

# Estimating marine species richness: an evaluation of six extrapolative techniques

Andrew Foggo<sup>1,\*</sup>, Martin J. Attrill<sup>1</sup>, Matthew T. Frost<sup>1</sup>, Ashley A. Rowden<sup>1,2</sup>

<sup>1</sup>School of Biological Sciences, University of Plymouth, Drake Circus, Plymouth PL4 8AA, United Kingdom

<sup>2</sup>Present address: Marine Biodiversity Group, National Institute of Water and Atmospheric Research, PO Box 14-901, Kilbirnie, Wellington, New Zealand

**ABSTRACT:** The number of species in an assemblage at a given point in time is a fundamental property of ecological systems, yet it is hard to quantify for many marine systems. We studied the performance of 6 techniques ('estimators') for extrapolating species richness from limited numbers of samples, using 3 datasets for which an absolute value for total species richness could be determined. We propose that the ideal estimator should always slightly overestimate species richness compared to any observed maximum species richness derived from sampling, as sampling error will always lead to underestimation of true richness. We quantified performance of the estimators relative to the sampled total richness in the assemblage across a range of efforts up to 80% of that required to achieve the asymptote of the species accumulation. We used 3 measures: bias (mean deviation of an estimate from the known richness), precision (variance of repeated estimates based upon a subset of the available pool of samples), and overall accuracy (a combination of bias and precision). No single estimator performed best in all cases, and estimator performance was affected by sampling effort. The estimator *Chao1* performed best at intermediate sampling efforts, with *LAG S<sub>∞</sub>* also performing well at high relative effort. *S<sub>∞</sub>* consistently underestimated, whilst *Chao2* and *ICE* both overestimated and displayed poor precision and accuracy, especially at intermediate sampling efforts and in datasets with uneven patterns of species incidence. Species abundance and incidence amongst samples of a dataset were shown to affect performance of most of the estimators, with the exception of the recently proposed *S<sub>∞</sub>* family of techniques. We conclude that *Chao1* represents the best compromise choice of estimator, and that such nonparametric techniques may represent useful tools for rapid estimation of species richness for some marine assemblages, based on limited sampling effort.

**KEY WORDS:** Biodiversity · Extrapolation · Accumulation · Sampling

Resale or republication not permitted without written consent of the publisher

## INTRODUCTION

Species richness is a fundamental property of any biotic assemblage. Quantifying species richness is a valuable means of characterising assemblages, either for use as a biotic index of the condition of the ecosystem or as a summary variable used in studies of other ecosystem properties or processes (Gaston 1996). However, inventories of richness are costly, are time consuming, and demand enormous resources in terms of taxonomic expertise (Gaston & May 1992). This problem is probably more acute in marine ecosystems

than in any other (Grassle et al. 1991), due to high phyletic diversity coupled with a comparatively limited knowledge of groups such as porifera. Therefore any technique that offers a rapid solution to quantification of richness based on limited sampling effort may have a great deal to offer the field of marine ecology.

A set of techniques has been developed to extrapolate species richness in a discrete assemblage from limited numbers of replicate samples. The theoretical underpinning of these techniques lies in fundamental research into the nature of species-area curves (Arrhenius 1921, 1923, Gleason 1922, Kilburn 1966). In excess

of 20 different techniques have been described that can be used to estimate total richness from limited numbers of samples. The most frequently encountered techniques of species richness estimation ('estimators') found throughout the literature belong to 4 different categories: (1) extrapolations of species-area curves; (2) fitting of species-abundance distributions; (3) modelling species accumulation curves; (4) non-parametric techniques (Colwell & Coddington 1994, Walther et al. 1995, Walther & Morand 1998). Taxa for which comparative studies of the performance of these estimators has been attempted include plants in both tropical and temperate forests (Palmer 1990, Condit et al. 1996, Hellman & Fowler 1999), parasites of vertebrates (Walther & Morand 1998), zooplankton (Dumont & Segers 1996, Arnott et al. 1998), tropical freshwater invertebrates (Melo & Froehlich 2001), and marine macrobenthos (Rumohr et al. 2001). However, the validity of the evaluation techniques employed by these studies varies greatly, depending upon the ease with which accurate values for total assemblage richness can be determined.

The evaluation of any technique aimed at determining the total species richness of an assemblage from limited sampling effort depends upon an accurate value for the total number of species in the assemblage, as this is the yardstick by which all low-effort extrapolations must be judged. Whilst figures for total species richness might be only fleetingly correct, thanks to immigration, emigration and local extinction, such figures in any one period of sampling do provide the best means of judging an estimate. Total species richness of a sampled site may be approximated from a number of replicate samples by plotting the curve of the mean cumulative number of species encountered in incrementally aggregated samples ( $S_{\text{obs}}$ ) over a large number of randomised permutations of the sample aggregation. The asymptote of this curve gives the total observed richness,  $S_{\text{tot}}$ . But, however efficient sampling may be, there will always be a probability of  $S_{\text{tot}}$  under-representing the actual species richness at the site. For that reason we have chosen to refer to the measured maximum richness as  $S_a$ , for *approximated* total species richness.

Establishing an absolute value for species richness of an assemblage is therefore costly and difficult if not essentially impossible using conventional sampling techniques, which begs the question of whether it is a truly worthwhile exercise? After all, various indices of both richness and diversity have been developed (see Magurran 1988, Clarke & Warwick 2001) that aim to facilitate comparisons across assemblages and sites and to reduce errors due to sampling bias. However, even widely accepted and applied indices such as Simpson's  $D$  and the Shannon-Wiener  $H'$  can be more

sensitive to sampling effort effects than might be desirable in ranking of habitats (Lande et al. 2000, Rumohr et al. 2001, Foggo et al. unpubl.). In addition, whilst the use of indices of diversity remains commonplace (although compare the approach adopted by Gaston & Spicer 1998, and comments by Rosenzweig 1995 and Lande et al. 2000), they may nonetheless be criticised on the basis of the *a priori* assumptions they make about the relative importance of spread of abundance amongst species. The influence of this *a priori* weighting can result in inconsistent rankings of a set of habitats depending upon which index is used or how many samples are taken (Foggo et al. unpubl.). By contrast, a simple expression of site total richness makes no such *a priori* assumptions, and it is not a product of operator choice as are the available diversity and richness indices. Total species richness therefore represents a parsimonious alternative to diversity and richness indices, and estimations or extrapolations of total richness might in turn be reliable best approximations for actual total richness.

In this paper we review the performance of 6 methods of estimating total species richness when applied to aquatic datasets for which an accurate value for total richness can be derived from a mean randomised species accumulation curve. We consider the accuracy and precision of the estimators, consider which properties of the sampled assemblages affect their performance, and review their applicability and ease of use.

## MATERIALS AND METHODS

**Choice of datasets for study.** We screened 32 datasets describing discrete marine and estuarine assemblages for which a large number of samples had been taken. These datasets came from both the published literature and our own studies. In each case we employed the software 'EstiMateS' Ver. 5 (Colwell 1997) to produce a mean species accumulation curve over 100 randomised iterations. We sought datasets for which a viable estimate for total species richness was available by applying the following 3 *a priori* criteria:

- The dataset had to show a clear asymptote with at least the final 2 values being identical.
- The variance of the mean estimate of  $S_{\text{obs}}$  at the asymptotic point defined above must be 0.
- The final 20% of the values of the curve preceding the first asymptotic sample had to be within 5% of the asymptotic value.

Out of the 32 datasets examined, only 3 satisfied the criteria, all of which derived from our own studies.

In August 2000 we collected 60 samples of sandy beach macrofauna comprising 15 samples from each of

4 open-coast, intermediate macrotidal sandy beaches on the south coast of the border between Devon and Cornwall, UK. Sampling was adapted from Larsen & Dogget (1990), with the beach being divided into 4 zones: lower shore, lower middle shore, upper middle shore and upper shore. We excluded the upper shore from our sampling plan, as terrestrial vagrants influence the fauna there considerably (M. Frost unpubl. data). Five  $\times$  0.1 m<sup>2</sup>, 24 cm deep box cores were taken at random within each of the 3 remaining height bands on each beach (15 samples per beach); this methodology should be adequate to effectively sample the diversity of each beach (Jaramillo et al. 1995). Sediment was sieved through a 1 mm mesh, and the retained portion was preserved with formalin before being returned to the laboratory for identification. Twenty Eight species were recorded after an average of 59 aggregated samples, giving a reliable stable estimate at 60 samples. This dataset is referred to hereafter as 'Beach macrofauna'.

In 1999 we collected a set of 40 samples of sublittoral benthic sediment from Rainham Creek, a tidal tributary of the Thames Estuary, UK (Attrill et al. 2000). Sampling depended upon depth of water, which varied between 0.3 and 2.5 m. Eighteen samples from water shallower than 0.75 m were collected using a 200 cm<sup>2</sup> surber sampler; the remaining 22 samples were collected using a 200 cm<sup>2</sup> Ekman grab. Samples were sieved (500  $\mu$ m mesh) and then preserved in formalin. Twelve species of oligochaetes were identified, with the mean species accumulation curve for the dataset reaching an asymptote after 39 samples, and a reliable estimate therefore using all 40 samples. This dataset is hereafter referred to as 'Estuarine oligochaetes'.

Between June and August 1999, 112 modified line intercept transect surveys of butterfly fish (Chaetodontidae) were conducted along the outer slope of the fringing reef lying to the west of Pulau Hoga, part of the Tukang Besi peninsula, south-east of the Island of Sulawesi, Indonesia (R. M. Burton unpubl. data). The surveys, following the method of English et al. (1994), were conducted over 50 m at a standard depth of 3 m below the reef crest, using teams of trained SCUBA divers working with the conservation organisation Operation Wallacea. We used the data from each 50 m survey as a single sample to calculate a species accumulation asymptote of 33 species after 109 samples, giving a nominal sampling effort to guarantee accuracy of 110 samples. This dataset is referred to as 'Reef fish'.

**Homogeneity of the datasets.** Homogeneity, for example, across beaches in the composite 'Beach macrofauna' dataset, is not a prerequisite of effective use of extrapolative techniques for estimating species richness (cf. Rumohr et al. 2001). Randomisation of the

order of addition of samples in a large number of iterations of index calculation obviates the requirement for such homogeneity (note that characteristic 'steps' in species accumulations and extrapolations indicate that this randomisation has not been performed—see Gray 2000), and non-homogeneity is reflected in lower precision and accuracy. However, non-homogeneity of datasets in our study could arise either from between-site differences in the 'Beach macrofauna' dataset or from between-sampling-system variability in the 'Estuarine oligochaetes' dataset. Variation from either source would be reflected in the measures of precision and accuracy used in this study, and therefore we were concerned that samples taken by different sampling systems should be as comparable as possible and that there be no wide disparity amongst assemblages in the samples of any of the datasets in order that these factors should not overly influence the performance of the estimators employed.

To determine whether the 2 potentially heterogeneous datasets used in the calculations contained any discrete sampling-derived sub-groupings that might introduce error into the estimations, we produced ordination plots of each using non-metric multidimensional scaling (n-MDS) in PRIMER (Plymouth Routines In Marine Ecological Research) ver. 5.0, employing the Bray-Curtis index on double square-rooted abundance data (Clarke & Warwick 1994). For the 'Estuarine oligochaetes' and 'Beach macrofauna' data the samples were ordered according to sampling system (all of one system first, then all of the second system) and beach respectively. We then plotted the means and standard deviations of the subdiagonals of the multivariate association matrix used to produce the n-MDS plot for these 2 datasets, with sample lags between 1 and 25 samples for the estuarine dataset and 1 and 30 samples for the beaches (Rumohr et al. 2001). With the samples corresponding to the 2 techniques and the 4 beaches input as 2 discrete blocks, these plots should show either a distinct gradient across the sampling lag axis or a series of steps if the techniques or habitats produce different results in terms of the community sampled.

**The estimators.** We reviewed the literature for candidate estimators that performed well at sampling efforts up to 75% of the level required to sample all the species in the assemblage. If estimation is to be used to reduce the time and effort spent in deriving a value for absolute species richness, then this saving should be as great as possible; hence estimators that perform well at low sampling effort are of greatest interest. *Chao1* (Chao 1984) is an estimator based on the number of species in a sample that are represented only by 1 or 2 individuals, whilst *Chao2* (Chao 1987) applies the same approach to species that occur only in 1 or 2 sam-

ples. *Chao1* is thus an abundance-based estimator, and *Chao2* an incidence-based approach. Both *Chao1* and *Chao2* take the general form

$$S_{\text{tot}} = S_{\text{obs}} + (a^2/2b) \quad (1)$$

where  $S_{\text{tot}}$  is the predicted total species richness;  $S_{\text{obs}}$  is the observed number of species at a given sampling effort;  $a$  is the number of species represented by a single individual (number of singletons) in *Chao1* and the number of species present in exactly 1 sample (unique species) in *Chao2*; and  $b$  is the number of species represented by exactly 2 individuals (number of doubletons) in *Chao1* and the number of species present in exactly 2 samples in *Chao2*.

Chao and her co-workers have also developed new coverage-based richness estimators using a more flexible interpretation of 'rare' and 'infrequent' (Chao & Lee 1992, Chao et al. 1993, Lee & Chao 1994). These techniques require an arbitrary cut-off level to be imposed to determine the abundance or incidence level at which a species is considered to be rare or infrequent, adding an element of subjectivity to the index. However Chazdon et al. (1997) have studied the effects of changing the value of this cut-off, and they recommend a value of 10, which we have employed in this study. Of these 2 estimators *ACE* (abundance-based coverage estimator) and *ICE* (incidence-based coverage estimator), the former was previously found to perform poorly when applied to freshwater benthic assemblages (Foggo et al. unpubl.), whilst the latter performed very well, especially at low relative sampling effort. We therefore chose only *ICE* for evaluation in this study.

In *ICE*, the extrapolated value of  $S_{\text{tot}}$  assumes that the observed number of species in a sample comprises a number of common and a number of infrequent taxa. Thus:

$$S_{\text{tot}} = S_{\text{infrequent}}(n) + S_{\text{common}}(n) \quad (2)$$

where  $S_{\text{infrequent}}(n)$  is the number of species that occur in fewer than  $n$  of the pooled samples and  $S_{\text{common}}(n)$  is the number of species that occur in greater than  $n$  of the samples. *ICE* determines the coverage of moderately infrequent species within the 'infrequent' category as the proportion of all species in the category that are not unique (occurring in only 1 sample), and it estimates the total richness based on the relative proportions of common, infrequent, and unique species respectively.

The next estimator is based upon the series of Jack-knife estimators for mark-recapture estimation of population size developed following the work of Tukey (1958). *Jack* calculates a first-order Jack-knife (Heltshel & Forester 1983), based upon calculation of the number of species that only occur in 1 sample:

$$S_{\text{tot}} = S_{\text{obs}} + Q_1 \left( \frac{m-1}{m} \right) \quad (3)$$

where  $Q_1$  is the number of species that occur in exactly 1 of the pooled samples and  $m$  is the total number of samples (Colwell & Coddington 1994). This estimator performed well at low sampling efforts in studies by Walther & Morand (1998) and Foggo et al. (unpubl.), and it has previously been applied by to a benthic marine dataset (Rumohr et al. 2001).

The final 2 estimators belong to the  $S_{\infty}$  family of estimators developed by Karakassis (1995) and Rumohr et al. (2001). The estimator  $S_{\infty}$  (Karakassis 1995) is derived from the theoretical upper limit of the asymptote of the species-area curve, which is estimated by calculating the point at which 2 successive samples are expected to produce an identical cumulative number of species. The value of  $S_{\text{obs}}$  for  $n$  samples is regressed against  $S_{\text{obs}}(n+1)$ , and the equation of this regression is then solved for the theoretical value of the asymptote of the species accumulation, termed  $S_{\infty}$ , by finding the intercept point of this regression line with the line bisecting the angle of the coordinates axes. Rumohr et al. (2001) explored the potential for developing the  $S_{\infty}$  technique through both truncation of the data series and introducing variable sample lag, i.e. regressing  $S_{\text{obs}}(n)$  against  $S_{\text{obs}}(n+k)$  where  $k = 1 \dots n-1$ . Of these, sample lag produced the greatest improvement in the accuracy of the estimator, but necessitated greater overall number of samples. We therefore chose to investigate the estimator *Lag  $S_{\infty}$* , employing a moderate lag of 10 samples.

**Calculation of the estimators.** We wrote Visual Basic<sup>®</sup> routines to iteratively calculate  $S_{\infty}$  and *Lag  $S_{\infty}$*  following the procedure described by Karakassis (1995) and Rumohr et al. (2001). There are 2 possible ways of deriving variance estimates for the  $S_{\infty}$  family of estimators (Karakassis 1995): either through calculation of the index for each permutation of the species accumulation (which occasionally produces highly variable, often negative estimates) or through pseudo-replication. In this latter case a number of species accumulations is performed to derive a mean value for  $S_{\text{obs}}$  at each level of sampling effort, producing a value for  $S_{\infty}$ ; this process is then repeated a number of times to gain a mean value of  $S_{\infty}$  and hence a variance estimate. Clearly the former process is more demanding and critical of the consistency of the estimator, and the latter has a confounded variance estimate. We nonetheless chose the latter technique to produce our variance estimate because Karakassis (1995) originally intended  $S_{\infty}$  to be calculated using a large number of iterations of the species accumulation to produce mean values of  $S_{\text{obs}}$ , resulting in a single value of  $S_{\infty}$ . The pseudo-replicated technique therefore best approaches an estimate of the variance of  $S_{\infty}$  as it will be

applied in practice. Means and variances of *Chao1*, *Chao2*, *ICE* and *Jack* were calculated using the software 'EstiMateS' Ver. 5 (Colwell 1997).

Because of the differing underlying sources of variation in the modes of calculation of the estimators, we decided upon a number of permutations to perform in each case after first checking the shape of the curve of variance plotted against number of permutations. For the non-parametric estimators, there was no detectable decrease in variance once the number of permutations exceeded 20 to 30 (depending upon the level of sampling effort studied). For the  $S_{\infty}$  estimators, variance of the final mean value of permuted  $S_{\infty}$  declined rapidly with the number of permutations contributing to each individual estimate, stabilising around  $n \approx 50$  permutations. Final variance also declined asymptotically with the number of repeated individual estimates contributing to the mean, stabilising when  $n \approx 70$ . We therefore decided to use 100 permutations of the species accumulation in calculating the non-parametric estimators, but to ensure a critical test of the  $S_{\infty}$  estimators we used only 50 permutations in calculation of each individual  $S_{\infty}$  and *Lag*  $S_{\infty}$  value. This represents half the number of permutations recommended by Rumohr et al. (2001), and it provides for a more critical appraisal of the techniques. We calculated 100 such replicate values of  $S_{\infty}$  and *Lag*  $S_{\infty}$  to produce a mean and variance estimate for each.

In order to approximate performance of the estimators, we calculated mean and variance values for each estimator at every 10th percentile of the sample effort required to produce a reliable species accumulation asymptote, up to a maximum of 80%. The sample lag of 10 for *Lag*  $S_{\infty}$  constrains this ceiling, and in fact for the 'Estuarine oligochaetes' data set the lag employed to calculate *Lag*  $S_{\infty}$  at 80% was only 8 samples. Thus for 'Reef fish', with an asymptotic sampling effort of 110, we calculated mean and variance for the estimators at 11, 22, 33, 44 ... 88 samples.

**Quantifying the performance of the estimators.** We calculated 3 performance indicators following the rationale of Hellmann & Fowler (1999):

*Bias* indicates whether an estimator consistently over- or underestimates the true total species richness:

$$\text{Bias} = \left[ \left( \frac{\sum_{i=1}^n S_i}{n} \right) - S_a \right] \quad (4)$$

where  $S_i$  are individual richness estimates from 1 of the  $n$  permutations and  $S_a$  is the total species richness of the assemblage approximated from the asymptote.

The Monte Carlo variance of the estimator over the 100 permutations provides an indicator of *precision* (how close are repeated estimates based on different species accumulations to one another?):

$$\text{Precision} = \left[ \frac{\sum_{i=1}^n S_i^2 - \left( \frac{\sum_{i=1}^n S_i}{n} \right)^2}{n} \right] / (n-1) \quad (5)$$

The predictability of repeated estimates' proximity to the true species richness provides a measure of overall *accuracy*. We calculated this as mean square error:

$$\text{Accuracy} = \text{variance}(S_{\text{estimated}}) + \left[ \left( \frac{\sum_{i=1}^n S_i}{n} \right) - S_a \right]^2 \quad (6)$$

Because both bias and variability are incorporated into this latter measure of performance, it can be regarded as the most desirable property of an estimator. It also down-weights variability in favour of bias, reducing the influence of the confounded variance estimates from the 2 different forms of estimator calculation.

**Factors affecting performance of the estimators.** We wanted to establish how the properties of the entire assemblage sampled, in terms of diversity, equitability and relative species incidence across samples affect the performance of the estimators. We calculated values for diversity (Shannon-Wiener  $H'$ ), and equitability (Simpson's evenness  $\hat{D}/S_a$ , where  $\hat{D}$  is the inverse form of Simpson's dominance index) for each assemblage. Rank-occupancy distribution was used to characterise the pattern of occurrence of species amongst samples; rank-occupancy in most large datasets approximates to a logarithmic series (Krebs 1999). We calculated the proportional sample occupancy (proportion of all species' occurrences attributable to any individual species) for each species in the entire dataset, and we used these figures to compute a theoretical rank-occupancy distribution based on the logarithmic series (Krebs 1999). Thus the proportion of samples in which each species occurs in nature can be compared to that predicted by theory, and deviation from this theoretical pattern of proportional occupancy indicates that the occurrence of the species is either more or less common than that of the theoretical population. We performed a Kolmogorov-Smirnov test (SPSS ver. 10.0) comparing the observed and theoretical distributions, and we used the calculated value of  $Z$  and its associated probability as a measure of the conformity the pattern of sample occupancy to that expected.

Both rarity and frequency of species' occurrence have been postulated to affect the reliability of species richness estimates and comparisons between assemblages (Gentil & Dauvin 1988, Riddle 1989, Cao et al. 1998, Gotelli & Colwell 2001). We therefore tested for relationships between our performance parameters and the mean number of singleton and unique species encountered at each level of sampling effort studied. We conducted linear regressions for bias, precision and accuracy of each estimator with number of singleton (log transformed) and unique species as predictors.

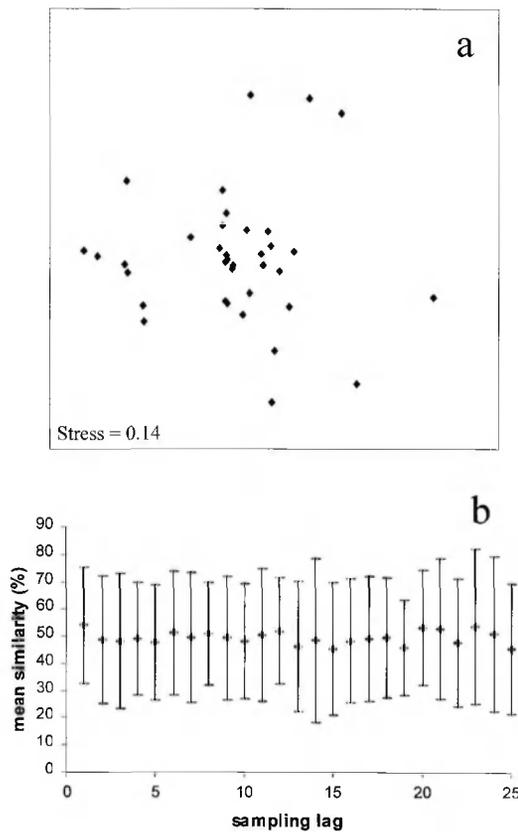


Fig. 1. (a) *n*-MDS ordination plot of the 40 replicate samples in the 'Estuarine oligochaetes' dataset; (b) mean ( $\pm$  SD) values of the multivariate similarity matrix for sample lags from 1 to 25 samples

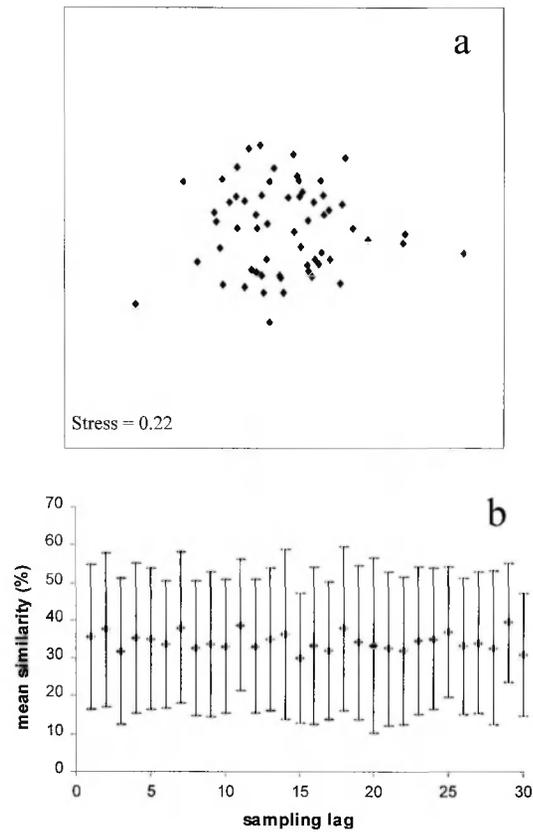


Fig. 2. (a) *n*-MDS ordination plot of the 60 replicate samples in the 'Beach macrofauna' dataset; (b) mean ( $\pm$  SD) values of the multivariate similarity matrix for sample lags from 1 to 30 samples

## RESULTS

The *n*-MDS plots and the means and standard deviations of the subdiagonals of the multivariate association matrix for varying sample lags indicate no lack of homogeneity in either the 'Estuarine oligochaetes' or 'Beach macrofauna' datasets (Figs. 1 & 2), with no apparent clustering of samples or steps in the plots of subdiagonals which would indicate such non-homogeneity.

The 3 datasets represent quite markedly different assemblages (Table 1), and the species accumulation curves approach the asymptote differently (Fig. 3). The 'Reef fish' dataset plateaus early in the species accumulation, but then continues to creep upwards to the asymptote. This dataset is the most equitable in terms of species' abundance, and the least patchy in terms of species' occurrence, with a rank-occurrence profile approaching deviation from the logarithmic prediction (Kolmogorov-Smirnov  $Z = 1.231$ ,  $n = 66$ ,  $p = 0.097$ ). The 'Estuarine oligochaetes' dataset was depauperate in comparison with the other 2, and accordingly less

diverse. It was also the patchiest in terms of species occurrence (Kolmogorov-Smirnov  $Z = 0.816$ ,  $n = 24$ ,  $p = 0.518$ ), indistinguishable from the theoretical logarithmic distribution. In fact half of the species only occurred in 2 of the 40 samples, although 2 of these species only co-occurred in the same sample twice. This patchiness may be expected to pose a challenge to the incidence-based estimators, certainly in terms of

Table 1. Shannon-Wiener diversity ( $H'$ ), Simpson's evenness ( $\hat{D}/\hat{S}$ ) and deviation of rank-occurrence distribution from a logarithmic prediction (Kolmogorov-Smirnov  $Z$ ; numbers of observations and significance of the deviation are given in parentheses)

	Estuarine oligochaetes	Beach macrofauna	Reef fish
$H'$	1.32	2.26	2.77
$\hat{D}/\hat{S}$	0.245	0.219	0.345
$Z$	0.816 ( $n = 24$ ; $p = 0.518$ )	0.935 ( $n = 56$ ; $p = 0.346$ )	1.231 ( $n = 66$ ; $p = 0.097$ )

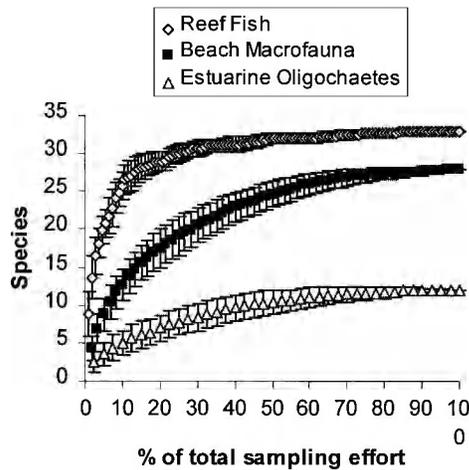


Fig. 3. Observed species accumulation curves ( $\pm 1$  SD) of the 3 datasets studied

precision, yet should not affect the  $S_{\infty}$  family of estimators greatly. The predominance of species occurring in low abundances in this dataset might also influence the behaviour of *Chao1*, which incorporates an abundance-based component.

The 'Beach macrofauna' dataset was intermediate in terms of species accumulation, with a steady asymptote. The species occurrence profile was not distinguishable from the logarithmic prediction (Kolmogorov-Smirnov  $Z = 0.935$ ,  $n = 56$ ,  $p = 0.346$ ), and diversity was intermediate with the lowest evenness of the 3 datasets in terms of abundance. Fig. 4 show the calculated values for bias of the 6 estimators for each dataset respectively; estimator precision is shown in

Fig. 5 and accuracy in Fig. 6 (note logarithmic y-axes in plots of accuracy and precision to improve resolution). Significant regressions of estimators bias, precision and accuracy across the 3 datasets against number of unique species and  $\log(\text{number of singleton species})$  are shown in Table 2.

Fig. 4 indicates that the same trends in bias are apparent across all 3 datasets.  $S_{\infty}$  consistently underestimates at all sampling efforts, whilst *Chao2* performs badly in the second quartile of the sampling effort, producing considerable overestimation. *ICE*, *Chao2* and *Jack* are consistent over-estimators after 30 to 40% of the total sampling effort is reached. *Chao1* has a slight tendency to overestimate in the 2 datasets with more even patterns of occurrence, and tends to underestimate in the 'Estuarine oligochaetes' dataset, with its high unpredictability in species occurrence. The typical magnitude of bias in the incidence- and non-incidence-based estimators is not different, but it often has opposite sign. *Chao1* and *Lag S<sub>∞</sub>* are the most consistently unbiased estimators in 2 of the datasets, but the performance of *Chao1* in the 'Estuarine oligochaetes' dataset is poor, whilst *Lag S<sub>∞</sub>* performs badly in the more equitable 'Reef fish' case, as does the straightforward  $S_{\infty}$ . The  $S_{\infty}$  family estimators are arguably more predictable in their response to increasing sampling effort than the other estimates, showing consistent improvements in bias with increasing sampling effort, but it requires up to 50% of the asymptotic sampling effort for *Lag S<sub>∞</sub>* to begin to mildly overestimate. In these crucial second to third quartiles of the curves, *Lag S<sub>∞</sub>* represents the least-biased and most-consistent estimator across datasets followed by *Chao1*. However *Chao1* is a more predictable overestimator than

Table 2. Bias, precision and accuracy figures from the 3 datasets pooled for each estimator and regressed against number of unique and singleton ( $\log$ -transformed) species in the assemblage at each sampling effort tested. In all regressions,  $n = 24$

	<i>ICE</i>	<i>Chao1</i>	<i>Chao2</i>	<i>Jack</i>	$S_{\infty}$	<i>Lag S<sub>∞</sub></i>
<b>Bias</b>						
Unique spp.	-	-	-	-	$R^2 = 0.252$ $p = 0.012$	-
Log(singleton spp.)	-	$R^2 = 0.23$ $p = 0.018$	-	-	-	-
<b>Precision</b>						
Unique spp.	-	$R^2 = 0.652$ $p < 0.000$	$R^2 = 0.493$ $p < 0.000$	$R^2 = 0.888$ $p < 0.000$	-	-
Log(singleton spp.)	-	$R^2 = 0.467$ $p < 0.000$	$R^2 = 0.240$ $p < 0.000$	$R^2 = 0.548$ $p < 0.000$	-	-
<b>Accuracy</b>						
Unique spp.	$R^2 = 0.187$ $p = 0.035$	$R^2 = 0.552$ $p < 0.000$	$R^2 = 0.509$ $p < 0.000$	$R^2 = 0.175$ $p = 0.042$	-	-
Log(singleton spp.)	-	$R^2 = 0.259$ $p = 0.011$	$R^2 = 0.236$ $p = 0.016$	-	-	-

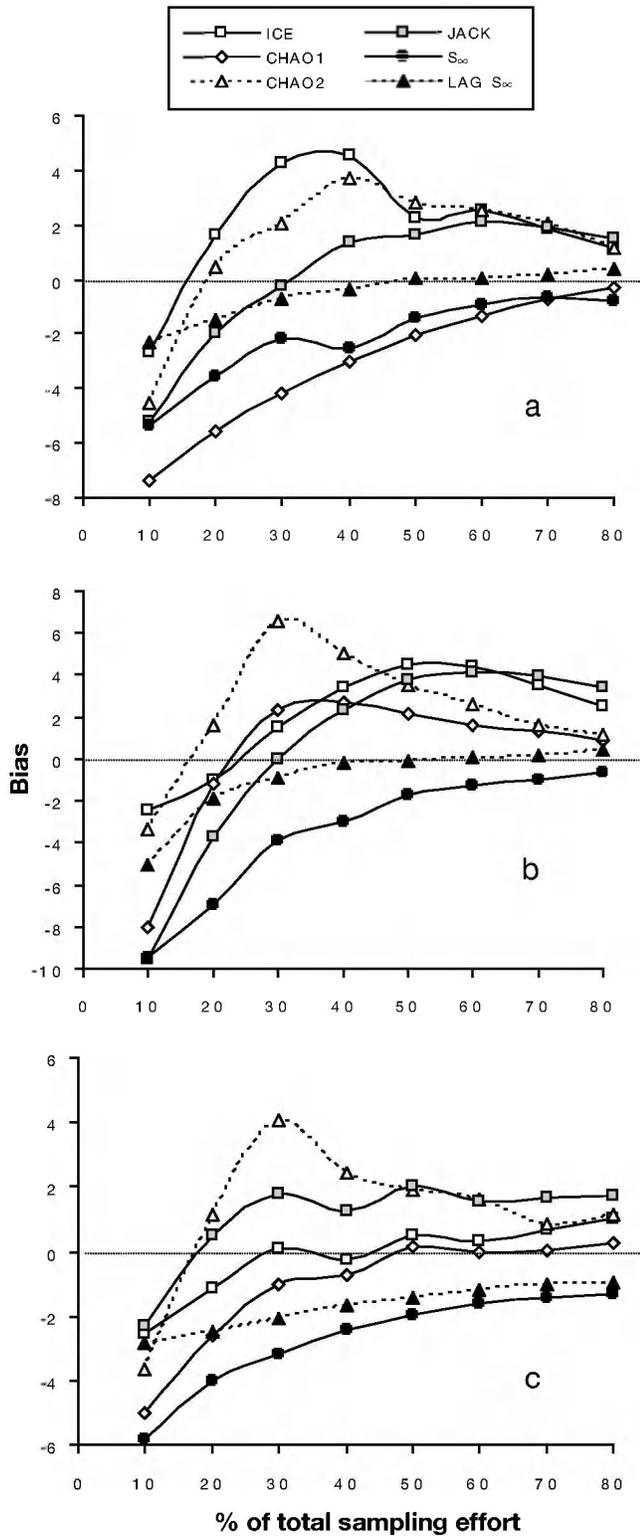


Fig. 4. Bias of the 6 estimators every 10th percentile of the total sampling effort up to 80%. (a) Estuarine oligochaetes, (b) Beach macrofauna; (c) Reef fish

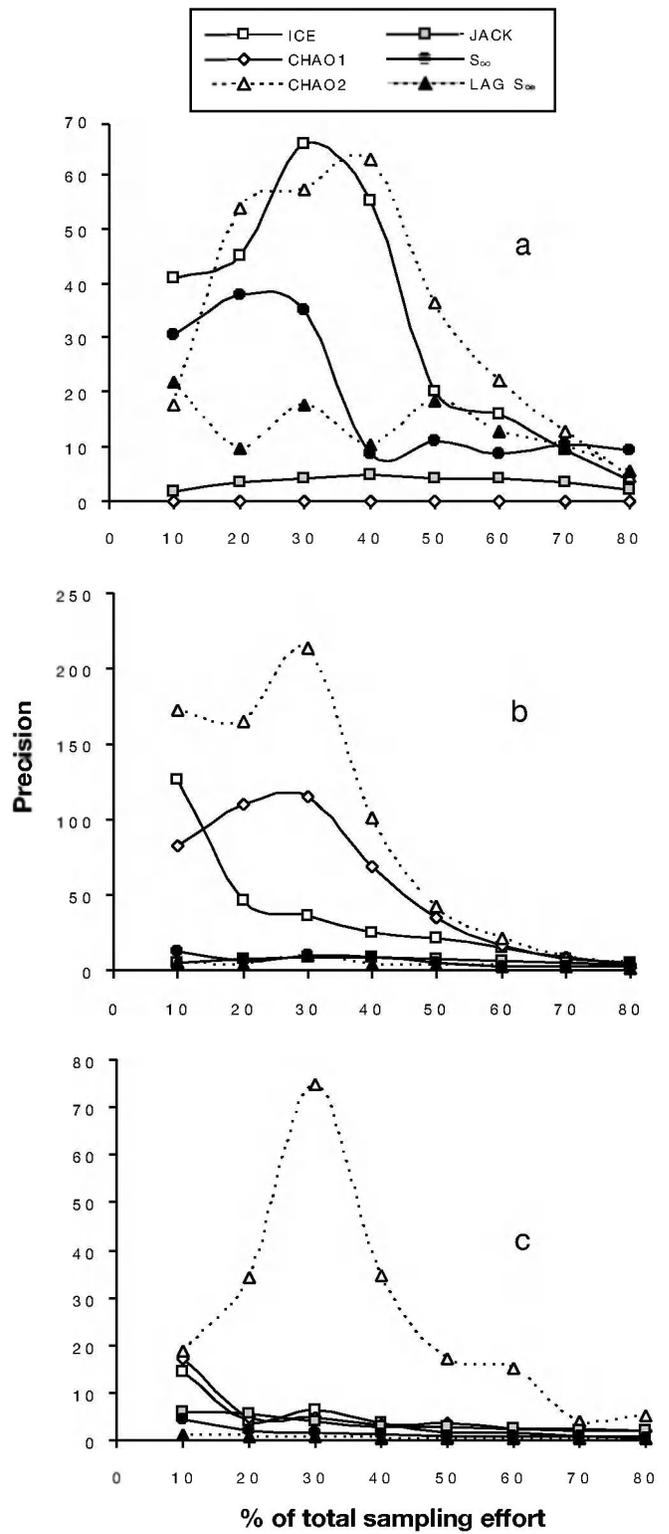


Fig. 5. Precision (Monte Carlo variance) of the 6 estimators every 10th percentile of the total sampling effort up to 80%. (a) Estuarine oligochaetes, (b) Beach macrofauna; (c) Reef fish. (NB: Low values indicate low variance, and hence high precision)

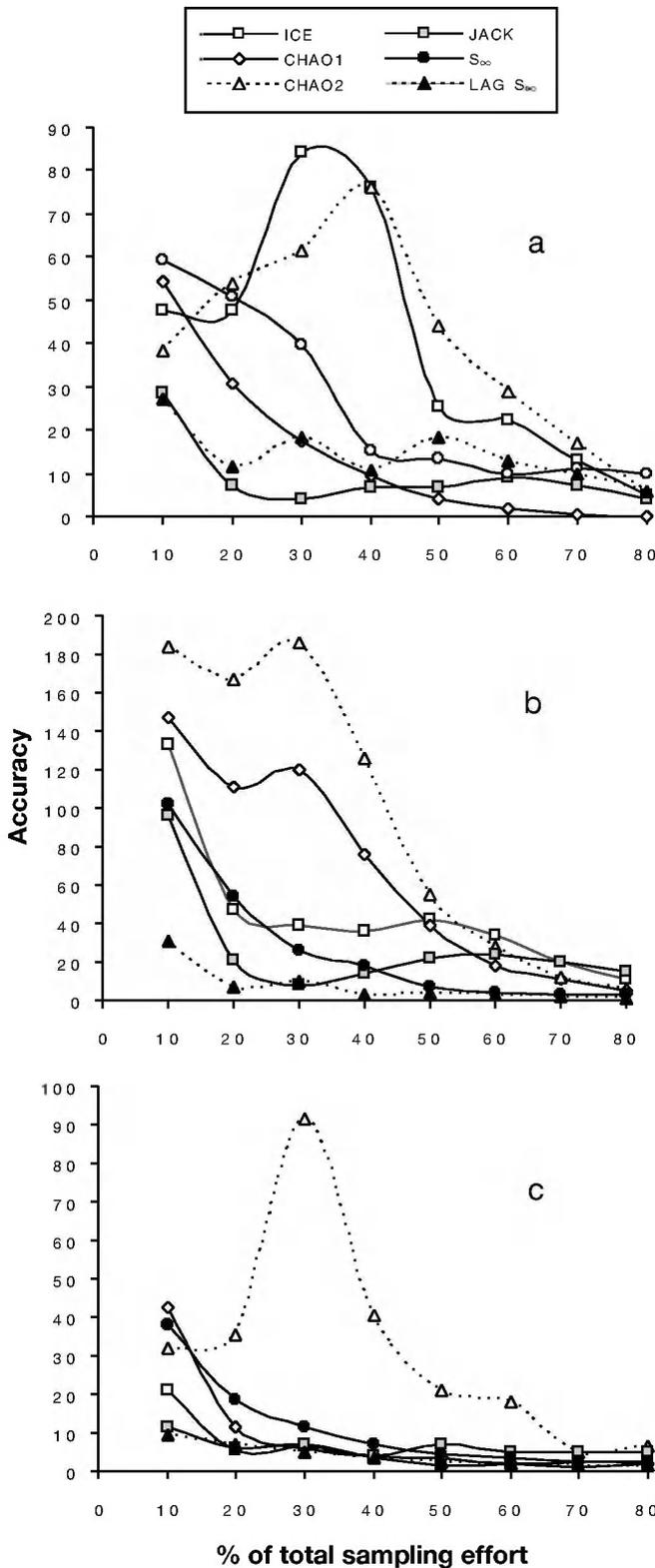


Fig. 6. Accuracy (mean square error) of the 6 estimators every 10th percentile of the total sampling effort up to 80%. (a) Estuarine oligochaetes, (b) Beach macrofauna; (c) Reef fish. (NB: Low values indicate high accuracy)

*Lag S<sub>∞</sub>*, which is significantly devalued by under-estimation.

The precision of the estimators (Fig. 5) reflects their performance in terms of bias. Equitability of the assemblage greatly affects precision levels, other than in the case of *Chao2*, which is consistently variable in all datasets. The *S<sub>∞</sub>* family estimators reflect their reduced variance thanks to the 2-stage process involved in the calculation; however, the case of the inequitable 'Estuarine oligochaetes' dataset demonstrates the potentially high variance produced by this estimator. *Chao1* produced incredibly low values for precision in the low-richness, low-abundance 'Estuarine oligochaetes' dataset. This probably reflects the index's response to the number of samples with abundance counts of 1 and 2 for the majority of the species; this dataset, however, produced the worst bias performance in this estimator. *Jack* was notable for the consistently low variability it produced.

With the exception of the *Chao1* in the final stages of the sampling effort curve for the 'Estuarine oligochaetes' dataset, *Lag S<sub>∞</sub>* was statistically the most accurate estimator, despite its bias being negative, but in the important second quartile of the effort range, *Jack* was the most consistently accurate in this phase of sampling with positive bias, despite it being surpassed by other estimators at higher effort levels. In contrast to their performance in the less even datasets, *ICE* and *Chao1* also performed well for the 'Reef fish' data.

Table 2 illustrates some of the underlying reasons for the general observations above.

Only 2 weak relationships are apparent which affect the bias of the estimators: the number of unique species increases bias in *S<sub>∞</sub>*, and the number of singleton species increases bias in *Chao1*. Increasing bias in *S<sub>∞</sub>* in response to numbers of unique species is probably an artefact.

Because incidence-based estimators rely on extrapolating from numbers of unique species in a set of samples, *Chao2* and *Jack* show decreasing precision with increasing numbers of unique species in the dataset. *Chao1* shows the same trend, probably caused by the fact that many unique species are also singletons.

In terms of outright accuracy, only *S<sub>∞</sub>* and *Lag S<sub>∞</sub>* are not affected by the numbers of singletons or unique species; all of the incidence- and abundance-based estimators are influenced to some degree by one or the other, with the marginal exception of *ICE*, with its unique/infrequent differentiation diluting the influence of unique species. The working value of 10 for the incidence cut-off chosen in this study (based upon the conclusions of Chazdon et al. 1997) may have introduced some bias into the comparison of this technique across the 3 datasets; the rank orders of accuracy of the estimators apparent in Fig. 6 show *ICE* signifi-

cantly higher in the hierarchy for the 'Reef fish' dataset, with its equitable incidence compared to the other 2 datasets.

## DISCUSSION

All of the estimators tested were capable of producing accurate and consistent estimations of species richness in the crucial middle quartiles of the sampling effort curve. They offered reliable and quite accurate predictions of richness based upon only half the number of samples required for the observed species richness to reach the asymptote. It is debatable whether this represents the kind of benefit that might warrant a reduction of the scale of sampling in species richness assessments. Perhaps a more pertinent question would be whether application of estimation techniques to inventories based upon limited sampling would improve the overall accuracy of such inventories, and perhaps also improve quantitative rankings of habitats inventoried.

The performance of the estimators applied to the 3 datasets described contrasts with that found by other authors (e.g. Walther & Morand 1998, Hellmann & Fowler 1999, Foggo et al. unpubl.). These results represent the first comparative, quantitative assessment of *Lag S<sub>∞</sub>*; the initial performance of this technique is encouraging, although a tendency to underestimation at lower relative sampling efforts is disappointing. This may be corrected by increasing the lag size, but this is likely to increase the variance of estimates.

Occasionally, some of the estimators cannot be calculated, this was not the case for our data. For example, in the low species richness parasite datasets studied by Walther & Morand (1998), *ICE* proved to be difficult to calculate, possibly due to the chosen cut-off value for the common/infrequent distinction. The implication of this is that a working cut-off value for the incidence distinction would bias against some habitats or assemblages and in favour of others, in a similar way to the arbitrary weighting of species richness and relative abundance in most diversity indices. Such subjectivity is undesirable if extrapolative estimation of species richness is to be used for monitoring purposes, and it might thus reduce the potential for the use of this estimator.

Walther & Morand's (1998) overall result placed *Chao2*, *Jack*, and *Chao1* in that rank order of performance across the second and third quartiles of the species accumulation curve for 5 datasets describing vertebrate parasites sampled using individual hosts as samples. The current study does not support this as a general ranking. Condit et al. (1996) found that *Jack* and *Chao1* performed poorly for their tropical tree

assemblages, which are also characterised by highly patchy distributions of individuals and high degrees of rarity. Walther & Morand's (1998) analysis of computer-simulated datasets in the same study supported their conclusions; it may thus be the case that the responses to assemblage incidence for their data all agreed more with those of our 'Reef fish' data than those of the other 2 datasets. The patchy nature of incidence in many macrobenthic assemblages might therefore represent a hurdle to the use of estimators such as *ICE* and *Chao2*.

In a recent study of freshwater benthic assemblages Foggo et al. (unpubl.) found *ICE*, *Chao1* and *Chao2* to be the most consistently accurate and precise estimators; they also highlighted the potential for the use of richness estimates as an alternative to indices of richness and diversity that employ *a priori* weightings of numbers of species and relative abundances. In their study, *Jack* performed poorly and *S<sub>∞</sub>* was discounted from the quantitative assessment at an early stage as it consistently under-estimated; the modified *Lag* form of *S<sub>∞</sub>* was not considered. Other studies (Melo & Froehlich 2001) have attempted to perform such a quantitative evaluation of estimator applicability for aquatic taxa, but have to date only presented data for incomplete species accumulations, and therefore cannot quantify the performance of the estimators against a reliable value for *S<sub>tot</sub>*. In comparison with our own unpublished data describing freshwater benthic taxa, the present study comparing *ICE*, *Chao1* and *Chao2* to *Lag S<sub>∞</sub>* demonstrates that the best performances previously available may be improved upon by newly developed approaches, especially in highly equitable datasets and at higher sampling efforts.

The data presented derive from habitats that share one important characteristic with most of the other habitats in which estimators have been shown to be effective: they are quite small and discrete, with relatively low absolute species richness. The performance of the estimators is likely to differ substantially in cases where different scales, both spatial and of species richness apply. These scales will influence the frequency of encounter of rare and infrequent taxa, with deleterious consequences for the performance of the estimators (Gotelli & Colwell 2001, J. S. Gray pers. comm.).

The properties of the assemblages sampled, their diversity, evenness and profile of species occurrence influence the performance of all of the species richness estimators to some degree, but the variability in both accuracy and precision across datasets of the abundance- and incidence-based estimators renders decisions about choice of estimators subjective and thus biased. Only the *S<sub>∞</sub>* estimators are unperturbed by these properties of assemblages, and hence these estimators may represent parsimonious choices for appli-

cation to a wide variety of different scenarios where species richness estimates are desired from limited sampling efforts, especially in inequitable assemblages such as the 'Beach macrofauna' and 'Estuarine oligochaetes' cases studied here. It is clear that, despite the disadvantages of subjective selection of an estimator for standardisation of richness extrapolations, there can be no unequivocal choice made on the basis of the data presented. However, given that slight overestimation is always better than underestimation, we believe that *Chao1* constitutes the best choice of estimator in the second and third quartiles of the sampling effort range, with the addition of *Lag S<sub>∞</sub>* as an option representing the estimators of choice at the highest sampling efforts if the possibility of underestimation is not considered an impediment.

The other main gain to be made from application of these techniques is elimination of sample number and sampling technique biases when comparing data from different sources. Because all of these techniques are based around standard areas and/or volumes of samples, however, the mitigation potential for sampling equipment differences may be limited.

Of the estimators tested, *Lag S<sub>∞</sub>* is certainly one of the most consistently accurate. It is also conservative, more frequently producing a small negative bias, which may be deemed a desirable quality in some circumstances. It does, however, employ an arbitrarily selected sample lag, which demands greater replication in sampling. Rumohr et al. (2001) have tested variable sample lags and found consistent gains in predicted richness. Our data indicate that increasing the sample lag also produces a desirable slightly positive bias. Increasing lag could also, however, incur penalties in terms of reduced applicability of the technique to sets of data differing in the numbers of samples taken; increasing lag reduces the end point at which extrapolations are possible, as in our 40-replicate dataset.

One different application of these techniques is in determining the efficiency of sampling. Asymptotic curves for estimators may not always be reliable estimates of true richness as indicated by the performance graphs; however, with datasets for which the true species richness has been reached by the sampled assemblage, the curves of observed and estimated richness should coalesce, irrespective of which estimator is employed. Separation of these 2 values indicates the potential for increasing the species count with further sampling.

Extrapolation appears to be capable of producing accurate, consistent estimates of maximum species richness for these 3 varied aquatic assemblages, and it may extend the potential practical applications of species accumulation studies explored by other authors

(Soberón & Llorente 1993, Hammond 1994, Gray 2000). Accurate estimates of richness, based on the presence or absence of data which are rapid and cost-effective to collect, may in turn be reliable correlates of abundance (Simberloff & Moore 1997), adding to the potential value of such techniques in biodiversity inventory. Gaston (1996) points out that elucidation of species richness may also reveal information about numbers of higher taxonomic units and morphological diversity, about degrees of phylogenetic disparity (see Williams & Humphries 1994), and about aspects of functional diversity. Extrapolation in order to generate estimates of total species richness from limited sampling could be used to test these postulates in hard-to-sample habitats such as the marine benthos, contributing to the pace of development of our knowledge of marine biodiversity in such areas.

*Acknowledgements.* The authors wish to thank the many people who have contributed in some way to the production of this paper. The staff and students of the Benthic Ecology Research Laboratory of the Department of Biological Sciences, University of Plymouth, helped in both the lab and field, and with comments on manuscript drafts. We are grateful for emails regarding the subject matter from Prof. J. S. Gray. Comments from 4 referees helped to substantially improve an earlier draft of the manuscript, The University of Plymouth assisted A.F. and M.J.A. financially. The Environment Agency funded the Rainham Creek survey. We gratefully acknowledge the assistance of Operation Wallacea for making data available to R. M. Burton during production of an undergraduate Honours thesis at the University of Plymouth. EstiMateS is made available by the University of Connecticut: <http://viceroy.eeb.uconn.edu/estimates>.

#### LITERATURE CITED

- Arnott SE, Magnuson JJ, Yan ND (1998) Crustacean zooplankton species richness: single- and multiple-year estimates. *Can J Fish Aquat Sci* 55:1573–1582
- Arrhenius O (1921) Species and area. *J Ecol* 9:95–99
- Arrhenius O (1923) Statistical investigation in the constitution of plant associations. *Ecology* 4:68–73
- Attrill MJ, Bilton DT, Chapman AS, Foggo A, Rowden AA, Rundle SD (2000) Rainham Creek/River Darent Ecological Survey Report. The Environment Agency, Bishops Square Business Park, Herts, St. Albans
- Cao Y, Williams DD, Williams NE (1998) How important are rare species in aquatic community ecology and bioassessment? *Limnol Oceanogr* 43:1403–1409
- Chao A (1984) Non-parametric estimation of the number of classes in a population. *Scand J Stat* 11:265–270
- Chao A (1987) Estimating the population-size for capture recapture data with unequal catchability. *Biometrics* 43:783–791
- Chao A, Lee SM (1992) Estimating the number of classes via sample coverage. *J Am Stat Assoc* 87:210–217
- Chao A, Ma MC, Yang MCK (1993) Stopping rules and estimation for recapture debugging with unequal failure rates. *Biometrika* 80:210–217
- Chazdon RL, Colwell RK, Denslow JS, Guariguata MR (1997).

- Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of NE Costa Rica. In: Dallmeier F, Corniskey J (eds) *Forest biodiversity in North, Central and South America and the Caribbean: research and monitoring*. Parthenon Publishing, Paris, p 285–309
- Clarke KR, Warwick RM (1994) *Change in marine communities: an approach to statistical analysis and interpretation*. NERC, Plymouth
- Clarke KR, Warwick RM (2001) A further biodiversity index applicable to species lists: variation in taxonomic distinctness. *Mar Ecol Prog Ser* 216:265–278
- Colwell RK (1997) *EstiMateS: statistical estimation of species richness and shared species from samples*. Version 5. University of Connecticut, Storrs, CT; available at <http://viceroy.eeb.uconn.edu/estimates>
- Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. *Phil Trans R Soc Lond B* 345:101–118
- Condit R, Hubbell SP, Lafrankie JV, Sukumar R, Manokaran N, Foster RB, Ashton PS (1996) Species-area and species-individual relationships for tropical trees: a comparison of three 50-ha plots. *J Ecol* 84:549–562
- Dumont HJ, Segers H (1996) Estimating lacustrine zooplankton species richness and complementarity. *Hydrobiologia* 341:125–132
- English S, Wilkinson C, Baker V (1994) *Survey manual for tropical marine resources*. Australian Institute of Marine Science, Townsville
- Gaston KJ (1996) Species richness: measure and measurement. In: Gaston KJ (ed) *Biodiversity: a biology of numbers and difference*. Blackwell Science, Oxford, p 77–113
- Gaston KJ, May RM (1992) Taxonomy of taxonomists. *Nature* 356:281–282
- Gaston KJ, Spicer JI (1998) *Biodiversity: an introduction*. Blackwell Science, Oxford
- Gentil F, Dauvin JC (1988) Is it possible to estimate the total species number of a macrobenthic community? Application to several soft-bottom communities from the English Channel. *Vie Milieu* 38:207–212
- Gleason HA (1922) On the relation between species and area. *Ecology* 3:158–162
- Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol Lett* 4:379–391
- Grassle JF, Lasserre P, McIntyre AD, Ray GC (eds) (1991) *Marine biodiversity and ecosystem function. A proposal for an international programme of research*. Biology International, Special Issue No. 23
- Gray JS (2000) The measurement of marine species diversity, with an application to the benthic fauna of the Norwegian continental shelf. *J Exp Mar Biol Ecol* 250:23–49
- Hammond PM (1994) Practical approaches to the estimation of the extent of biodiversity in speciose groups. *Phil Trans R Soc Lond B* 345:119–136
- Hellmann JJ, Fowler GW (1999) Bias, precision, and accuracy of four measures of species richness. *Ecol Appl* 9:824–834
- Heltshel JF, Forester NE (1983) Estimating species richness using the jack-knife procedure. *Biometrics* 39:1–11
- Jaramillo E, McLachlan A, Dugan J (1995) Total sample area and estimates of species richness in exposed sandy beaches. *Mar Ecol Prog Ser* 119:311–314
- Karakassis I (1995) S<sub>h</sub>: a new method for calculating macrobenthic species richness. *Mar Ecol Prog Ser* 120:299–303
- Kilburn PD (1966) Analysis of the species-area relation. *Ecology* 47:831–843
- Krebs CJ (1999) *Ecological methodology*, 2nd edn. Benjamin Cummings, Menlo Park, CA
- Lande R, DeVries PJ, Walla TR (2000) When species accumulation curves intersect: implications for ranking diversity using small samples. *Oikos* 89:601–605
- Larsen PF, Dogget LE (1990) Sand beach macrofauna of the Gulf of Maine with inference on the role of oceanic fronts in determining community composition. *J Coast Res* 6: 913–926
- Lee SM, Chao A (1994) Estimating population size via sample coverage for closed capture-recapture models. *Biometrics* 50:88–97
- Magurran AE (1988) *Ecological diversity and its measurement*. Princeton University Press, Princeton, NJ
- Melo AS, Froehlich CG (2001) Evaluation of methods for estimating macroinvertebrate species richness using individual stones in tropical streams. *Freshw Biol* 46:711–721
- Palmer MW (1990) The estimation of species richness by extrapolation. *Ecology* 71:1195–1198
- Riddle MJ (1989) Precision of the mean and the design of benthos sampling programs—caution advised. *Mar Biol* 103: 225–230
- Rosenzweig ML (1995) *Species diversity in space and time*. Cambridge University Press, Cambridge
- Rumohr H, Karakassis I, Jensen JN (2001) Estimating species richness, abundance and diversity with 70 macrobenthic replicates in the Western Baltic Sea. *Mar Ecol Prog Ser* 214:103–110
- Simberloff D, Moore J (1997) Community ecology of parasites and free-living animals. In: Clayton DH, Moore J (eds) *Host-parasite evolution: general principles and avian models*. Oxford University Press, Oxford, p 174–197
- Soberón J, Llorente J (1993) The use of species accumulation functions for the prediction of species richness. *Conserv Biol* 7:480–488
- Tukey J (1958) Bias and confidence in not quite so large samples. *Ann Math Stat* 29:614
- Walther BA, Morand S (1998) Comparative performance of species richness estimation methods. *Parasitology* 116: 395–405
- Walther BA, Cotgreave P, Gregory RD, Price RD, Clayton DH (1995) Sampling effort and parasite species richness. *Parasitol Today* 11:306–310
- Williams PH, Humphries CJ (1994) Biodiversity, taxonomic relatedness, and endemism in conservation. In: Forey PL, Humphries CJ, Vane-Wright RI (eds) *Systematics and conservation evaluation*. Oxford University Press, Oxford, p 269–287

*Editorial responsibility: Otto Kinne (Editor), Oldendorf/Luhe, Germany*

*Submitted: November 2, 2001; Accepted: October 25, 2002  
Proofs received from author(s): February 17, 2003*