

Comparative performance of species richness estimation methods

B. A. WALTHER^{1*}† and S. MORAND²

¹Department of Zoology, Oxford University, Oxford OX1 3PS, UK

²Centre de Biologie et d'Écologie tropicale et méditerranéenne, Université de Perpignan, 66860 Perpignan, France

(Received 13 June 1997; revised 9 August and 26 September 1997; accepted 26 September 1997)

SUMMARY

In most real-world contexts the sampling effort needed to attain an accurate estimate of total species richness is excessive. Therefore, methods to estimate total species richness from incomplete collections need to be developed and tested. Using real and computer-simulated parasite data sets, the performances of 9 species richness estimation methods were compared. For all data sets, each estimation method was used to calculate the projected species richness at increasing levels of sampling effort. The performance of each method was evaluated by calculating the bias and precision of its estimates against the known total species richness. Performance was evaluated with increasing sampling effort and across different model communities. For the real data sets, the Chao2 and first-order jackknife estimators performed best. For the simulated data sets, the first-order jackknife estimator performed best at low sampling effort but, with increasing sampling effort, the bootstrap estimator outperformed all other estimators. Estimator performance increased with increasing species richness, aggregation level of individuals among samples and overall population size. Overall, the Chao2 and the first-order jackknife estimation methods performed best and should be used to control for the confounding effects of sampling effort in studies of parasite species richness. Potential uses of and practical problems with species richness estimation methods are discussed.

Key words: species richness, estimation methods, sampling methods, diversity, species accumulation curves, jackknife, bootstrap.

INTRODUCTION

The concept of species richness is one of the oldest and most fundamental in community ecology (Peet, 1974), perhaps because it is such a simple concept: the total count of all species present. Total species richness can theoretically be determined for any community, because the number of species is limited. In practice, counting the number of species 'faces the operational disadvantage of being dependent on sample size' (Baltanás, 1992). For example, estimates of parasite species richness are often confounded by uneven sampling effort (Walther *et al.* 1995). Since the sampling effort needed to attain a total species richness count is excessive in most real-world contexts, it has become paramount to develop and test methods to estimate total species richness from incomplete collections.

The literature on estimating species richness is extensive (for recent reviews, see Bunge & Fitzpatrick, 1993; Colwell & Coddington, 1994; Walther *et al.* 1995; Chazdon *et al.* 1997). In ecological contexts, 3 types of methods have been used (Palmer, 1990; Baltanás, 1992; Colwell & Coddington, 1994; Walther *et al.* 1995): (1) the fitting of species-

abundance distributions, (2) the extrapolation of species accumulation curves and (3) non-parametric estimators. The fitting of species-abundance distributions faces serious questions of methodology (Colwell & Coddington, 1994) and performance (Palmer, 1990, 1991). For example, the most widely used model, the log-normal model (Preston, 1948), is a continuous model which nevertheless needs to be fitted to discrete data. Furthermore, estimates of species richness are affected by the choice of the width and the cut-off points of the intervals of the abundance categories (Colwell & Coddington, 1994). Finally, the log-normal model performed almost as badly as the species accumulation curve in Palmer's (1990, 1991) comparative study of species richness estimators. Since the methodological problems with this approach are not yet resolved, we did not fit any species-abundance distributions to our data. Rather, we used the computer program *EstiMateS*© (R. K. Colwell, unpublished) which includes 2 accumulation curve models and 7 non-parametric estimators.

Graphically, a species accumulation curve is a plot of cumulative species richness against sampling effort with the curve approaching the total species richness asymptote as sampling effort increases (Fig. 1). The 2 curve models are based on the Michaelis-Menten equation of enzyme kinetics, using a maximum likelihood transformation proposed by Raaij-

* Corresponding author: Hauptstrasse 25, 27432 Ebersdorf, Germany. Tel: +49 4765 431.

† Present address: Hauptstrasse 25, 27432 Ebersdorf, Germany.

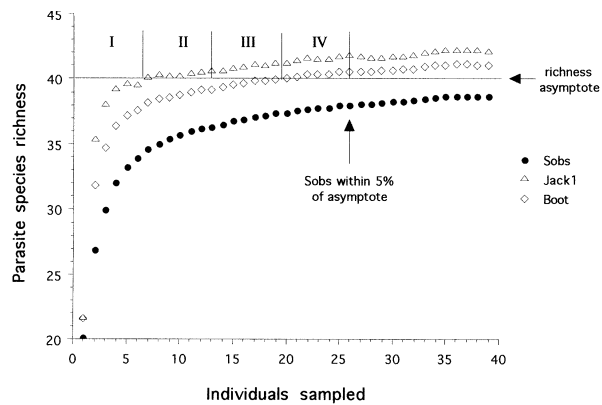


Fig. 1. Plot of cumulative parasite species richness against sampling effort (in this case, individuals sampled). The number of observed species (Sobs) is equivalent to the species accumulation curve which will merge with the total species richness asymptote, once sampling is complete. Species richness estimates of 2 non-parametric estimation methods (the first-order jackknife (Jack1) and the bootstrap (Boot) method; see Materials and Methods section) are also depicted. Data are taken from a simulated data set. Sampling effort prior to reaching within 5% of the asymptote was arbitrarily divided into 4 quarters (I–IV).

makers (1987). Estimated species richness is calculated by extrapolating the curve to its asymptote at each level of sampling effort. However, the order in which samples are drawn from the pool of samples influences the curve's shape and thus the estimated species richness. To avoid this sampling bias, the program samples the entire pool of samples repeatedly using different random orders of samples and then averages species richness estimates over many runs (e.g. 100 times). Averaging estimates can be done in 2 ways: (1) for each randomized run, estimated species richness is calculated for each sample size, and then the 100 estimates are averaged for each given sample size; or (2) the mean accumulation curve is calculated by averaging over 100 accumulation curves derived from 100 runs, and this curve is then used to estimate species richness for each sample size just once (these 2 methods correspond to MMRuns and MMMean, respectively; see Materials and Methods section).

Most non-parametric estimators, on the other hand, add to the number of observed species (which is equivalent to the species accumulation curve) an estimate which is based on the abundance or incidence of rare species (Colwell & Coddington, 1994; Chazdon *et al.* 1997). Abundance is the number of individuals of 1 species in all samples, and incidence is the number of samples containing individuals of 1 species. Thus, these methods use the number of rare species which occur just once, twice, three times, etc. to estimate the number of yet undiscovered species. Since the abundance and incidence of rare species changes with increasing sampling effort, these methods are expected to

calculate different species richness estimates as sampling effort increases.

Likewise, they should return different estimates if the relative abundance of rare species changes within the population. As different populations have different species-abundance distributions, estimator performance should depend on the species-abundance distribution of the data set being analysed (Bunge & Fitzpatrick, 1993; Soberón & Llorente, 1993; Colwell & Coddington, 1994; Walther *et al.* 1995). Thus, for different data sets, different estimators are expected to perform best. Preference for a specific estimation method should be a pragmatic one, testing performance of estimators for a wide variety of model communities with differing community parameters and species-abundance distributions (Colwell & Coddington, 1994). In this study, we compare the performance of 9 estimation methods when analysing data sets with species-abundance distributions typical of parasite populations among hosts.

Comparisons of methods have been few so far, and these were made using either computer-simulated (Chao, 1987; Raaijmakers, 1987; Baltanás, 1992) or real data sets (Palmer, 1990, 1991; Colwell & Coddington, 1994; Chazdon *et al.* 1997). Computer-simulated data allow control over the setting of community parameters (e.g. total species richness, dispersion of mean abundances) as well as the generation of bodies of data large enough for statistical analysis (Baltanás, 1992); however, they may not mimic real patterns of community structure (Palmer, 1990). Therefore, the performance of the 9 estimation methods was evaluated using simulated as well as real data sets.

The performance of each method was evaluated by calculating the bias and precision of its estimates against the known total species richness. Bias measures whether an estimate consistently under- or overestimates the parameter; precision measures the overall closeness of the estimate to the parameter without measuring bias (Zar, 1996). Good estimators should have zero bias and small precision values. These 2 measures allow an objective quantitative comparison of the performance of estimation methods.

MATERIALS AND METHODS

The analysis was deliberately structured to be similar to the analyses performed by Palmer (1990, 1991) and Baltanás (1992) allowing a comparison of the results. For this reason alone, several definitions and methods were borrowed from their work.

Real data sets

Five real data sets were available for analysis. (1) Abundance data for 3 chewing louse species

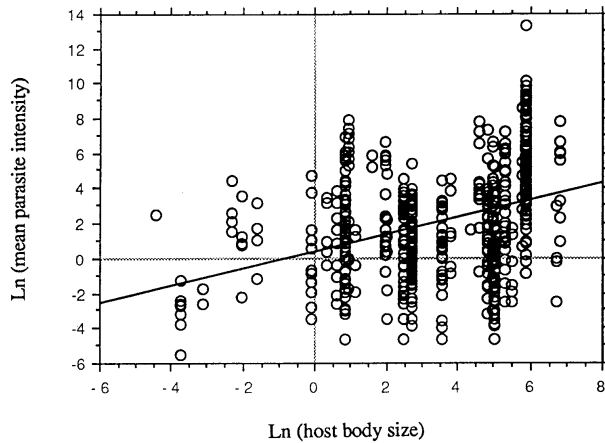


Fig. 2. Regression of Ln (host body size in kg) against Ln (mean intensity) of nematode populations occurring in 50 mammalian host species (range of body size: 0.01–900 kg). See text for details. Regression equation: $y = 0.49x + 0.38$, $n = 546$, $r^2 = 0.13$, $P < 0.0001$.

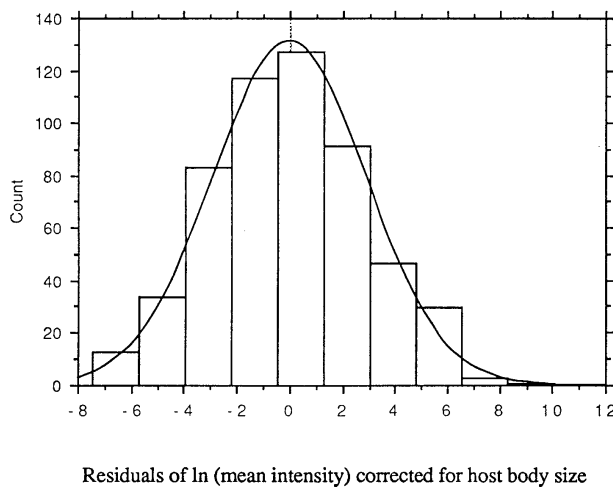


Fig. 3. Plot of residuals around the regression line of Fig. 2 (mean = 0.0, s.d. = 2.9, controlled for body size).

sampled from 67 individual woodcreepers *Glyphorhynchus spirurus* (Clayton, Gregory & Price, 1992). (2) Abundance data for 6 helminth species sampled from 396 individual fish *Gobius bucchichii* (Sasal, Faliex & Morand, 1996). (3) Abundance data for 9 helminth species sampled from 246 individual frogs *Rana temporaria* (E. Faliex & P. Durand, unpublished data). (4) Incidence data for 10 helminth species sampled from 127 individual rabbits *Oryctolagus cuniculus* (J. C. Casanova, unpublished data). (5) Incidence data for 4 blood parasite species sampled from 94 individual tawny owls *Strix aluco* sampled over 2 years resulting in 125 samples. (B. Appleby, unpublished data).

Total species richness for each real data set was determined as follows. For 100 randomized runs, *EstiMateS*© was used to calculate the mean species accumulation curve for each data set. We assumed that the total species richness (e.g. 3 species for the woodcreeper data set) of the community had been

determined if (1) the last 5% of the values of the accumulation curve had equal values (i.e. the curve had reached the horizontal asymptote) and (2) the last 50% of the accumulation curve values were within 5% of that final value of the accumulation curve. The 5 data sets above met these 2 criteria. Many more data sets were analysed but rejected because they did not meet these criteria.

Computer-simulated data sets

We created the simulated data sets with a PASCAL computer program using random number generators. In our simulation, each host individual represents 1 sampling unit in which all parasite species could theoretically occur. The simulation thus created an abundance matrix of parasite species versus individual hosts with each matrix cell containing the number of individual parasites (of the given parasite species found in the given host individual).

We modelled the distribution of the individuals of each parasite species within hosts using a negative binomial distribution (Elliot, 1977; Southwood, 1978). For all parasite species, the parameter k was held constant, while the mean intensity μ (the number of individual parasites of 1 parasite species divided by all sampled host individuals of the given host species, including uninfected individuals) was varied for each parasite species.

Mean intensities were varied according to a log-normal distribution (Southwood, 1978; Shaw & Dobson, 1995). We used values extracted from 32 published studies reporting the mean intensities of nematode species within 50 mammalian host species to yield a regression equation which relates host body size to mean intensity of parasites (Fig. 2). The residuals of this equation were plotted to yield the standard deviation around the regression line, i.e. a measure of variation of the mean intensities (Fig. 3). The standard deviation was held constant because the residuals have constant and normal variance around the regression line (Fig. 2).

This procedure caused both the distribution of the number of parasite species among hosts, as well as the distribution of the number of individual parasites of each parasite species among hosts, to be aggregated. This is in accordance with observations on distributions of many different parasite taxa within hosts (e.g. Anderson & May, 1985; Dobson & Keymer, 1990).

We imposed constraints on the simulation procedure to keep the data sets within realistic boundaries. The chosen numbers (for species richness, aggregation levels and population sizes) reflect levels commonly found in nematode communities of vertebrate hosts (e.g. Gregory, 1990; Poulin, 1993; Shaw & Dobson, 1995). We varied 3 community parameters: total species richness (10, 20 and 40

species), the parameter k (0.5, 1.0 and 5.0), and the mean intensity of each parasite species within each individual host. To determine parasite mean intensity, we chose 3 host body sizes (0.01, 0.1 and 1.0 kilogram). These values were entered into the regression equation (Fig. 2) to yield an average mean intensity for each parasite species (0.15, 0.47 and 1.46 parasites per host, respectively). We varied mean intensity because, presumably, larger hosts can sustain a higher number of individual parasites.

We resampled any individual host if the sum of its parasites exceeded the given total mean intensity, thus avoiding any unrealistically high intensities which would presumably result in the death of the host in the real world. Although a constant lethal level for a given host size is an oversimplification (Anderson & May, 1978; Dobson & Keymer, 1990), a more elaborate model was not necessary for our purposes. We also excluded any parasite species if the sum of its individuals was zero because this is the equivalent to an undetected parasite species which would not be entered into a real-life data set matrix. We sampled each of the 27 parasite communities (3 levels of species richness*3 values of k *3 mean intensities) 10 times for 100 individual hosts resulting in 27 000 sampled hosts in 270 data sets.

Estimation methods

The unpublished program *EstiMateS*© (available from R. K. Colwell, Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269-3042, USA) computes the following 9 species richness estimation methods for each data set (all abbreviations are taken from the user's guide, version 4.1): the extrapolated accumulation curves MMRuns and MMMean (Raaijmakers, 1987), and the non-parametric estimators Chao1 (Chao, 1984), Chao2 (Chao, 1987), ACE (Chao, Ma & Yang, 1993), ICE (Lee & Chao, 1994), the first- and second-order jackknives Jack1 and Jack2 and the bootstrap Boot (Burnham & Overton, 1978, 1979; Heltshé & Forrester, 1983; Smith & van Belle, 1984; Palmer, 1991). The program also calculates the number of observed species (Sobs) which is equivalent to the species richness accumulation curve. For comparative purposes, Sobs was also included as an 'estimator', even though Sobs is always a negatively biased estimator. For further details on these estimators, refer to Colwell & Coddington (1994) and Chazdon *et al.* (1997).

Version 3.1 of *EstiMateS*© included 2 estimators, CandL1 and CandL2 (Chao & Lee, 1992), which performed badly with several data sets (Colwell & Coddington, 1994; R. K. Colwell, personal communication). As they also performed badly with our simulated data sets (unpublished results), these estimators were replaced by the modified estimators ACE and ICE of version 4.1.

Program parameters of *EstiMateS*© were set as follows: 100 randomized runs were performed on each data set, the initial random number generator seed was 17, and the number of abundance/incidence classes for ACE and ICE was 20 unless stated otherwise. The patchiness parameter A was set at default, so that the patchiness (or aggregation) of the data was unaffected.

Performance evaluation

Species accumulation curves eventually approach the total species richness asymptote (Fig. 1). Once this asymptote has been reached, estimation methods are no longer of interest. To evaluate the performance of the estimators before the accumulation curve approaches the asymptote, we arbitrarily ended our sampling effort as soon as the accumulation curve came within 5% of the asymptote (Fig. 1). This is necessary so as not to include the infinitely long horizontal part of the accumulation curve which exists beyond the point when all species have been recorded. We then divided the curve into 4 parts of equal sampling effort (Fig. 1). Within the first quarter, the sampling effort expended is usually too low to yield reliable estimates. Within the fourth quarter, estimators usually approach the asymptote so closely that their performance is barely different. Therefore, we evaluated estimator performance in the second and third quarter. Basically, we wanted to determine which estimator works best when the species accumulation curve is still increasing and nowhere near the asymptote.

Definitions are as follows: E_j is the estimated species richness, A_j is the total species richness (= the asymptote), and n is the number of sampling units (i.e. host individuals sampled). Within the second and third quarter, we calculated the following performance measures. As measures of bias, we used: (1) BIAS = $\Sigma([E_j - A_j]/[A_j n])$ with $j = 1$ to $j = n$. Note that A_j is a constant when estimators are evaluated within 1 community, but not necessarily when they are evaluated across communities. This measure is equivalent to Palmer's (1990) mean deviation (MD) except that it is scaled by dividing the equation by the asymptotic value A_j . It is also equivalent to Baltanás' (1992) estimate PAR = $100 E_j/A_j$, except that our measure is divided by 100 and has 1 subtracted as a constant. (2) %OVER: another measure of bias is the percentage of overestimates. If the estimator always overestimates A_j , it will have positive bias and 100% overestimates, and if it always underestimates A_j , it will have negative bias and 0% overestimates. An unbiased estimator returns zero bias and 50% overestimates. As measures of precision, we used: (3) DEVIATION = $\Sigma([E_j - A_j]^2/[A_j^2 n])$ with $j = 1$ to $j = n$, which is equivalent to Palmer's (1990) mean square proportional deviation (MSPD). Deviation could also

Table 1. Performance of 10 estimators for 5 real data sets, averaging data for both the second and third quarter

(For each data set, the first column indicates the precision ranking of the estimators with the top estimator being the most precise. The second and third columns give the bias and deviation values of the respective estimator with some estimators returning undefined values (*und.*). The boldly printed estimators (Chao2 and Jack1) were the overall least biased and most precise estimators using either average ranking scores (not presented) or average bias and deviation values (presented in the last column).)

Woodcreeper parasites	Bias	Deviation	Fish parasites	Bias	Deviation	Frog parasites	Bias	Deviation
Chao2	-0.0304	0.0014	Chao2	0.0089	0.0019	MMRuns	-0.0297	0.0010
Jack1	0.0203	0.0030	Jack1	0.0405	0.0034	MMMean	-0.0358	0.0013
Jack2	0.0715	0.0073	MMMean	-0.0329	0.0037	Jack2	-0.0288	0.0014
Boot	-0.0765	0.0090	Chao1	-0.0375	0.0050	Jack1	-0.0427	0.0021
Chao1	-0.0644	0.0095	Boot	-0.0782	0.0101	Chao1	-0.0518	0.0032
MMMean	0.1112	0.0124	Jack2	0.0934	0.0104	Chao2	-0.0571	0.0036
Sobs	-0.1726	0.0334	MMRuns	0.0551	0.0415	ICE	-0.0653	0.0048
MMRuns	0.2251	0.0545	Sobs	-0.1973	0.0439	Boot	-0.0707	0.0053
ICE	<i>und.</i>	<i>und.</i>	ICE	0.2591	0.0857	Sobs	-0.1009	0.0106
ACE	<i>und.</i>	<i>und.</i>	ACE	<i>und.</i>	<i>und.</i>	ACE	<i>und.</i>	<i>und.</i>

Rabbit parasites	Bias	Deviation	Owl parasites	Bias	Deviation	Overall results	Average bias	Average deviation
Jack1	0.0173	0.0024	MMMean	-0.0410	0.0018	Chao2	-0.0012	0.0023
ICE	0.0204	0.0027	Chao2	0.0348	0.0019	Jack1	0.0099	0.0027
ACE	0.0078	0.0028	Chao1	0.0348	0.0019	Chao1	-0.0162	0.0045
Chao2	0.0374	0.0029	Jack1	0.0142	0.0027	MMMean	-0.0173	0.0056
Chao1	0.0374	0.0029	ACE	0.0671	0.0047	Jack2	0.0582	0.0076
MMRuns	-0.0687	0.0056	Boot	-0.0681	0.0071	Boot	-0.0750	0.0082
MMMean	-0.0885	0.0087	Jack2	0.0585	0.0083	Sobs	-0.1575	0.0287
Boot	-0.0821	0.0094	ICE	0.1319	0.0179	MMRuns	0.0890	0.0448
Jack2	0.0971	0.0104	Sobs	-0.1473	0.0242	ACE	<i>und.</i>	<i>und.</i>
Sobs	-0.01698	0.0316	MMRuns	0.2648	0.1223	ICE	<i>und.</i>	<i>und.</i>

be measured by adding absolute values $|E_j - A_j|$ without squaring them. However, we found it desirable to weigh those estimates more heavily which are far away from A_j . We disagree with Baltanás' (1992) measure of precision (see Appendix). (4) RANGE 5% is the percentage of estimators falling within the range $A_j \pm (A_j/20)$ which translates into a 5% range around the asymptote A_j . A perfect estimator returns zero deviation and 100% estimates falling within the 5% range.

Note that all 4 measures above were divided by n , the number of sampling units falling within each quarter. Therefore, bias and precision values are independent of sample size and can be compared directly. This is important because sample size varies for differently truncated accumulation curves.

The performance of estimators can be evaluated in 2 ways: (1) with increasing sampling effort (i.e. sample size) and (2) across different communities. Baltanás (1992) and Palmer (1990, 1991) evaluated estimators across different communities, but the advent of *EstiMateS*© also allowed the evaluation of estimator performance with increasing sampling effort. Since the performance of each estimator was tested for each of the 270 data sets, we ended up with

270 data points for bias and precision for each estimator. All given values are the averages across all data sets included in the respective analysis.

RESULTS

Real data sets

Chao2 and Jack1 were the overall least biased and most precise estimators (Table 1). On average, Chao2 had a slightly negative bias while Jack1 had a slightly positive bias. Chao2 and Jack1 were among the 4 most precise estimators for each data set with only 1 exception (frog parasites, Table 1). Some estimators yielded good estimates for some kinds of data sets, but rather bad estimates for others (e.g. MMRuns and MMEan). Other estimators had consistently medium-range results (e.g. Chao1). All estimation methods performed better than the number of observed species (Sobs) except ICE and MMRuns which were more biased and less precise than Sobs in several cases. These results were not tested statistically as only 1 data point per data set is available for each estimator. Two estimators, ACE and ICE, often returned undefined values, indicating that their use was not appropriate for these particular data sets (see Discussion section).

Table 2. Performance of 10 estimators for computer-simulated data sets

(All given values are averages across all respective data sets. 'Both quarters' refers to the average of the data from the second and third quarter; 'und.' denotes estimators returning undefined values. Boldly printed results indicate the best numerical performance within the respective category. Superscript letters denote those estimators whose bias and deviation values ($n = 270$) were not significantly different from each other at the $P < 0.01$ significance level.)

Estimator	Second quarter				Third quarter				Both quarters			
	Bias	% Over	Deviation	Range 5 %	Bias	% Over	Deviation	Range 5 %	Bias	% Over	Deviation	Range 5 %
Jack1			0.00117	88.8	0.04271	96.5	0.00275	65.5	0.02123 ^a	75.1	0.00196	77.2
Boot	-0.0003 ^a	53.8	0.00462	55.3	-0.0075	32.7	0.00042	97.5	-0.03122	17.2	0.00252	76.4
Chao1	-0.05490	1.8	0.00259 ^a	74.5	0.03346 ^a	73.6	0.00417 ^b	70.5	0.01583^a	60.0	0.00338 ^a	72.5
Chao2	-0.00179 ^a	46.5	0.00270 ^a	74.4	0.03627 ^a	76.4	0.00452 ^a	67.9	0.01952 ^a	64.0	0.00361 ^a	71.1
Jack2	0.00278 ^a	51.6	0.00361	62.1	0.08247	97.9	0.01043	34.8	0.06181	91.7	0.00702 ^c	48.5
MMMean	0.04116	85.5	0.01152 ^b	28.2	-0.05350 ^b	7.5	0.00515 ^b	49.6	-0.06880 ^b	6.8	0.00830 ^{b,c}	38.9
MMRuns	-0.08410 ^b	6.0	0.01696 ^b	29.2	-0.05127 ^b	7.8	0.00514 ^b	50.6	-0.06390 ^b	7.2	0.01105 ^b	39.9
Sobs	-0.07653 ^b	6.6	0.01696 ^b	29.2	-0.05127 ^b	7.8	0.00514 ^b	50.6	-0.06390 ^b	7.2	0.01105 ^b	39.9
ACE	-0.16280	0.0	0.2911	0.0	-0.09538	0.0	0.00951	0.0	-0.12909	0.0	0.01931	0.0
ICE	und.	und.	und.	und.	und.	und.	und.	und.	und.	und.	und.	und.
	und.	und.	und.	und.	und.	und.	und.	und.	und.	und.	und.	und.

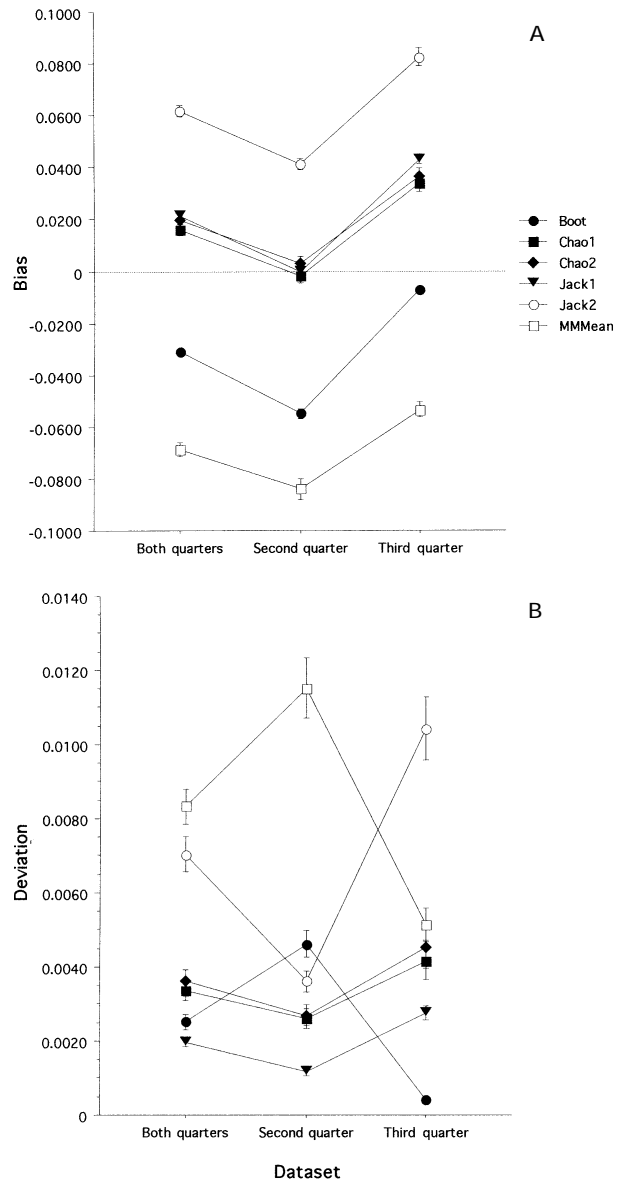


Fig. 4. Plot of (A) bias and (B) deviation of 6 estimators within the second and third quarter of sampling effort, and for both quarters combined. Note that bias values close to zero and low deviation values are associated with good performance. Error bars indicate 1 standard error.

Computer-simulated data sets : increasing sampling effort

Table 2 gives average values for bias and precision measures for each estimator. For the simulated data sets, Sobs performed worse than any of the estimation methods. Three estimation methods also performed badly. MMRRuns was very biased and imprecise. ACE and ICE, on the other hand, returned undefined values for most simulated data sets, again indicating that their use was not appropriate for these data sets (see Discussion section). Therefore, these 4 estimators were excluded from further consideration.

Performance rankings are based on statistically significant differences (Table 2). For the second

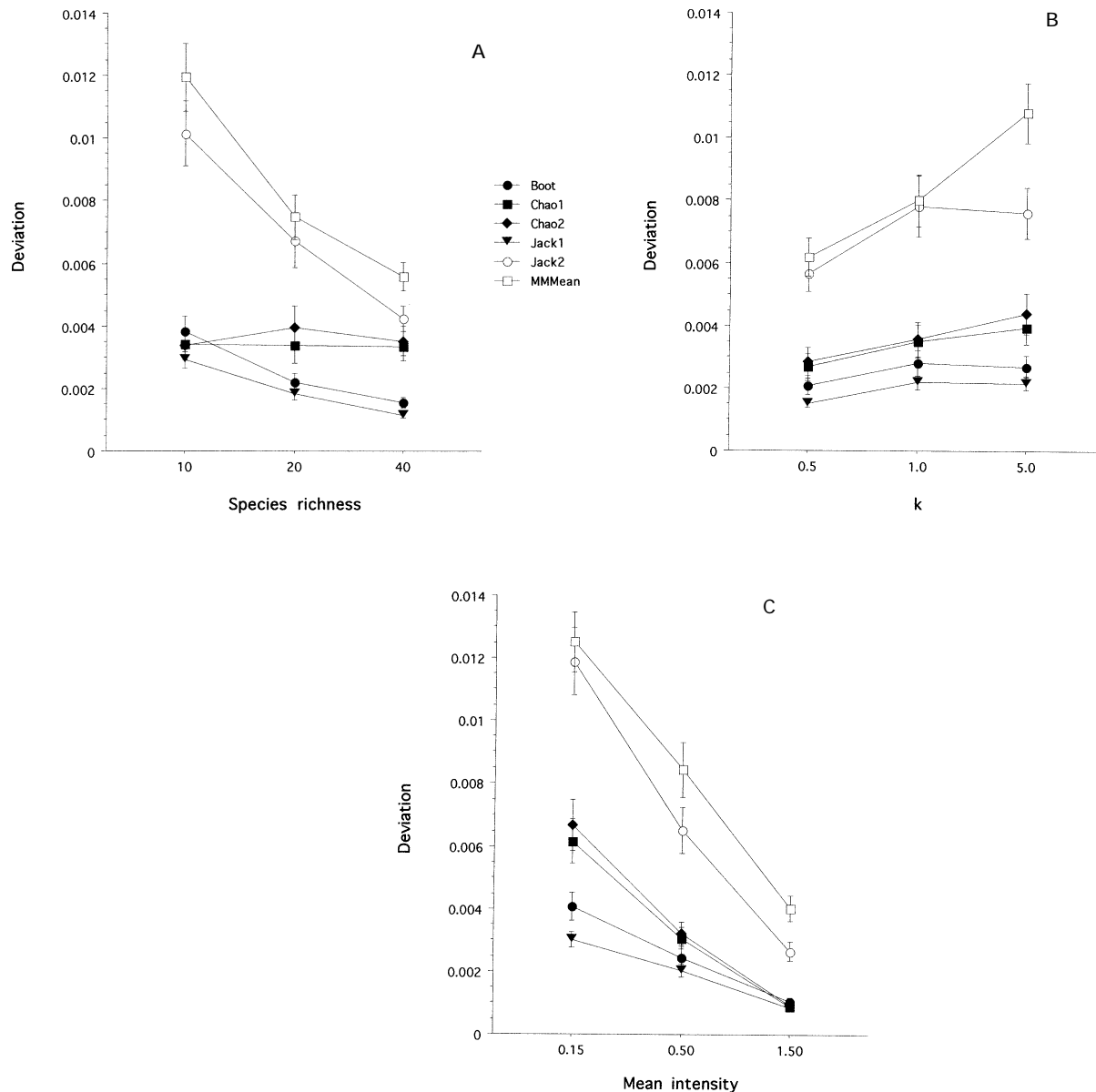


Fig. 5. Plot of estimator deviation versus (A) species richness, (B) aggregation parameter k , and (C) mean intensity, combining data for both quarters. Error bars indicate 1 standard error.

quarter, Chao1, Chao2 and Jack1 were the least biased estimators while Jack1 was the most precise estimator. For the third quarter, Boot was the least biased and most precise estimator. Averaging data from both the second and third quarter, Chao1, Chao2 and Jack1 were the least biased estimators (all with a small positive bias) while Jack1 was the most precise estimator. Boot, Chao1, Chao2 and Jack1 outperformed all the other estimators except for Boot being only the fifth best estimator in the second quarter (Fig. 4).

Computer-simulated data sets : across model communities

Since results for bias and precision were similar in the following analyses (i.e. did not change the relative ranking of the estimators), only results for the latter

are reported. Model communities differed in 3 parameters: total species richness, the aggregation parameter k , and mean intensity (which is equivalent to the average population size within each host, see Materials and Methods section).

Increasing parasite species richness increased the precision of most estimators, with the exception of the Chao estimators (Fig. 5A). The ranking among estimators did not change with increasing species richness with the notable exception of improved precision of Boot at higher species richness relative to the precision of the Chao estimators.

Increasing aggregation of parasites among hosts (i.e. decreasing k) caused all estimators to increase in precision (Fig. 5B). Again, the ranking among estimators did not change with increasing aggregation.

Increasing mean intensity led to pronounced

improvement in the precision of all estimators (Fig. 5C). While a clear ranking of estimators was evident at small mean intensity, the estimators Boot, Chao1, Chao2 and Jack1 did not differ in precision at higher mean intensities.

These results indicate that, for the simulated data sets, the precision values of the estimators Boot, Chao1, Chao2 and Jack1 are less affected by changes of parameter settings than the values of the other estimators (Fig. 5).

DISCUSSION

The non-parametric bootstrap (Boot), Chao1, Chao2, and first-order jackknife (Jack1) estimators had the best overall performance for the computer-simulated data sets (Table 2). These estimators were less biased, more precise, and less affected by changes of parameter settings than the other examined methods. The performance of Boot and Chao1, however, was consistently worse than that of Chao2 and Jack1 for the real data sets (with 2 minor exceptions: see frog and owl parasites, Table 1). Thus, for any study of parasite species richness, the Chao2 or the Jack1 estimator should be used to control for the confounding effects of sampling effort on estimates of total species richness.

Other studies support this recommendation. Baltanás (1992) found Jack1 to be the least biased estimator when compared to 2 other estimation methods, and Palmer (1990, 1991) found that Jack1 was the most precise and the second least biased estimator when compared to 6 other estimation methods. Baltanás and Palmer did not test the Chao estimators, but Colwell & Coddington (1994) and Chazdon *et al.* (1997) recommended the Chao2 estimator based on the analysis of seed bank and tree seedling data sets. Chazdon *et al.* (1997) specifically found that the Chao2 estimator is much less sensitive to aggregated data sets than the Chao1 estimator.

Overall, neither of the 2 accumulation curve models performed well, although MMMean was the fourth best estimator for the real data sets, and MMRuns and MMMean were the best estimator for the frog and owl parasites, respectively (Table 1). Perhaps others published accumulation curve models (Raaijmakers, 1987; Palmer, 1990; Baltanás, 1992; Bunge & Fitzpatrick, 1993; Soberón & Llorente, 1993; Colwell & Coddington, 1994; Walther *et al.* 1995; Chazdon *et al.* 1997) would do better and should be tested comparatively. Nevertheless, *EstiMateS*© includes almost every available non-parametric estimation method (R. K. Colwell, personal communication), so that the results of this study are a comprehensive comparison of the performance of non-parametric estimators.

Almost all estimation methods performed better than the number of observed species (Sobs). Sobs is always a negatively biased estimate of the total

species richness and very imprecise unless sampling effort has been exhaustive. Estimation methods which perform worse than Sobs (e.g. ICE and MMRuns, Table 1) clearly fail the grade.

The new estimators ACE and ICE (*EstiMateS*©, version 4.1) did not perform well. Only in 1 case did they perform well (rabbit parasites, Table 1), otherwise they performed badly or returned undefined values. These estimators might not be appropriate for the analysis of data sets based on parasite abundance distributions because their calculation requires too many classes of rare species (≥ 20). When used with data sets of higher species richness, ACE and ICE performed much better, with ICE actually outperforming all other estimators (Chazdon *et al.* 1997). However, these estimators also have the unpleasant characteristic that the number of classes of rare species which are used in their calculation is not fixed but can be varied. For several real data sets, we increased the number of classes from 5 to 20 in increments of 5, which caused species richness estimates of ACE and ICE to vary by about 5% and to return fewer undefined values (but usually still too many for proper performance evaluation).

Furthermore, the calculation of these estimation methods is very complicated while the calculation of Chao2 and Jack1 is relatively straightforward as it only requires knowledge of the number of observed species, the sample size and the number of species which occur in exactly 1 and exactly 2 samples. Finally, *EstiMateS*© calculates a variance estimate only for Chao1, Chao2 and Jack1. This variance estimate can be used to calculate a confidence interval attached to the species richness estimate (Heltshe & Forrester, 1983; Chao, 1987; Krebs, 1989). All other estimation methods calculated by *EstiMateS*© lack such a variance estimate, although future versions of *EstiMateS*© may incorporate such estimates (R. K. Colwell, personal communication).

The analyses indicate that estimators perform better in species-rich sampling communities with large populations whose individuals are aggregated among samples. This observation contradicts Baltanás' (1992) results findings for 1 parameter in that he found better performance in species-poor sampling communities. This contradiction may be due to Baltanás' use of 2 estimation methods not tested in this study (a curve model and a fit of a log-normal distribution) or his use of bias instead of precision for performance evaluation (see Appendix). Differential estimator performance may also have resulted from his use of a different model community derived from the log-normal model (Preston, 1948).

The presented results are specific to the analysis of data sets with species-abundance distributions typical of parasite populations. Estimator performance cannot be extrapolated to the analysis of *other* species-abundance distributions as estimator per-

formance depends on the species–abundance distribution of the data set being analysed (Bunge & Fitzpatrick, 1993; Soberón & Llorente, 1993; Colwell & Coddington, 1994; Walther *et al.* 1995). Thus, for different species–abundance distributions, different estimators may perform better than Chao2 and Jack1. Although the presented results are strictly speaking only applicable to parasite communities, they achieve wider importance when compared with other similar studies. Analyses of data sets of plant and bird communities suggest that the 2 recommended estimators Chao2 and Jack1 do well for a wide variety of ecological communities, usually being among the best estimators (Palmer, 1990, 1991; Colwell & Coddington, 1994; Chazdon *et al.* 1997; B. A. W. and J.-L. Martin, unpublished results). Nevertheless, further analyses of both real and simulated data sets are required to establish the most reliable estimators for a wide variety of ecological communities, as this problem is far from resolved (Colwell & Coddington, 1994).

To not further complicate the presentation of the results, we chose to set the patchiness parameter at default, thus leaving the original patchiness of the data unaffected. The effect of data patchiness on estimator performance is certainly another area which should be explored as the patchiness of the data affects the species–abundance distributions (Chazdon *et al.* 1997).

Both Palmer (1990, 1991) and Baltanás (1992) used r^2 values of the regression between estimated and total species richness to evaluate the ability of estimators to reliably rank communities according to their total species richness. However, estimated species richness is dependent on sampling effort (see Results section), and therefore, such a ranking should only be done at constant sampling effort. An arbitrary level of sampling effort could be chosen for this kind of analysis, but overall analysis of bias and precision over a large range of sampling effort appeared to be the more comprehensive approach.

However, even the presented results are to some extent dependent on the pre-determined choice of the boundaries of the quarters (Fig. 1), although their choice appears reasonable from a practical point of view (see Materials and Methods section). In the end, comparative performance is always dependent on the definition of performance. For example, Chazdon *et al.* (1997) used a different definition of performance which is not based on bias and precision, but on the sensitivity of the estimator to increases in sample size and data patchiness. This approach, however, is questionable in that a stable estimator (whose estimates do not change with increasing sample size or data patchiness) may still yield a consistently biased or imprecise estimate. Researchers should always compare the species accumulation curve with the species richness estimates to check for possible inconsistencies.

Reliable and accurate estimates of total species richness are important to researchers in various fields. The estimation methods used in this study may represent some of the most powerful statistical tools to derive such estimates even when sampling is nowhere near complete. Furthermore, researchers should note that these methods also allow a *post hoc* check whether the species accumulation curve (Sobs) has indeed reached its asymptote by observing whether Sobs values have converged with Chao2 values, for example.

Potential uses include studies of species richness and diversity (Cornell & Washburn, 1979; Gregory, 1990; Coddington, Young & Coyle, 1996; Siemann, Tilman & Haarstad, 1996), biogeographic patterns (Rahbek, 1995), biodiversity assessment and monitoring (Coddington *et al.* 1991; Gardner & Campbell, 1992; Colwell & Coddington, 1994) and global species richness assessment (May, 1994). Since many conservation biologists have recently shifted their focus from single-species conservation to entire ecosystems and landscapes (Franklin, 1993), reliable species richness estimates will be important in determining areas of highest conservation priority. Furthermore, in many biological communities, indices such as species richness, abundance and diversity may be correlated with each other (Southwood, Moran & Kennedy, 1982; Stork, 1991; Siemann *et al.* 1996; Simberloff & Moore, 1997; Chazdon *et al.* 1997). Therefore, accurate estimates of species richness may be reliable correlates of species abundance and diversity which are usually more difficult to estimate than species richness (Simberloff & Moore, 1997).

It is straightforward to set total species richness in a computer-simulated community, but a number of practical problems remain when defining total species richness in the real world. At any instant, species richness within a given area is a finite number. However, sampling inevitably is a continuous exercise, and total species richness usually increases with the time-interval of sampling as unrecorded species continuously wander into the sampled region. Also, the investigator has to define which species are actually biologically meaningful in the research context. For example, the total species richness of a region could be defined as all breeding species or as all species present. For birds or mammals, this distinction may be possible, but for most other taxa (e.g. parasites) it may not be possible to determine whether their presence is actually meaningful in the investigated biological context. Sampling for long time-periods and including every recorded species may thus lead to over-sampling (Harrison & Martinez, 1995; Walther *et al.* 1995; Elphick, 1997).

The appropriate use of estimation methods requires presence/absence or abundance data for each taxon sampled and the establishment of a unit of

sampling effort. The choice of the former depends on the methodology used. The latter requires keeping records of sampling units separately. Sampling units may be fixed numbers of individuals or species examined or encountered (sampling-effort-dependent), equal time-intervals (time-dependent), or standardized substrate samples (space-dependent). Individual faecal samples, hosts, or areas are also space-dependent sampling units. The choice of sampling unit again is a practical field problem.

The next practical problem is to determine the number of sampling units which will yield a reliable species-richness estimate. Two approaches appear reasonable. Researchers could continue sampling until the variance estimate of either Jack1 or Chao2 falls below a threshold which was set before sampling was begun (e.g. stop sampling once the variance is less than 5% of the estimated species richness). Alternatively, Chazdon *et al.* (1997) suggested to sample a representative community, and then to select a sample size which incorporates a pre-determined portion of the total species richness (as derived from the species-richness estimate). The selected sample size could then be used to comparatively sample other similar communities. Further practical suggestions, e.g. how to estimate the accumulation rate of new species, have been given by Coddington *et al.* (1991, 1996) and Soberón & Llorente (1993).

Although practical problems will remain for the proper use of estimation methods, the theoretical framework for comparative evaluation of estimator performance has been presented and tested in this study. Further research will hopefully identify reliable and accurate estimation methods for a wide variety of communities.

We thank Bridget Appleby, Juan Carlos Casanova, Dale Clayton, Patrick Durand, Elizabeth Faliex, Jean-Louis Martin and Pierre Sasal for providing data. We thank Angel Baltanás, Dale Clayton, Robert Colwell, Clive Hamblen, Sean Nee, Mike Packer, Mark Pagel, Jorge Soberón Mainero, Richard Southwood and 2 anonymous referees for comments and discussion. Robert Colwell provided the unpublished richness estimator program *EstiMateS*©, versions 3.1 and 4.1. We owe special thanks to him and Mark Pagel for help and encouragement, and to Carsten Rahbek for starting it all. B.A.W. has been supported by an Evan Carroll Commager Fellowship and a John Woodruff Simpson Fellowship granted by Amherst College, MA, USA.

APPENDIX

Definitions are as follows: E_j is the estimated species richness, A_j is the asymptote of total species richness, and n is the number of sampling units, with $j = 1$ to $j = n$. Baltanás (1992) evaluated the performance of estimators by calculating the ratio called PAR = $u = 100 E_j/A_j$ which is a measure of bias (see Materials and Methods section). He also measured the dispersion of the PAR estimates around

the mean by calculating the standard deviation of the PARs = $\sqrt{\Sigma[u - \bar{u}]^2/n}$ which can be algebraically transformed into

$$\begin{aligned} &= \sqrt{\frac{100}{n A_j}} \times \sqrt{\Sigma((E_j - (\Sigma E_j/n))^2)} \\ &= \text{constant} \times \sqrt{\Sigma\{(E_j - (\Sigma E_j/n))^2\}}. \end{aligned}$$

It is evident from this formula that Baltanás' measure depends solely on the calculated values of the estimated species richness, not on the distance between estimated and total species richness. It is thus a measure of the closeness of repeated measurements of the same quantity (to add confusion, this measure is called the 'precision of a measurement', but it is not a measure of the 'precision of a statistic', see Zar (1996)). Precision, as we define it, is not measured by either PAR or the standard deviation of the PARs.

To illustrate, assume that total species richness A_j is 100, and, that in the first case, estimated species richness E_j is 90 and 95, and in the second case, it is 95 and 110. In both cases, precision should be equal (using our formula of deviation, it is 125/20000 for both). In the first case, mean PAR is 92.5 and the value for the standard deviation is $\sqrt{12.5/2}$. In the second case, mean PAR is 102.5 and the value for the standard deviation is $\sqrt{112.5/2}$.

REFERENCES

- ANDERSON, R. M. & MAY, R. M. (1978). Regulations and stability of host-parasite population interactions. *Journal of Animal Ecology* **47**, 219–242.
- ANDERSON, R. M. & MAY, R. M. (1985). Helminth infections of humans: mathematical models, population dynamics and control. *Advances in Parasitology* **24**, 1–101.
- 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**, 623–633.
- BURNHAM, K. P. & OVERTON, W. S. (1979). Robust estimation of population size when capture probabilities vary among animals. *Ecology* **60**, 927–936.
- CHAO, A. (1984). Non-parametric 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. & LEE, S.-M. (1992). Estimating the number of classes via sample coverage. *Journal of the American Statistical Association* **87**, 210–217.
- 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. (1997). Statistical methods for estimating species richness of woody regeneration in

- primary and secondary rain forests of NE Costa Rica. In *Forest Biodiversity in North, Central and South America and the Caribbean: Research and Monitoring* (ed. Dallmeier, F. & Comiskey, J.). Parthenon Press (in the Press).
- CLAYTON, D. H., GREGORY, R. D. & PRICE, R. D. (1992). Comparative ecology of Neotropical bird lice (Insecta: Phthiraptera). *Journal of Animal Ecology* **61**, 781–795.
- CODDINGTON, J. A., GRISWOLD, C. E., DAVILA, D. S., PENARANDA, E. & LARCHER, S. F. (1991). Designing and testing sampling protocols to estimate biodiversity in tropical ecosystems. In *The Unity of Evolutionary Biology: Proceedings of the Fourth International Congress of Systematic and Evolutionary Biology* (ed. Dudley, E. C.), pp. 44–60. Dioscorides Press, Portland, Oregon.
- CODDINGTON, J. A., YOUNG, L. H. & COYLE, F. A. (1996). Estimating spider species richness in a Southern Appalachian cove hardwood forest. *Journal of Arachnology* **24**, 111–128.
- COLWELL, R. K. & CODDINGTON, J. A. (1994). Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London, B* **345**, 101–118.
- CORNELL, H. V. & WASHBURN, J. O. (1979). Evolution of the richness-area correlation for cynipid gall wasps on oak trees: a comparison of two geographic areas. *Evolution* **33**, 257–274.
- DOBSON, A. P. & KEYMER, A. E. (1990). Population dynamics and community structure of parasite helminths. In *Living in a Patchy Environment* (ed. Shorrocks, B. & Swingland, I. R.), pp. 107–125. Oxford: Oxford University Press.
- ELLIOT, J. M. (1977). Statistical analysis of samples of benthic invertebrates. *Freshwater Biological Association, Scientific Publication* **25**, 1–156.
- ELPHICK, C. S. (1977). Correcting avian richness estimates for unequal sample effort in atlas studies. *Ibis* **139**, 189–190.
- FRANKLIN, J. F. (1993). Preserving biodiversity: species, ecosystems, or landscapes? *Ecological Applications* **3**, 202–205.
- GARDNER, S. L. & CAMPBELL, M. L. (1992). Parasites as probes for biodiversity. *Journal of Parasitology* **78**, 596–600.
- GREGORY, R. D. (1990). Parasites and host geographic range as illustrated by waterfowl. *Functional Ecology* **4**, 645–654.
- HARRISON, J. A. & MARTINEZ, P. (1995). Measurement and mapping of avian diversity in southern Africa: implications for conservation planning. *Ibis* **137**, 410–417.
- HELTSHE, J. H. & FORRESTER, N. E. (1983). Estimating species richness using the jackknife procedure. *Biometrics* **39**, 1–11.
- KREBS, C. J. (1989). *Ecological Methodology*. Harper and Row, New York.
- LEE, S.-M. & CHAO, A. (1994). Estimating population size via sample coverage for closed capture-recapture models. *Biometrics* **50**, 88–97.
- MAY, R. M. (1994). Conceptual aspects of the quantification of the extent of biological diversity. *Philosophical Transactions of the Royal Society of London, B* **345**, 13–20.
- 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.
- PEET, R. K. (1974). The measurement of species diversity. *Annual Review of Ecology and Systematics* **5**, 285–307.
- POULIN, R. (1993). The disparity between observed and uniform distributions: a new look at parasite aggregation. *International Journal for Parasitology* **23**, 937–944.
- PRESTON, F. W. (1948). The commonness, and rarity, of species. *Ecology* **29**, 254–283.
- RAAIJMAKERS, J. G. W. (1987). Statistical analysis of the Michaelis-Menten equation. *Biometrics* **43**, 793–803.
- RAHBEK, C. (1995). The elevational gradient of species richness: a uniform pattern? *Ecography* **18**, 200–205.
- SASAL, P., FALIEUX, E. & MORAND, S. (1996). Parasitism of *Gobius bucchichii* Steindachner 1870 (Teleostei, Gobiidae) in protected and unprotected marine environments. *Journal of Wildlife Diseases* **32**, 607–613.
- SHAW, D. J. & DOBSON, A. P. (1995). Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. *Parasitology* **111**, S111–S133.
- SIEMANN, E., TILMAN, D. & HAARSTAD, J. (1996). Insect species diversity, abundance and body size relationships. *Nature, London* **380**, 704–706.
- SIMBERLOFF, D. & MOORE, J. (1997). Community ecology of parasites and free-living animals. In *Host-Parasite Evolution: General Principles and Avian Models* (ed. Clayton, D. H. & Moore, J.), pp. 174–197. Oxford: University Press, Oxford.
- 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.
- SOUTHWOOD, T. R. E. (1978). *Ecological Methods: With Particular Reference to the Study of Insect Populations*. London: Chapman and Hall.
- SOUTHWOOD, T. R. E., MORAN, V. C. & KENNEDY, C. E. J. (1982). The richness, abundance and biomass of the arthropod communities on trees. *Journal of Animal Ecology* **51**, 635–649.
- STORK, N. E. (1991). The composition of the arthropod fauna of Bornean lowland rain forest trees. *Journal of Tropical Ecology* **7**, 161–180.
- WALTHER, B. A., COTGREAVE, P., GREGORY, R. D., PRICE, R. D. & CLAYTON, D. H. (1995). Sampling effort and parasite species richness. *Parasitology Today* **11**, 306–310.
- ZAR, J. H. (1996). *Biostatistical Analysis*. Upper Saddle River, New Jersey: Prentice Hall International.