NOTE

S_∞: a new method for calculating macrobenthic species richness

Ioannis Karakassis

Institute of Marine Biology of Crete, PO Box 2214, GR-710 03 Heraklion, Crete, Greece

ABSTRACT: A simple mathematical method is proposed to calculate the total number of macrobenthic species present within a community when quadrat sampling is involved. The total number of species (S_∞) is calculated by extrapolation as the point beyond which no further increase in the cumulative number of species is expected by additional sampling effort. To this end the cumulative number of species in k samples is plotted against the cumulative number of species in $k + 1$ samples. The equation of the resulting line is solved for $y = x$, thus obtaining the intercept point with the bisector of the angle between the x and y axes. Different benthic data sets were used for testing the method, all of which gave satisfactory results. The method may be used in biodiversity and biogeography studies to compare species richness between different areas and in ecological studies to define the minimum sampling area.

KEY WORDS: Species richness · Diversity · Species-area curves · Quadrat sampling

The increasing interest in biodiversity imposes the need for a further refinement of the mathematical tools ecologists use to quantify the complexity of the living world. Within a given area, a certain number (S) of species is present, but its direct determination is not always feasible especially in cases where sampling (as opposed to census) is involved. This number is an indispensable ecological parameter and is used in ecology in association with various diversity indices. For instance, several evenness indices (Sheldon 1969, Heip 1974, Pielou 1975) are based on the total number of species in the community. In practice they are calculated using the total number of species in the sample, thereby introducing a numerical bias which, in some cases, might render an index useless (Peet 1975).

The estimation of species richness based on the jackknife estimator (Heltshe & Forrester 1983a) could be an alternative to these problems but this method overestimates the number of species in a community (Heltshe & Forrester 1983b).

This paper presents a possible solution to this problem by proposing a method for the estimation of the total number of species (S_∞) within a benthic community by extrapolation from a number of equal size quadrats taken at random.

Method. *Concept:* The species-area curves resulting from the analysis of a number of replicates from a given area are increasing curves approaching a maximum limit asymptotically. In most of the cases encountered in the literature the following procedure was followed: an arbitrary number of samples was taken and analyzed and a species-area curve was plotted. The minimum number of samples required was calculated as a percentage of the total number of species in the data set as if the community were totally sampled. In most cases however the curve was far from reaching its limit and therefore the total number of species was severely underestimated.

The method proposed in this paper addresses this problem by permitting the calculation of a number that is theoretically the upper limit of the asymptote. This is done by plotting the cumulative number of species present in k samples against the cumulative number of species present in $k + 1$ samples. By solving the equation of the regression line for $x = y$, the intercept with the bisector of the angle between the x and y axes is obtained i.e. the point where 2 successive samples are expected to present the same cumulative number of species.

Computations of the method involve randomization to obtain an estimate of the cumulative number of species for a certain number of samples, as previously done by Holme (1953), and a smoothing of the curve, as previously done by Weinberg (1978).

The merits of this method are: (1) it is not sensitive to the distribution of individuals among the species, (2) it can be used in cases of clumped distribution, (3) in theory it may be used to calculate the expected number of species from any number of quadrats; however, very small numbers of quadrats would tend to underestimate, (4) it does not require the assumption of an underlying statistical distribution of individuals in the community.

The minimum number of quadrats required in order to obtain a reliable estimate of S_∞ depends on the mean and the variance of the number of species per quadrat as well as on the homogeneity of the sampled community.

This method is in general similar to the one used to calculate L_∞ in animal growth models (Walford 1946, Jones 1981), where L_∞ is the asymptotic length that is attained if the animal is assumed to grow to an infinite age.

The S_∞ method has no meaning at the one-sample level. Therefore comparison with a species richness index such as that introduced by Margalef (1958) is not possible.

The 'pooled quadrat method' introduced by Pielou (1966a, 1969, 1975) is a method operating on a set of quadrats (as S_∞ and jackknife methods do) but its output (HB_{pop}) is a diversity index and not a number of species (as in the S_∞ and jackknife methods), therefore the comparison between HB_{pop} and S_∞ is not possible either.

What is examined in this method is the rate at which different species are incorporated into the data set with increasing sampling effort. There is no reason to assume that this method is specific to macrobenthic communities. It is important though that each quadrat contains an adequate number of species as in the case of macro- or meiobenthic samples obtained by means of standard samplers.

Algorithm: The method proposed in this paper is performed on presence-absence data and consists of the following procedural steps (Fig. 1).

(1) A number of different random permutations (P) of the sampling units are performed using a random numbers generator. ($N!$ is the maximum number of different arrangements which can result from N sampling units).

(2) The cumulative number of species (S_{ik}) present in k quadrats ($k = 1 \dots N-1$) in the ith permutation (out of P) of the sampling units is computed.

(3) The mean species content of all possible sets of k quadrats is computed as:

$$\overline{S}_k = \frac{1}{P} \sum_{i=1}^{P} S_{ik}$$

(4) The regression line for the pairs ($\overline{S}_k, \overline{S}_{k+1}$) is plotted using the equation:

$$\overline{S}_{k+1} = a + b\overline{S}_k$$

(5) The coordinates of the intercept of the regression line with the dichotomous line of the angle of the co-ordinates axes are calculated by solving Eq. (2) for $\overline{S}_k = \overline{S}_{k+1}$:

$$S_\infty = \frac{a}{1-b}$$

The resulting value S_∞ is the theoretical limit of the asymptote in the species-area curve.

There is no need to calculate S_{iN} since this is constant for all permutations and equals the total number of species in the sample set.

An alternative way of obtaining the same result was also examined: the calculation of S_∞ for each permutation. However although the linear correlation coefficient was also rather high (>0.9) in all permutations, and although this alternative could also provide confidence limits, it was rejected because when it was used with a small number of quadrats (less than 9), a small
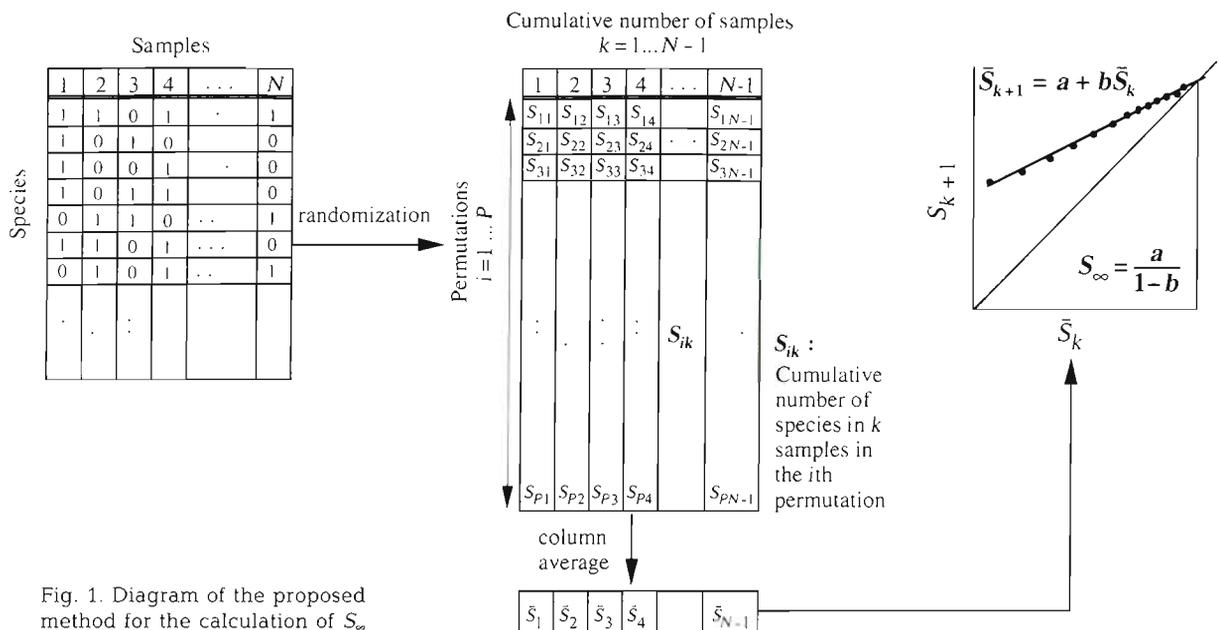

Fig. 1. Diagram of the proposed method for the calculation of S_∞

proportion of the permutations (<0.001%) produced negative values of S_∞.

Confidence intervals could be calculated using this alternative through pseudo-replication (averaging S_k for different sub-permutations) although in this case the mutual independence of the samples is also not met.

Results. The method has been previously tested with different data sets (Karakassis 1991, Tselepides 1992, A. Elefthesiou, Institute of Marine Biology of Crete, unpubl.), and all produced an excellent linear correlation between S_k and S_{k+1}. Two data sets are presented here. (1) Data from the English Channel off Plymouth, UK (Holme 1953) consisting of 40 benthic samples taken with a scoop sampler of $\frac{1}{20}$ m^2 at 60 m depth from a clean fine sand bottom and sieved through a 1.2 mm mesh. The 3 dominant species were *Lumbriconereis* sp., *Ampelisca tenuicornis* and *Magelona papillicornis*. (2) Data from the continental shelf of Crete, Greece (Karakassis 1991) consisting of 19 samples taken with a Smith-McIntyre grab of $\frac{1}{10}$ m^2 at the

70 m depth contour from a muddy bottom community and sieved through a 0.5 mm mesh. The 3 dominant species were *Onchnesoma steenstrupi*, *Tharyx heterochaeta* and *Golfingia procera*.

In each data set 150 different permutations were used to obtain the mean species content in each set of quadrats.

The results in both cases revealed a high linear correlation coefficient (0.999 and 0.998 respectively) between S_k and S_{k+1} (Fig. 2). The predicted S_∞ in both cases was higher than the actual numbers of species found in the respective surveys and lower than the total number of species predicted by jackknife estimation.

The number of permutations used did not influence the mean S_k but it did influence the variance. The 95% confidence intervals (Fig. 2c, d) are narrow at the edges of the plots and wider in the middle because the possible combinations of quadrats are less than 150 for sample sets of 1 and 39 in the first data set and 1 and 18 in the second. In the middle however there are
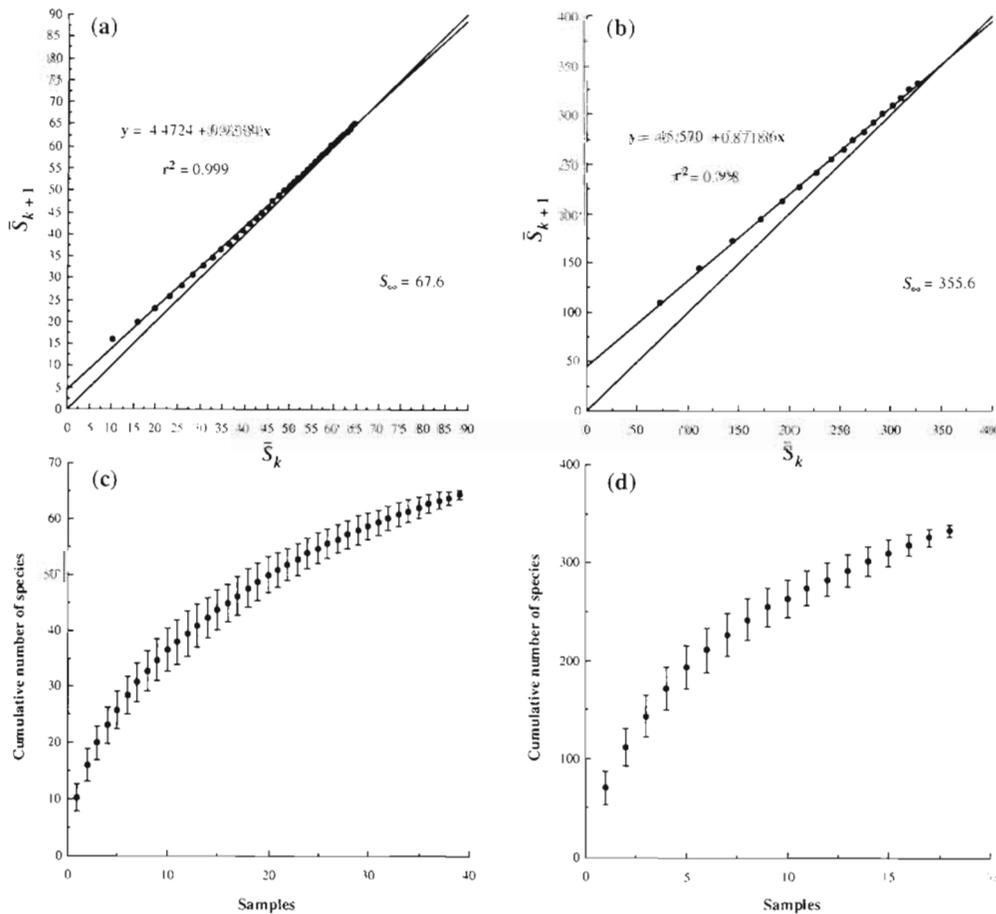

Fig. 2. Plots of (a, b) $\bar{S}_k$ against $\bar{S}_{k+1}$ and (c, d) cumulative number of species against number of samples for data sets from the English Channel (Holme 1953) (a, c) and the continental shelf of Crete (b, d)

1.3×10^{11} possible combinations of 20 samples out of 40 (Holme) and 92 378 possible combinations of 10 samples out of 19 (Karakassis). Therefore some values are inevitably repeated in the case of 1 or $N-1$ data sets, reducing the variance.

The numbers of species predicted by jackknife estimation, the actual numbers of species in the data sets and the predicted S_∞ are presented in Table 1. In Holme's study S_∞ is only slightly higher than the number of species in the data because some species (e.g. unidentified polychaetes) were excluded from the analysis. In comparison to the data set from Crete the English Channel community seems very poor in species, but the mesh size used is larger. Therefore S_∞ in the English Channel seems to be underestimated. It could be argued however that this S_∞ reflects the larger mesh size and the taxonomic level of identification. There is considerable distance between S_∞ and the jackknife estimation for the 70 m depth muddy community in Crete (Table 1).

Discussion. The proposed method has some elements in common with numerical techniques previously used for the same or different purposes: the randomization technique of the samples sequences (Holme 1953, Pielou 1966b, 1975) and the determination of the mean species content of all possible subsets of quadrats (Weinberg 1978). The proposed method could also be used in combination with the 'semicumulative curves' for mutually independent sets of quadrats proposed by Ursin (1960), but this would reduce the number of points in the regression line by half since after $N/2$ there would be only 1 independent set of samples.

Gentil & Dauvin (1988) suggested that rare species be excluded from the analysis of the cumulative species area curve. Apart from the arbitrariness of the term 'rare', their attempt to predict, by extrapolation, the total number of species present in a community is not very different from the ones already in use.

The disadvantages of the method proposed here are related to the problems of the community or assemblage concepts. Heck et al. (1975) noted that in a very patchy environment the species accumulation curve could become asymptotic before many of the species were sampled. Therefore it seems that all analyses dealing with these sorts of curves require a certain amount of 'community uniformity'. The latter however can be regarded merely as a convenient simplification. The universal communities of Petersen (1913), or Thorson (1955), were used with less enthusiasm after Mills (1969) defined the assemblage and the continuum concept. Lambshead et al. (1983) finally define an assemblage as the result of adequate sampling of all organisms of a specific category in a defined place.

If quadrat size is too small and the number of samples too low an underestimation of the total number of species might result this would also result in a species-area curve. This method is not directly influenced by the aggregation pattern of the species in the community, provided that the number of samples is not too small.

This proposed method does not predict the number of samples required to catch all the species in the community. Theoretically, this is done only by infinite sampling effort. However it does provide a more stable number on which the proportion caught by sampling effort made can be calculated.

The method can be used to compare species richness in different environments or biogeographical areas when similar quantitative samplers have been used (e.g. grabs, cores, etc.) but different amounts of samples have been taken. The method proposed in this study could also be used for samples taken by means of trawls, dredges, plankton nets etc., provided that the sampled volume is equal between samples.

Since no adequate estimation of species richness exists, comparison between communities is based on various methods that are mainly 'evenness-influenced'. Rarefaction estimates for instance (Hurlbert 1971, Heck et al. 1975) are mainly based on the distribution of individuals among species and therefore the presence of clumped species with high abundance in 1 sample would severely modify the expected number of species. In the proposed method a different approach to the between-communities comparison is adopted: the accumulation of species in relation to the sampling effort and the rate at which new species are incorporated into the data set are taken into account.

Table 1. Comparison between the 2 data sets used with respect to total area sampled, species numbers, and species-numbers estimations

	English Channel	Crete
Total area sampled (m^2)	2.0	1.9
Species number in data set	64	340
S_∞	67.6	355.6
S (jackknife)	84.5	464.1

LITERATURE CITED

Gentil, F., Dauvin, J.-C. (1988). Peut-on estimer le nombre total d'espèces d'un peuplement macrobenthic? Application aux peuplements de substrat meuble de la Manche. Vie Milieu 38(3/4): 207–212

Heck, K. L. Jr, van Belle, G., Simberloft, D. (1975). Explicit

calculation of the rarefaction diversity measurement and the determination of sufficient sample size. Ecology 56: 1459–1461

Heip, C. (1974). A new index measuring evenness. J. mar. biol. Ass. U.K. 54: 555–557

Heltshe, J. F., Forrester, N. E. (1983a). Estimating species richness using the jackknife procedure. Biometrics 39: 1–11

Heltshe, J. F., Forrester, N. E. (1983b). Estimating diversity using quadrat sampling. Biometrics 39: 1073–1076

Holme, N. A. (1953). The biomass of the bottom fauna in the English Channel off Plymouth. J. mar. biol. Ass. U.K. 32: 1–49

Hurlbert, S. H. (1971). The non-concept of species diversity: a critique and alternative parameters. Ecology 52: 577–586

Jones, R. (1981). The use of length composition data in fish stock assessments (with notes on VPA and cohort analysis). FAO Fish. Circ. 734, Rome

Karakassis, I. (1991). Contribution to the study of the benthic ecosystem of the continental shelf of Crete. Ph.D. thesis, University of Crete

Lambshead, P. J. D., Platt, H. M., Shaw, K. M. (1983). Detection of differences among assemblages of benthic species based on an assessment of dominance and diversity. J. nat. Hist. 17: 859–874

Margalef, R. (1958). Information theory in ecology. Gen. Syst. 3: 36–71

Mills, E. L. (1969). The community concept in marine zoology with comments on continua and instability in some marine communities. A review. J Fish. Res. Bd Can. 26: 1415–1428

Peet, R. K. (1975). The measurement of species diversity. A. Rev. Ecol. Syst. 5: 285–307

Petersen, C. G. J. (1913). Valuation of the sea. II. Animal communities of the sea bottom and their importance for marine zoogeography. Rep. Dan. Biol. Stn 21. 1–14

Pielou, E. C. (1966a). Species diversity and pattern diversity in the study of ecological succession. J. theor. Biol. 10: 370–383

Pielou, E. C. (1966b). The measurement of diversity in different types of biological collections. J. theor. Biol. 13: 131–144

Pielou, E. C. (1969). An introduction to mathematical ecology. Wiley, New York

Pielou, E. C. (1975). Ecological diversity. Wiley, New York

Sheldon, A. L. (1969). Equitability indices: dependence on the species count. Ecology 50: 466–467

Thorson, G. (1955). Modern aspects of marine level bottom communities. J. mar. Res. 14(4): 387–397

Tselepides, A. (1992). Ecological study of the bathyal ecosystem of the Aegean Sea. Ph.D. thesis, University of Crete

Ursin, E. (1960). A quantitative investigation of the echinoderm fauna of the central North Sea. Medd. Denmarks Fisk og Havunders. N.S. 2(24)

Walford, L. A. (1946). A new graphic method of describing the growth of animals. Biol. Bull. mar. biol. Lab., Woods Hole 90: 141–147

Weinberg, S. (1978). The minimal area problem in invertebrate communities of Mediterranean rocky substrata. Mar. Biol. 49: 33–40

This note was submitted to the editor