Estimating species richness, abundance and diversity with 70 macrobenthic replicates in the Western Baltic Sea

Heye Rumohr[1,*], Ioannis Karakassis[2], Jørgen Nørrevang Jensen[3]

[1]Institut für Meereskunde, Düsternbrooker Weg 20, 24105 Kiel, Germany
[2]Institute of Marine Biology of Crete, PO Box 2214, 71003 Heraklion, Crete, Greece
[3]International Council for the Exploration of the Sea, Palægade 2-4, Copenhagen 1261, Denmark

ABSTRACT: An unusually large number of replicated macrofaunal samples (70) was taken from the Western Baltic in May 1995 for a ringtest in an ICES/HELCOM intercalibration exercise. This data set was employed in this study in order to investigate the performance of numerical methods used for predicting species richness and to assess the accuracy of the estimates of abundance and diversity currently used in benthic ecology. The results of this study indicate that: (1) more than 10 replicates are required in order to include in the data set more than two-thirds of the species found in 70 replicates, and more than 53 replicates are required in order to include 95 % of the species; (2) estimates of average abundance and of average Shannon-Wiener diversity index using 5 replicates could result in less than 40 % error; this could be less than 30 % for 10 replicates and less than 5 % for 70 replicates; (3) both types of species-richness predictions (jackknife estimate and S_∞) increased with increasing number of samples used in the calculations, indicating that their ability to assess overall species richness in the community is rather limited; in particular, it is shown that jackknife overestimates and S_∞ slightly underestimates species richness. Different configurations of the S_∞ method were tested in order to optimize its performance, and it was found that both truncation and increasing sampling lag result in increased and stabilized estimates of species richness.

KEY WORDS: Macrobenthos sampling methodology · Species-area curves · Abundance · Diversity · Western Baltic

INTRODUCTION

The species-area curve is a pattern that has been known for a long time in ecology (Arrhenius 1921, Rosenzweig 1995), and in fact seems to consist of 3 or 4 different patterns (Williams 1943, Rosenzweig 1995), i.e. it can be calculated (1) among tiny pieces of a single biota, (2) among larger pieces of a single biota, (3) among islands of 1 archipelago and (4) among areas with independent evolutionary histories. These patterns are shaped by processes operating at different spatial and temporal scales, ranging from stochastic within-site variability to speciation and extinction. At the low end of this spectrum, species-area curves are often used as a criterion for decisions on sampling strategy, particularly with respect to the minimum sampling effort required. However, the number of species in a community to which intermediate yield-per-sample is compared may also be considered as an intermediate result in an infinite series of samples that could be taken from a given site at a given time. Therefore, the number of species included in the calibration sample represents only a subset of the species richness in the community (in this paper the term 'species richness' is used to refer to the total number of species in the community and not to the index introduced by Margalef 1958).

*E-mail: hrumohr@ifm.uni-kiel.de

Two methods have been proposed as a means of overcoming this problem by extrapolating the existing information: (1) the jackknife estimate (Heltshe & Forrester 1983a) predicting total species richness from a series of samples using the number of species present in n samples, the number of samples and the number of *unique* species (i.e. of species present in 1 and only 1 sample), and (2) the S_∞ method (Karakassis 1995), whereby total species richness is calculated by regression of the cumulative number of species in $k + 1$ samples against that in k samples and by calculating the intercept point with the bisector of the angle between the coordinates, thus producing an estimate of the number of species beyond which no further increase in the cumulative number of species is expected using additional sampling effort. None of these methods has been calibrated against a really high number of replicated samples.

The data set of macrobenthic replicates presented here is one of the largest multi-species ecological replication data sets in scientific literature — it contains 70 true replicates (taken in sequence from the same site) or a total of 7 m² of sea bed. This data set has been produced using a quality assurance protocol and originates from an area where the fauna is largely known. Patchiness in this particular station, or other sources of spatial heterogeneity (Morrisey et al. 1992), may be considered low; the species richness in the entire area is relatively low, so the cumulative species richness curve should be closer to the asymptote than in most other known similar data sets. All these qualities/properties make the data set appropriate for testing the performance of statistics developed to predict community attributes from a limited number of samples.

MATERIALS AND METHODS

Sampling site. The macrofaunal data presented in this study came from the ICES/HELCOM intercalibration study, wherein these samples had been used for a ringtest between 13 participating laboratories (ICES 1996). The Station 'Millionenviertel 14' is located in the northern part of Kiel Bay (Western Baltic) at a depth of 24 m, and has a sandy/mud sediment. Maps indicating the position of the sampling station and the main biological features can be found in Arntz (1981) and Weigelt (1991).

Analysis of macrofauna. Macrofaunal samples were collected by means of a 0.1 m² van Veen grab. They were sieved on board through a 1 mm sieve, preserved in 4% formaldehyde, and identified to species by the environmental consultant ZOOTAX in Stockholm (Sweden). The methods generally followed the ICES

and HELCOM recommendations on sampling benthos and treatment of samples (Rumohr 1999).

Multivariate analysis. Abundance data of all macrobenthic replicates taken were analyzed by means of nonmetric multidimensional scaling (N-MDS) as recommended by Field et al. (1982). Similarities were calculated using the Bray-Curtis similarity index (Bray & Curtis 1957). Prior to the analysis, a double square-root transformation was applied to the abundance in order to normalize and avoid skew in the data set. This analysis was performed in order to detect whether the samples taken form discrete subgroups, thus introducing bias in subsequent analyses.

Species richness abundance and diversity estimates. The species-area curve was plotted using 100 computer-generated random sample sequences (Holme 1953, Karakassis 1995) and the mean species content of all possible subsets of samples by averaging as described in Weinberg (1978) and Karakassis (1995). This technique minimizes the potential for bias in the shape of the cumulative curve of species number against the number of samples which could result from the presence of extremely species-rich or species-poor samples at the beginning of the sequence.

The same type of procedure was used for the calculation of the average abundance and Shannon-Wiener index, H' (Shannon & Weaver 1949), performing 100 random permutations in the sample sequence. For each of these permutations, the average value was calculated for 1, 2, 3, ... up to 70 samples:

$$Y_{ij} = \sum_{k=1}^{i} \frac{x_{kj}}{i}$$

where x_{kj} = the value (abundance or diversity) of the ith sample in sequence during the jth random permutation, and y_{ij} = the average value for the first i samples of the jth random permutation.

In order to estimate the maximal possible error in either of these variables (abundance and diversity), we have defined as maximal deviation (mDi) the maximal absolute difference between the cumulative average abundance (or diversity) for each possible subset of species from the average value for 70 samples (y_{70j}) divided by the average value for 70 samples, which may be considered as the value closest to the 'true' mean of the community:

$$mD_i = \max_{j=1}^{100} \left| \frac{Y_{ij} - Y_{70j}}{Y_{70j}} \right| \times 100$$

The advantage of this approach is that it provides a measure of the potential error, with no assumption on the distribution of abundance among the replicates.

The calculation of the jackknife estimate was performed also with 100 computer-generated random-sample sequences, and during each permutation the

prediction of the estimate was recorded as the sample size increased from 1 to 70 grabs.

The S_∞ prediction was calculated in the same way (with increasing sample size) using the average of 100 permutations. In addition, 2 different configurations were tested with this method modifying the initial method proposed by Karakassis (1995):

Varying sample lag (SL): The initially proposed method was based on a regression of the species in $k + 1$ samples against the species contained in k samples. The main concept was that this index would provide the number of species expected to be attained when the difference in the cumulative number of species between 2 consecutive samples (sampling lag = 1) would be zero. However, in that case there should also be a zero difference between higher sampling lags. It could be expected that the increase in sampling lag would provide more accurate results, particularly in species-poor communities, since it could give a higher resolution in detecting trends in the increase of species richness. Therefore, we have chosen to test the behaviour of the S_∞ index for different values of SL (1, 10 and 30).

Truncation of the ends of the series: As has been stated before (Karakassis 1995), the standard deviation of species richness at the 2 ends of the series is lower than in the middle, since the number of combinations at those points is limited and, therefore, some of the values are repeated during the permutations. In a broader context, it could be argued that the 2 ends of the data series are more dependent on the values resulting from the particular sampling event, whereas the intermediate values would be more representative of the rate of increase in richness. In this particular study, we have tested the behavior of the S_∞ model when omitting 1 to 10 pairs of values.

RESULTS

Homogeneity of the site

The results from the multivariate analysis through MDS (Fig. 1) did not show any major clustering of the replicates into discrete groups. The relatively high stress value (although below 0.3) is typical for samples which do not represent series along a gradient or grouping into clusters. The average similarity between different samples was 67%, with a minimum of 38 and a maximum of 87%. In comparison to the first sample taken, similarity of subsequent samples ranged between 46 and 78%.

Homogeneity was further tested by comparing the subdiagonals of the association matrix. Since samples were arranged by sampling hour, the similarities of
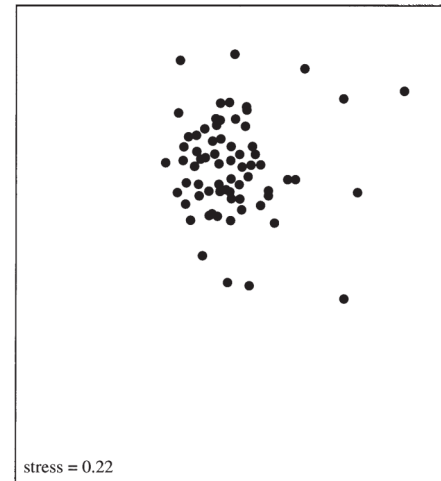

stress = 0.22

Fig. 1. MDS ordination plot of all 70 replicated samples taken

the first subdiagonal correspond to samples taken in sequence (lag = 1), those of the second subdiagonal to similarities among every second sample (lag = 2) and so on. Averaging similarities of these subdiagonals (Fig. 2) provides an additional means of testing for multivariate homogeneity: if the average similarity decreases with increasing lag, then it could be assumed that strong patchiness or a 'hidden gradient' reduces the homogeneity of the data. The data from the present study did not show such an effect, the average similarity remaining relatively constant for Lags 1 to 40, ranging from 63 to 68%.

Species occurrence

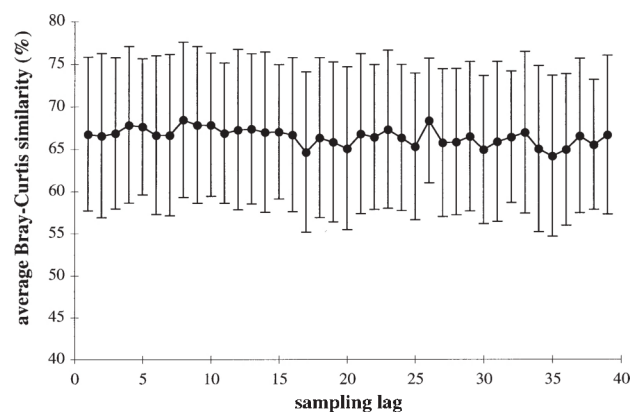Regarding species distribution among replicates (Fig. 3), it was found that a large proportion of the spe-

Fig. 2. Average (±SD) similarity between samples taken at different sampling lags (lag = 1 for samples taken in sequence)
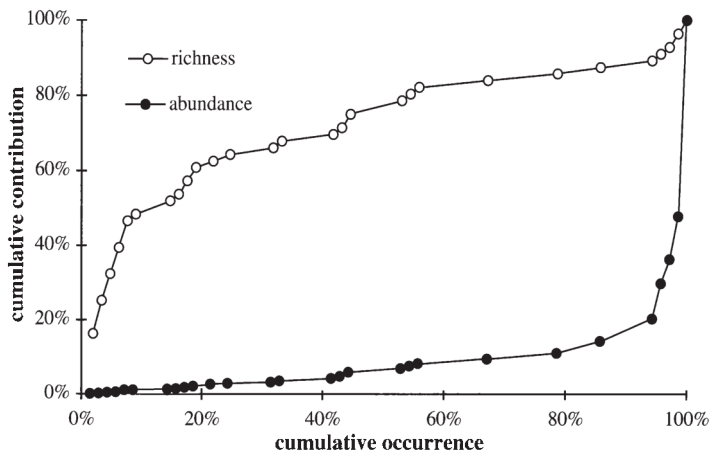
Fig. 3. Cumulative contribution of macrobenthic species to total abundance and to species richness in relation to cumulative occurrence in the replicates (i.e. horizontal axis indicates maximum proportion of the samples wherein the respective species were found)

Fig. 5. Maximal deviation of the average abundance and Shannon-Wiener diversity index (H') calculated as percentage over the respective values for 70 replicates

cies (16%) occurred in only 1 sample and 75% of the species were present in less than half of the samples. In total, 70% of the species richness contributed less than 5% to the total macrofaunal abundance.

Abundance and diversity

Typically (Fig. 4), as the number of replicates increases the average total abundance or diversity tends to stabilize, fluctuating around the average value for the 70 replicates. The results of different permuta-
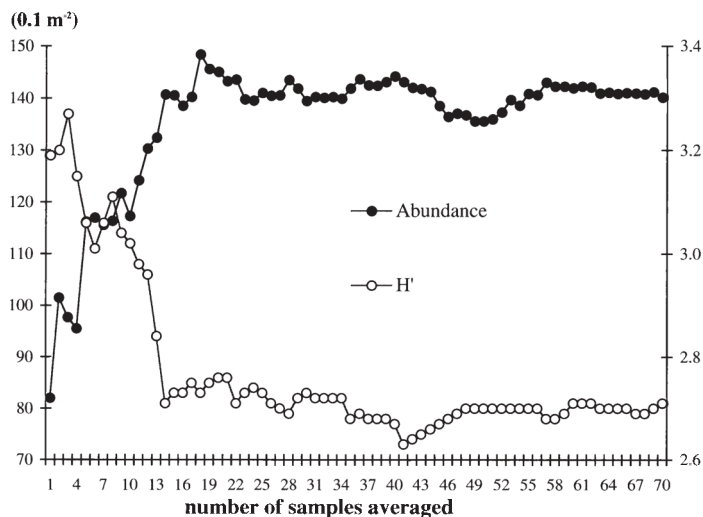

Fig. 4. Abundance and Shannon-Wiener diversity index, H' (per sample), averaged over a random series of replicates
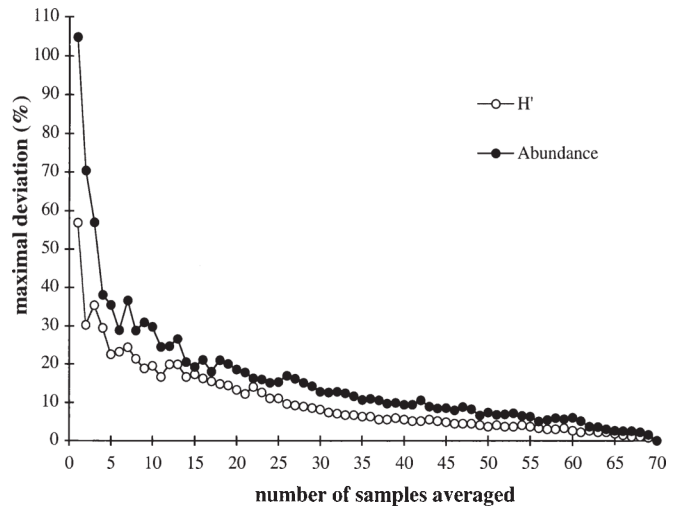
tions in the series of samples (Fig. 5) showed that there is a decreasing potential for deviation from the true value (measured as percent difference between the estimated value for a given number of samples and the value for all 70 replicates) with increasing sampling size for both abundance and diversity. In this type of analysis, the minimal deviation is obviously zero. The estimates of the average abundance and of the average Shannon-Wiener diversity index for 5 replicates could result in less than 40% deviation from the true value of these variables, whereas it could be less than 30% for 10 replicates. However, it seems that there is a need for more than 60 samples in order to avoid the possibility of deviation higher than 5%.

Species richness

Species richness was found to be an increasing function of sampling effort (Fig. 6) that did not level off until the end of the horizontal axis. The first 5 samples cover only 55% of the species found in 70 replicates, whereas in order to sample 70 or 90% of the species, it takes 12 and 35 samples respectively.

The estimates of species richness (Fig. 6) obtained through jackknife and the standard configuration of S_∞ were found to be highly dependent on the species already available in the data set, whereas they showed no signs of approaching the asymptote before the actual curve of species present in the data set. In fact, the jackknife exceeded in all cases the number of species present in the community and was approximately parallel to it after 30 replicates. The curve pro-
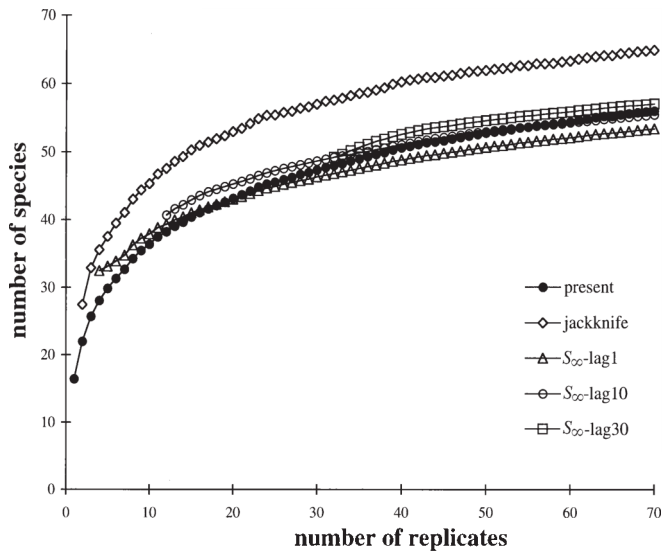
Fig. 6. Cumulative number of species present in an increasing number of replicates and prediction of the total species richness by the jackknife estimate and 3 configurations of the S_∞ method using different sample lag

Fig. 8. S_∞ estimate for different configurations regarding sample lags and truncation of the randomized sample series

duced through the standard configuration of S_∞ (Fig. 6: S_∞lag 1) was much closer to the actual species richness, whereas towards the end of the line it seemed to underestimate the number of species by approximately 3.

The S_∞ estimate (for all 70 replicates) under different sampling lags (Fig. 7) showed that there is a considerable increase in the expected species richness with increasing lag that becomes relatively stable after 30. The decrease found after the 59 lag is probably due to the very low number of points considered for the regression.
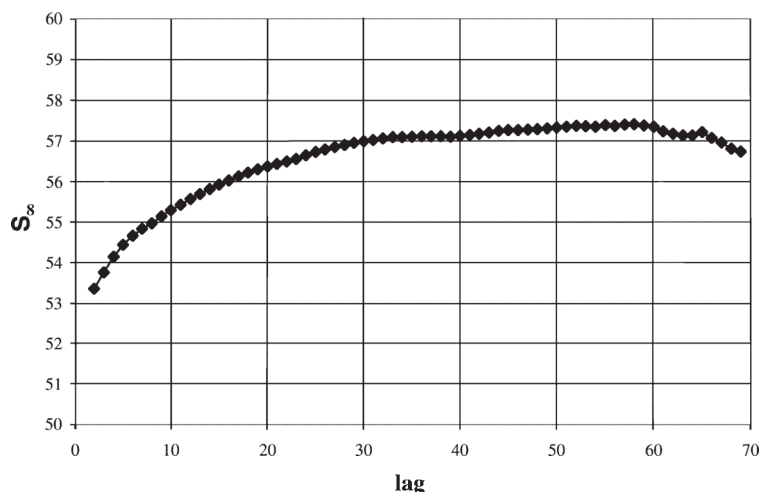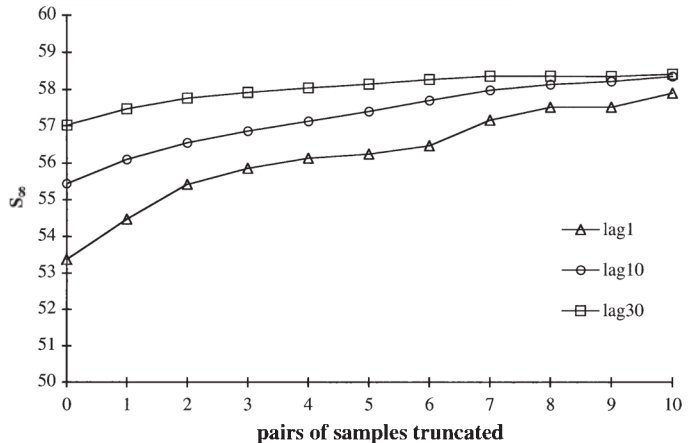
Different combinations of sampling lag and truncation of the entire data set were used in order to provide the best configuration for the S_∞ method (Fig. 8). It was found that both truncation and sampling lag lead to an increase in the species richness predicted by S_∞ in comparison to the standard configuration (SL = 1, no truncation). However, it was found that both treatments tend to produce similar results exceeding the initial estimate by approximately 8%, whereas when truncation exceeds 7 the maximal difference among all configurations is less than 1 species.

DISCUSSION

The sampler used in the present study was a 1000 cm^2 van Veen grab (modified according to Dybern et al. 1976), which is a standard means of sampling for numerous benthic surveys in the Baltic and around the world. Although Downing (1980) has emphasized the efficiency of samplers of smaller unit area for benthic abundance estimates, there is still no consensus as to how the use of such smaller samplers would affect the estimates of species richness, whereas the edge-effects of the sampler become more important as the area of the sampler decreases (Gray 1984). According to Riddle (1989), a sampler of smaller surface area could provide a saving of total area sampled of about 20%.

In terms of abundance, the function provided by Downing (1979) worked

Fig. 7. S_∞ estimate for different lags between successive samples

relatively well with the data of the present study, predicting that $\log_{10}$ of the standard deviation would be 2.5 whereas the actual value found (calculated per m^2) was 2.7. This is an indication that the distribution of abundance was not 'atypical' in comparison to other benthic studies.

The coefficient of variation (standard deviation/ mean) for the abundance data in the present study was 0.39 (mean = 140.1, standard deviation = 54.5). Applying these figures and the formula (Andrew & Mapstone 1987, Riddle 1989) relating the number of replicates required to estimate the mean abundance as a function of the desired precision to the mean and the sample variance, it was found that 4 samples (actually 3.9) would be enough to get a 20% precision of the mean. However, the results of the 100 randomized sequences of samples (Fig. 5) showed that abundance could deviate by more than 20% when the number of replicates was lower than 15. In this context, the formula for calculating the required number of samples provided by Krebs (1989) involving also 95% confidence limits was more realistic, since it would predict that 12 to 15 samples are required in order to achieve ±20% relative precision of the mean abundance estimate.

The presence of singletons (or rare or unique species), i.e. species occurring in 1 and only 1 quadrat (Krebs 1989), is a typical feature of collections of ecological samples. The presence of these species is in fact quite an important piece of information, allowing the prediction of the total number of species in the community. Both jackknife (Heltshe & Forrester 1983a) and S_∞ (Karakassis 1995) estimates of species richness are directly or indirectly influenced by the number of singletons in the community. It would be expected that in a relatively homogeneous area, in a species-poor macrobenthic community, and after such an exhaustive sampling, singletons would tend to approach zero. However, 9 species (16% of the species richness) were found in only 1 sample in the entire data set, whereas during the randomization process their percentage in any subset of samples was never lower than 6.5% of the respective species richness

Based on 12 macrofaunal replicates from the German Bight (Gerlach 1972), Gray (1984) assumed that for this particular area more than 60 and 90% of the species in the community would be collected with 2 and 7 replicates respectively. Holme (1953) sampled benthic macrofauna at 4 sites off Plymouth, using multiple replication (15 to 40 samples) with different sampling devices (200 cm^2 suction corer, 500 cm^2 scoop sampler and 1000 cm^2 Petersen grab). In none of these sites had the number of species leveled off, and the number of singletons at the most intensively sampled site was not less than 30% of the total number of species found in the 40 replicates.

Gentil & Dauvin (1988) presented a large data set of 32 replicates taken at the Bay of Morlaix (Station Pierre Noire) by means of a similar sampling device (0.1 m^2, Smith-McIntyre grab). The total number of species found within this data set (91) was higher than that in the present study and the number of singletons encountered was also high. Gentil & Dauvin suggested that the occurrence of rare species is largely unpredictable and therefore it is only possible to estimate the number of 'non rare' species. These authors partly attributed the difficulties in estimating the species richness of the entire community to particular characteristics of the site investigated, such as the macrotidal regime and the high diversity in the area which allows for the existence of species that are largely dispersed and of such low abundance that they rarely appear in the samples. This omission strategy was also recommended by Riddle (1989), who also tested the effect of different samplers (grabs and cores) on the estimation of species richness. Riddle emphasized the fact that there are more singletons in marine benthic communities than in freshwater ones.

Cao et al. (1998), in a river benthic data set of 24 replicates, also found a large number of singletons (ca 30%), and they tested how omission of rare (at different levels of rarity) species from the data sets would affect comparisons among different sites. They concluded that rare species are critical for accurate community studies and bioassessment. Cao et al. also emphasized the idea that although exclusion of rare species may be desirable for purely statistical reasons, this procedure seriously violates general ecological observations and theory, leading to an unacceptable loss of ecological information.

The results from the present study show that a large number of rare species may be present even in a relatively species-poor, microtidal, largely homogeneous benthic community and even after taking 70 true replicates, i.e. after sampling 7.0 m^2 of the sea bed.

The results of both methods for predicting species richness in the entire community, namely jackknife and S_{∞}, were not independent of the actual numbers already found in the data set. The jackknife estimate has been found to overestimate the number of species (Heltshe & Forrester 1983b, Karakassis 1995). As the number of species in the data set increases, this estimate tends to become equal to $S + u$ where S = number of species in the data set and u = number of unique species. However, since the number of the unique species (or even their proportion in the species richness) does not seem to decrease with increasing sampling effort, it is quite difficult to conclude on the reliability of the jackknife estimate results.

The main criterion for assessing the efficiency of a predictive method for species richness would be that

the cumulative curve produced should rise early (with the fewest samples possible) and level off before the actual species curve reaches the asymptote. From this point of view it was found that the predictive efficiency of the S_∞ method increased with increasing SL (Fig. 6), although the increase in SL inevitably delays the starting point, therefore limiting the potential of the method.

Different combinations of sampling lag and truncation in the entire data set were used in order to provide the best configuration for the S_∞ method (Fig. 8). It was found that both truncation and sampling lag result in increasing and stabilizing the species richness predicted by S_∞ in comparison to the standard configuration (Karakassis 1995), becoming higher than the actual number of species present in the data set and lower than the jackknife estimate. Of course, this new configuration increases the number of samples required in comparison to the standard configuration, since the increase in both truncation and the sampling lag results in a lower number of pairs for the S_∞ regression. Therefore, in order to obtain the best possible information out of the S_∞ method, it is suggested that a set of different configurations (depending on the number of available replicates) involving both truncation and increasing lag could be used in order to locate the point where the estimates tend to stabilize.

The idea that with a decent number of replicates biodiversity will be adequately sampled does not seem to be realistic. With 5 replicates we could only include 53 % of the species present in 70 replicates and undoubtedly it could be expected that the number of species found in 70 replicates is only a proportion of the respective number for 100 or 200 replicates. Furthermore, the number or the proportion of singletons did not show any marked decrease with increasing sampling effort. Therefore, it may be concluded that selecting the number of replicates required should not be based on the criterion of including a large proportion of the community species in the set of samples (Gray 1984). One could be fairly confident that the abundant species are included in the data set with relatively low sampling effort, but it seems extremely laborious to take enough samples so as to include in the data set all the species of the community, even for calibration purposes. However, it is recommended that a large number of replicates is taken before standard monitoring starts so that at least the precision of the estimates is known.

This inadequacy in determining the number of species in a community is an important problem for the monitoring of macrofaunal biodiversity at the species level, since it is quite a difficult task to compare 2 numbers (e.g. temporal changes in biodiversity) when neither of them is known. In fact, we may only compare numbers for the most abundant species and then we may only assume (or hope) that the abundant species, play a more important role than the rare ones and, therefore, since the former are included in our calculations, that these conclusions are meaningful. In this context, the measurable part of species richness refers to 'ecodiversity' (Margalef 1997) rather than to biodiversity. However, it should still be kept in mind that not all important species (e.g. keystone species) are abundant or of high biomass, and therefore estimates of species richness do not necessarily reflect ecological importance.

On the other hand, average H' diversity and abundance per m^2 may be estimated fairly easily with a reasonable number of replicates, and therefore the information contained in such estimates is quite reliable. In this context, the number of replicates required largely depends on the type of hypothesis tested, always being a compromise between precision and the time taken to obtain and process the samples (Gray 1984). However, a preliminary survey with a large series of replicates allows the determination of some important variables such as the standard deviation of the abundance or diversity so that differences between sets of samples are adequately documented.

LITERATURE CITED

Andrew NL, Mapstone DB (1987) Sampling and the description of spatial pattern in marine ecology. Oceanogr Mar Biol Annu Rev 25:39–90

Arntz WE (1981) Zonation and dynamics of macrobenthos biomass in an area stressed by oxygen deficiency. In: Barrett GW, Rosenberg R (eds) Stress effects on natural ecosystems. Wiley & Sons Ltd, London, p 215–225

Arrhenius Q (1921) Species and area. J Ecol 9:95–99

Bray JR, Curtis JT (1957) An ordination of the upland forest communities of southern Wisconsin. Ecol Monogr 27: 320–349

Cao Y, Williams DD, Williams NE (1998) How important are rare species in aquatic community ecology and bioassessment? Limnol Oceanogr 43:1403–1409

Downing JA (1979) Aggregation, transformation, and the design of benthos sampling programs. J Fish Res Board Can 36:1454–1463

Downing JA (1980) Corrections to recent publications. Aggregation, transformation, and the design of benthos sampling programs. Can J Fish Aquat Sci 37:1333

Dybern BI, Ackefors H, Elmgren R (1976) Recommendations

on methods for marine biological studies in the Baltic Sea. Baltic Mar Biol Publ 1:98

Field JG, Clarke KR, Warwick RM (1982) A practical strategy for analyzing multispecies distribution patterns. Mar Ecol Prog Ser 8:37–52

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

Gerlach SA (1972) Die Produktionsleistung des Benthos in der Helgoländer Bucht. Verh Dtsch Zool Ges 65:1–13

Gray JS (1984) Ökologie mariner Sedimente. Springer Verlag, Heidelberg

Heltshe JF, Forrester NE (1983a) Estimating species richness using the jackknife procedure. Biometrics 39:1–11

Heltshe JF, Forrester NE (1983b) Estimating diversity using quadrat sampling. Biometrics 39:1073–1076

Holme NA (1953) The biomass of the bottom fauna in the English channel off Plymouth. J Mar Biol Assoc UK 32:1–49

ICES (1996) Report on the ICES/HELCOM workshop on quality assurance of benthic measurements in the Baltic Sea. Int Counc Explor Sea Comm Meet 1996/E:2

Karakassis I (1995) S_∞: a new method for calculating macrobenthic species richness. Mar Ecol Prog Ser 120:299–303

Krebs CJ (1989) Ecological methodology. Herper & Row, New York

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

Margalef R (1997) Our biosphere. In: Kinne O (ed) Excellence in ecology, Book 10. Ecology Institute, Oldendorf/Luhe

Morrisey DJ, Howitt L, Underwood AJ, Stark JS (1992) Spatial variation in soft sediment benthos. Mar Ecol Prog Ser 81: 197–204

Riddle MJ (1989) Precision of the mean and the design of benthos sampling programmes: caution advised. Mar Biol 103:225–230

Rosenzweig ML (1995) Species diversity in space and time. Cambridge University Press, Cambridge

Rumohr H (1999) Soft bottom macrofauna: collection, treatment and quality assurance of samples. ICES Tech Mar Environ Sci 27:1–19

Shannon CE, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urbana

Weigelt M (1991) Short- and long-term changes in the benthic community of the deeper parts of Kiel Bay (Western Baltic) due to oxygen depletion and eutrophication. Meeresforschung 33:197–224

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

Williams CB (1943) Area and the number of species. Nature 152:264–267