of states in Mexico and United States by using records of automobile licence plates taken in two cities along several days and found the Chao 2 method was one of the best (they did not evaluate Jack 1, Jack 2 and Chao 1).

An ideal situation for evaluating richness estimators is to compare the estimated value to total species richness in an area. However, to know the total richness in rich communities with many rare species, such as tropical stream macroinvertebrates, an unfeasible large sample may be needed. As many species exist at larger space and time scales (regional pool), we will practically never know how many species there are in a given place because of the continuous appearance of rare species over time and with increasing catch effort.

Another problem of evaluating richness estimators are the criteria used to score the different methods. Palmer (1990, 1991) and Baltanás (1992) chose one fixed sub-sample size and applied several statistics to decide which method could estimate the richness in total samples with low bias and high precision. They did not present any explanation for choice of sub-sample size. A quick look at Figs 1 & 2 and Table 3 reveals that except for the Expo, MMMean and MMRuns methods, all other methods were capable of estimating the total sample richness, but at different sub-sample sizes. If we opt for an *a priori* sub-sample size of say 65% of samples, we would conclude that Boot was the best estimator.

An important characteristic expected from a estimation method, is that it should estimate total richness independently of sample size. However, as Figs 1 & 2 show, all evaluated methods in our study were strongly dependent of sub-sample size. In practice, this means that the richness estimate for a given sample size is not a reliable estimate of the true richness in the studied assemblage. It is likely that a different sub-sample size would produce a different estimated richness. Thus, agreeing to Baltanás (1992) and Schmit, Murphy & Mueller (1999), caution should be taken when using the absolute values produced by some estimation method from species rich assemblages to assess total richness in an area.

On the other hand, the very close sub-sample sizes needed for some methods to estimate the observed richness in the total sample (Table 3), even through rather different data sets as the six used samples (Table 2), reinforces the suggestions of Palmer (1990) and Baltanás (1992) about the good reliability of using such methods for comparison purposes. In the case of Jack 2, Jack 1 and Boot estimators, the range of sub-sample sizes needed to estimate richness in total samples were, respectively, 22.4–26.7, 35.6–41.3 and

63.6–66.7% of total samples. In a similar study, Hellmann & Fowler (1999) used five data sets of plants with different species richness and proportions of rare species and found that for Jack 1, Jack 2 and Boot, the sub-sample size needed to estimate richness in the total sample were, respectively, 22.6–29.1, 36.8–43.9 and 63.1–69.0% of total samples, values very close to those found in our study. Thus, it seems that these estimation methods are not only strongly dependent on sub-sample size but also that this dependence occurs in a predictable way. In other words, at least for assemblages where species accumulation curves are not approaching an asymptote, it is possible to estimate the richness in a sample of a given size by using a sub-sample of a constant proportion. Using the Jack 2 method, the richness expected in a sample of 40 sampling units can be estimated by using only 10 sampling units. As highlighted before, in many cases there is a reason for using an estimated richness instead of the observed richness. If samples do not have an appropriate size, potential erroneous conclusions can be obtained as a result of differences in equitability and proportion of rare species between data sets. For example, despite the Carmo sample being richer than the Japi sample (Table 2), this was apparent in the curve of accumulated observed richness only after 38 stones had been pooled. Using the estimated values of the Jack 2, this was apparent after the pooling of 14 stones (see Lande, DeVries & Walla, 2000) for a more comprehensive statement of the problem and an alternative solution using a diversity index). Anyway, further studies addressing specifically this question are necessary for a reliable conclusion.

Despite the structural differences among sampled assemblages highlighted previously, Jack 2 and, to a lesser extent, Jack 1, Chao 1 and Chao 2 presented very similar results across all six data sets. Taking into account their good performance in other evaluations cited above, these methods seen robust to variation in data structure and hence should be used for comparative purposes in assemblages similar to those tested here and in those kinds of assemblages for which no previous evaluation is available.

Acknowledgments

The unusually large samples used in this work were collected with the friendly aid of Rodolfo Lima, Pedro Castelo-Branco, André Ribeiro-Chaves, Flávia Batista, Anders Silva, Fábio Vanini, José Flóido, Dito Amaral, Elias, Eliseu, Faustino, Luiz, and the staff at Parque Estadual Intervales. Votorantim Celulose e Papel and CEMASI permitted collections at the Pinda site. The Guarda Municipal de Jundiai and the Fundação Florestal gave permission to work in Japi and Intervales, respectively. The manuscript was greatly improved by comments from Adalberto Santos, Hermes Medeiros, Ricardo Sawaya, Robert Colwell, George Shepherd, Colin Townsend and three anonymous referees. A.S.M. received student fellowships from CAPES and FAPESP (proc. 98/00590-0) and C.G.F. a research grant from FAPESP (proc. 98/05073-4) and a research fellowship from CNPq (proc. 301247/1996-0). This work is part of the BIOTA/FAPESP – The Biodiversity Virtual Institute Program (http://www.biotasp.org.br).

References

Baltanás A. (1992) On the use of some methods for the estimation of species richness. *Oikos*, **65**, 484–492.

Bunge J. & Fitzpatrick M. (1993) Estimating the number of species: a review. *Journal of the American Statistical Association*, **88**, 364–373.

Burnham K.P. & Overton W.S. (1978) Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika*, **65**, 625–633.

Burnham K.P. & Overton W.S. (1979) Robust estimation of population size when capture probabilities vary among animals. *Ecology*, **60**, 927–936.

Cao Y., Williams D.D. & Williams N.E. (1998) How important are rare species in aquatic community ecology and bioassessment? *Limnology and Oceanography*, **43**, 1403–1409.

Chao A. (1984) Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, **11**, 265–270.

Chao A. (1987) Estimating the population size for capture–recapture data with unequal catchability. *Biometrics*, **43**, 783–791.

Chao A., Ma M.-C. & Yang M.C.K. (1993) Stopping rules and estimation for recapture debugging with unequal failure rates. *Biometrika*, **80**, 193–201.

Chazdon R.L., Colwell R.K., Denslow J.S. & Guariguata M.R. (1998) Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forest of northeastern Costa Rica. In: *Forest Biodiversity Research, Monitoring and Modelling: Conceptual Background and Old World Case Studies* (Eds F. Dallmeier & J.A. Comiskey), pp. 285–309. Pantheon Press, Paris.

Clench H.K. (1979) How to make regional lists of butterflies: some thoughts. *Journal of the Lepidopterists' Society*, **33,** 216–231.

Colwell R.K. (1997) *EstimateS: Statistical Estimation of Species Richness and Shared Species from Samples*. User's Guide and Application. http://viceroy.eeb.uconn.edu/estimates.

Colwell R.K. & Coddington J.A. (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions Royal Society of London Series B*, **345,** 101–118.

Flather C.H. (1996) Fitting species-accumulation functions and assessing regional land use impacts on avian diversity. *Journal of Biogeography*, **23,** 155–168.

Haynes A. (1987) Species richness, abundance and biomass of benthic invertebrates in a lowland tropical stream on the island of Viti Levu, Fiji. *Archiv für Hydrobiologie*, **110,** 451–459.

Hellmann J.J. & Fowler G.W. (1999) Bias, precision, and accuracy of four measures of species richness. *Ecological Applications*, **9,** 824–834.

Keating K.A. & Quinn J.F. (1998) Estimating species richness: the Michaelis–Menten model revisited. *Oikos*, **81,** 411–416.

Keddy P.A. & Drummond C.G. (1996) Ecological properties for the evaluation, management, and restoration of temperate deciduous forest ecosystems. *Ecological Applications*, **6,** 748–762.

Lake P.S., Schreiber E.S.G., Milne B.J. & Pearson R.G. (1994) Species richness in streams: patterns over time, with stream size and with latitude. *Verhandlungen der Internationalen Vereinigung für Theoretische und Angewandte Limnologie*, **25,** 1822–1826.

Lamas G., Robbins R.K. & Harvey D.J. (1991) A preliminary survey of the butterfly fauna of Pakitza, Parque Nacional del Manu, Peru, with an estimate of its species richness. *Publicaciones Del Museo de Historia Natural, Universidad de San Marcos, Peru*, **40,** 1–19.

Lande R., DeVries P.J. & Walla T.R. (2000) When species accumulation curves intersect: implications for ranking diversity using small samples. *Oikos*, **89,** 601–605.

Lee S.-M. & Chao A. (1994) Estimating population size via sample coverage for closed capture–recapture models. *Biometrics*, **50,** 88–97.

Magurran A.E. (1988) *Ecological Diversity and its Measurement*. Croom-Helm, London.

Melo A.S. & Froehlich C.G. (2000) Macroinvertebrates in neotropical streams: richness patterns along a catchment and assemblage structure between two seasons. *Journal of the North American Benthological Society*, **20,** 1–16.

Minshall G.W., Petersen R.C.J. & Nimz C.F. (1985) Species richness in streams of different size from the same drainage basin. *American Naturalist*, **125,** 16–38.

Morellato L.P.C. (1992) *História Natural Da Serra Do Japi*. Editora da Unicamp/Fapesp, Campinas.

Mueller-Dombois D. & Ellenberg H. (1974) *Aims and Methods of Vegetation Ecology*. Wiley, New York.

Palmer M.W. (1990) The estimation of species richness by extrapolation. *Ecology*, **71,** 1195–1198.

Palmer M.W. (1991) Estimating species richness: the second-order jackknife reconsidered. *Ecology*, **72,** 1512–1513.

Peterson A.T. & Slade N.A. (1998) Extrapolating inventory results into biodiversity estimates and the importance of stopping rules. *Diversity and Distributions*, **4,** 95–105.

Prendergast J.R., Quinn R.M., Lawton J.H., Eversham B.C. & Gibbons D.W. (1993) Rare species, the coincidence of diversity hotspots and conservation strategies. *Nature*, **365,** 335–337.

Pressey R.L., Humphries C.J., Margules C.R., Vane-Wright R.I. & Williams P.H. (1993) Beyond opportunism: key principles for systematic reserve selection. *Trends in Ecology and Evolution*, **8,** 124–128.

Raaijmakers J.G.W. (1987) Statistical analysis of the Michaelis–Menten equation. *Biometrics*, **43,** 793–803.

Schmit J.P., Muphy J.F. & Mueller G.M. (1999) Macrofungal diversity of a temperate oak forest: a test of species richness estimators. *Canadian Journal of Botany*, **77,** 1014–1027.

Scrimgeour G.J., Culp J.M. & Glozier N.E. (1993) An improved technique for sampling lotic invertebrates. *Hydrobiologia*, **254,** 65–71.

Smith E.P. & van Belle G. (1984) Nonparametric estimation of species richness. *Biometrics*, **40,** 119–129.

Soberón J. & Llorente J. (1993) The use of species accumulation functions for the prediction of species richness. *Conservation Biology*, **7,** 480–488.

Stout J. & Vandermeer J. (1975) Comparison of species richness for stream-inhabiting insects in tropical and mid-latitude streams. *American Naturalist*, **109,** 263–280.

Walther B.A. & Morand S. (1998) Comparative performance of species richness estimation methods. *Parasitology*, **116,** 395–405.

Walther B.A., Cotgreave P., Price R.D., Gregory R.D. & Clayton D.H. (1995) Sampling effort and parasite species richness. *Parasitology Today*, **11,** 306–310.

(Manuscript accepted 1 September 2000)