

Differences in performance among four indices used to evaluate diversity in planktonic ecosystems

Grégory BEAUGRAND^{a,b*}, Martin EDWARDS^a

^a Sir Alister Hardy Foundation for Ocean Science, The Laboratory, Citadel Hill, Plymouth PL1 2PB, United Kingdom

^b Observatoire des sciences de l'univers, Laboratoire d'océanographie biologique et d'écologie du plancton marin, BP 28, 06234 Villefranche-sur-Mer cedex, France

Abstract – The performance of four common estimators of diversity are investigated using calanoid copepod data from the Continuous Plankton Recorder (CPR) survey. The region of the North Atlantic and the North Sea was divided into squares of 400 nautical miles for each 2-month period. For each 144 possible cases, Pielou's pooled quadrat method was performed with the aims of determining asymptotic diversity and investigating the CPR sample-size dependence of diversity estimators. It is shown that the performance of diversity indices may greatly vary in space and time (at a seasonal scale). This dependence is more pronounced in higher diverse environments and when the sample size is small. Despite results showing that all estimators underestimate the 'actual' diversity, comparison of sites remained reliable from a few pooled CPR samples. Using more than one CPR sample, the Gini coefficient appears to be a better diversity estimator than any other indices and spatial or temporal comparisons are highly satisfactory. In situations where comparative studies are needed but only one CPR sample is available, taxonomic richness was the preferred method of estimating diversity. Recommendations are proposed to maximise the efficiency of diversity estimations with the CPR data. © 2001 Ifremer/CNRS/IRD/Éditions scientifiques et médicales Elsevier SAS

Résumé – **Comparaison de la performance de quelques indices de diversité en milieu pélagique.** Les performances de quatre indices de diversité sont étudiées utilisant les données relatives aux copépodes calanoïdes provenant du programme *Continuous Plankton Recorder* (CPR). L'Atlantique Nord et la mer du Nord sont divisés en carré de 400 milles nautiques de côté et une grille spatiale est réalisée pour chaque période de 2 mois. Dans chaque carré, la méthode des quadrats cumulés de Piélou est employée avec pour objectif l'estimation de la diversité asymptotique et l'étude de la dépendance vis-à-vis de l'effort d'échantillonnage de certains estimateurs de la diversité. La performance et la robustesse des indices peuvent varier dans le temps (à l'échelle saisonnière) et dans l'espace. Cette dépendance est plus prononcée dans les milieux riches en espèces et lorsque la taille de l'échantillon est petite. Bien que tous les estimateurs sous-estiment la diversité réelle du milieu, la comparaison de sites reste possible dès un nombre faible d'échantillons CPR cumulés. Utilisant plusieurs échantillons CPR, le coefficient de Gini est le meilleur des indices étudiés et les comparaisons spatiales ou temporelles donnent des résultats très satisfaisants. Cependant, lorsqu'un seul échantillon CPR est considéré, l'utilisation de la richesse taxonomique devra être préférée pour des études comparatives. Des recommandations sont proposées pour optimiser les estimations de diversité lorsque les données du *Continuous Plankton Recorder* sont utilisées. © 2001 Ifremer/CNRS/IRD/Éditions scientifiques et médicales Elsevier SAS

Continuous Plankton Recorder survey / copepods / diversity indices / sample size / sensitivity

programme continuous plankton recorder / copépodes / indices de diversité / taille de l'échantillonnage / sensibilité

*Correspondence and reprints.

1. INTRODUCTION

Since evidence has arisen that human influence is strongly modifying the biosphere, the assessment of global biodiversity has become an important issue in ecology (Lovejoy, 1994; Williamson, 1998). At an organismal level (Harper and Hawksworth, 1994), diversity can be assessed in several ways, for example by fitting models of relative abundance, by extrapolating species accumulation curves or by using diversity indices (Magurran, 1988; Colwell and Coddington, 1994; He and Legendre, 1996). The latter has been the most commonly followed approach in the marine realm (particularly in benthic ecology) to detect effects of pollution on natural communities (Warwick and Clarke, 1991; Heip et al., 1998) and in conservation biology (Magurran, 1988).

However, diversity indices are inherently sensitive to the sampling size involving difficulties in comparing different locations or the same site at different time. Despite numerous studies on sensitivities of these estimators (Soetaert and Heip, 1990; Palmer, 1991), only a few studies have focussed on the variation of this dependence in space and time. This is however relevant if anyone desires to compare diversity of a location with another, to examine temporal changes in diversity or to map diversity of a taxonomic group in a region. In addition, most of these studies were theoretical or based on simulated data that may have badly mimicked natural communities (Tong, 1983; Bunge and Fitzpatrick, 1993; Lande, 1996). This must be particularly apparent for the pelagic domain and in temperate regions where the distribution of organisms is patchy and where seasonal variability is high.

The Continuous Plankton Recorder (CPR) survey, which began in 1931, has routinely identified approximately 400 planktonic taxa at monthly interval in the North Atlantic and the North Sea. Thus, a large amount of samples has been collected at an oceanic scale and represents an invaluable database that has been largely used to investigate relationships between climate, hydrography and plankton (e.g. Fromentin and Planque, 1996; Reid et al., 1998; Beaugrand et al., 2000). However, as the CPR sampling procedure is unique, it is important to evaluate the performance of diversity indices from such data. In this paper, the performance of some commonly used diversity indices were investigated through space and time with the following questions addressed:

- Which diversity indices are the most sensitive to the CPR sampling size?
- How does this sensitivity change with the CPR sample size?
- Are results reproducible from one region/month to another?
- Despite the bias of estimators, is it possible to make comparisons between regions or through months?

Finally, implications of these results for the study of spatial and temporal changes in pelagic diversity from CPR data will also be discussed.

2. MATERIALS AND METHODS

2.1. Sampling

The CPR survey is a plankton monitoring programme which has been mainly run in the North Atlantic ocean and the North Sea. Samples are collected by a high-speed plankton recorder (about 20 km·h⁻¹) towed behind ‘ships of opportunity’ at a depth of approximately 10 m (Hays, 1994; Warner and Hays, 1994). Water enters the recorder through a square aperture of 1.62 cm² and is filtered by a continuously moving band of silk with an average mesh size of 270 µm (Hays, 1994). A second band of silk covers the organisms that are reeled into a tank containing formaldehyde. On return to the laboratory, the silk is unwound and cut into sections corresponding to 18.5 km of tow that correspond to one sample and approximately 3 m³ of filtered seawater. Positions and times of samples are estimated from knowledge of the beginning and the end of deployment, changes in the CPR course and speed of the ship. Calanoid copepods were selected for this study because it is a taxon for which CPR sampling is the most quantitative. In most cases (92.5 %), identification goes to species levels except for some small calanoids which are identified to genus. Due to this procedure, diversity indices calculated from one CPR sample underestimates total species diversity of the epipelagic environment. However, as the identification procedure has not been modified since 1958, the underestimation is consistent and does not prevent comparisons between sites or for different time periods.

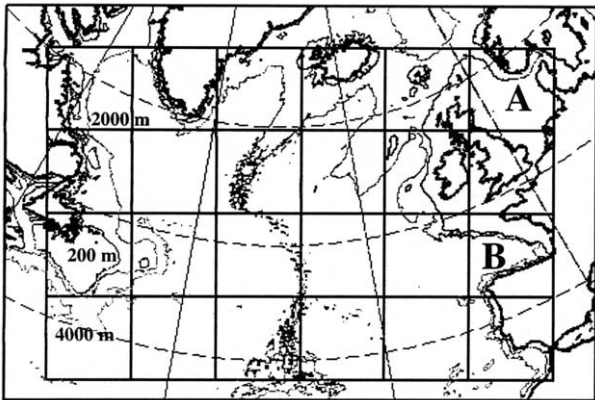


Figure 1. Continuous Plankton Recorder survey area divided into squares of 400 nautical miles for each 2-month period, integrating 40 years of sampling (from 1958 to 1997). Squares A (North Sea, lower diversity) and B (Bay of Biscay, higher diversity) represent regions used in the examples shown by figures 2, 3 and 4. Bathymetry is indicated.

2.2. Analysis procedure

2.2.1. Building of the spatio-temporal grid

The survey area was first divided into squares of 400 nautical miles (figure 1). The square size was chosen as a compromise between the spatial resolution and the number of samples needed to perform the analyses. Six different spatial grids, one for each 2-month period, were built to take seasonal variations into account. To eliminate variability due to diel vertical migrations of many calanoid species, only night samples were used. Thus, the grids consisted of 144 squares (24 squares per grid \times 6 different 2-month periods), which encompassed different pelagic assemblages of calanoids, from neritic to oceanic regions and took into account latitudinal spatial and seasonal variations.

2.2.2. Selected species diversity indices

Four non-parametric measures of diversity were used: taxonomic richness, Shannon information index, Gini coefficient and Brillouin diversity index. Taxonomic richness is merely a measure of the number of taxa in a defined sampling unit. Proposed by Margalef (1958) to be used as an diversity index, the Shannon information index (Shannon and Weaver, 1962) calculates the average information per individual:

$$H' = - \sum_{i=1}^s p_i \ln p_i$$

where p_i is the proportion of individuals found in the i th species and s the total number of species in the sample. \ln is the natural logarithm.

The Brillouin index (Brillouin, 1956) was also selected because it is more appropriate when estimation is based on samples for which random-sampling can not be guaranteed (Magurran, 1988). This is calculated as follows:

$$HB = \frac{\ln N! - \sum_{i=1}^s \ln n_i!}{N}$$

where n_i is the number of individuals in the i th species and N the total number of individuals in the sample. The Brillouin index is more exact than the Shannon index since it corresponds to a sampling without replacement in contrast to the Shannon information index (Legendre and Legendre, 1998).

Simpson's concentration index λ (Simpson, 1949), for a finite community is calculated by:

$$\lambda = \sum_{i=1}^s \frac{n_i(n_i - 1)}{N(N - 1)}$$

Instead of using the Simpson's concentration index, $1-\lambda$ also known as the Gini coefficient was used (Lande, 1996). It corresponds to the probability that two randomly chosen individuals from a given community are different species. Furthermore, it has been demonstrated that for this index a non-biased estimator exists (Lande, 1996).

2.2.3. Pielou's pooled quadrat method

For each square of the six grids, Pielou's pooled quadrat method was applied (Pielou, 1966, 1977). The goal of this analysis was first to determine the nearest approximation of the asymptotic diversity for all estimators. This value was then considered as the reference against which sampling-size dependence of the diversity indices were investigated. This method consists of pooling samples in a random sequence, re-estimating the abundance of species (in this case, 108 taxa) and re-calculating Brillouin index for each pooled sample. The Brillouin cumulative diversity HB_k is then plotted against the number of samples k and the portion where the curve flattens off allows the determination of the community's diversity

(Pielou, 1977; Magurran, 1988). As in the building of a species accumulation curve, the order in which samples are added affects the shape of the curves and the estimation of diversity (Palmer and White, 1994). That is why 100 curves were randomly generated and the mean value was taken for each pooled sample. In addition, according to Magurran (1988), this method can also be used with other diversity indices. Thus, it was applied for the four selected diversity estimators. Asymptotic diversity is assessed in two steps:

– Step 1. From $k = t + 1$ to $k = z$, where t is the number of pooled samples to be needed for the curve to flatten off and z the total number of pooled samples, h_k is calculated from the formula:

$$h_k = \frac{N_k D_k - N_{k-1} D_{k-1}}{N_k - N_{k-1}}$$

where D_k is the estimated diversity and N_k the number of individuals for the k th pooled sample.

– Step 2. Inference on the diversity of the community is given by:

$$H = \frac{1}{z - (t + 1)} \sum_{k=1}^z h_k$$

Calculations were based on a matrix of 50 384 samples \times 108 taxa for each square and for each 2-month periods for the four selected diversity indices, giving a total of $24 \times 6 \times 4 = 576$. Thus, it was not possible to graphically determine t for all these cases. Hence, t was fixed to 200 samples from empirical examination and z had a minimum of 250 values. As the maximum number of samples per curve was 300, $z - t$ varied between 50 and 100. Even graphically, the determination of t is subjective (Magurran, 1988) but Pielou (1975) assumed that the estimation of H is still accurate if the values of h_k between t and z are not serially correlated (Magurran, 1988).

2.2.4. Determination of the number of samples needed to reach a given value of diversity

The determination of the number of samples needed to reach the asymptotic diversity (parameter t) from Pielou’s method is graphic. Due to the high number of cases considered in this study, it was not possible to

determine the sensitivity of diversity indices to sampling size in this manner. Instead, it was evaluated numerically. Different models already exist that could have been used to answer this question. For example, the Michaelis-Menten model has often been used to estimate the richness from a species-area curve. By fitting this model to the data, inference could be made to determine the number of samples needed to reach the asymptotic value or a fraction of that. However, Keating (1998) showed that its performance largely depends on the community structure and that the best predictions are achieved when the community has species-abundance distributions based on MacArthur’s broken-stick model with a high number of species, a situation rarely met in natural communities. This same constraint occurs for other models (He and Legendre, 1996). Due to the large number of cases to compare, it was not reliable to use such models. Instead, an original analysis was developed. It consisted of making step-by-step calculations of the average slope S_i for the i th pooled samples in the diversity-sample curve.

Thus, for $p < S_i < \max - p$, S_i was estimated from:

$$S_i = \sum_{k=v}^{k=p} \frac{y_{i+k} - y_{i-k}}{x_{i+k} - x_{i-k}}$$

where y_{i+k} and y_{i-k} represented the superior and inferior value of diversity, respectively and similarly, x_{i+k} and x_{i-k} the number of pooled samples. p is the maximum number of samples above and below the pooled samples i (in this case $p = 10$) and v the number of samples from which the calculation of the slope starts at the pooled sample i (in this case $v = 2$). Thus, at a pooled sample i , slopes are computed in the interval $[i+v, i+p]$ for superior values and $[i-v, i-p]$ for inferior ones. Then, the sum of all these slopes directly gives the average slope S_i . No weight was attributed in the calculation of S_i since each individual slope was judged to be equally important. So, this average slope was calculated for each pooled samples from $p = 10$ to a maximum of 300 samples and the number of samples needed to reach a particular value of this slope (from 0.005 to 0.001). These latter values were selected empirically. It represented a compromise between a slope tending to 0 and the efficiency to discriminate between estimators.

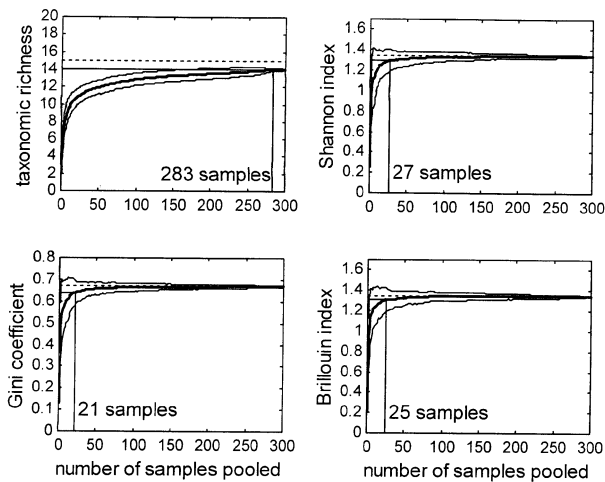


Figure 2. Diversity-sample curves for calanoid copepods in the North Sea for the period July-August (see *figure 1*). Each curve in bold represents an average of 100 other ones generated by randomising the sample pooling order. Each finer curve shows the confidence interval (\pm one standard deviation). For each case, a dashed line denotes the estimated value by using Pielou's pooled quadrat method. The number of samples needed to reach an average slope of 0.002 is given for each case (see Materials and Methods).

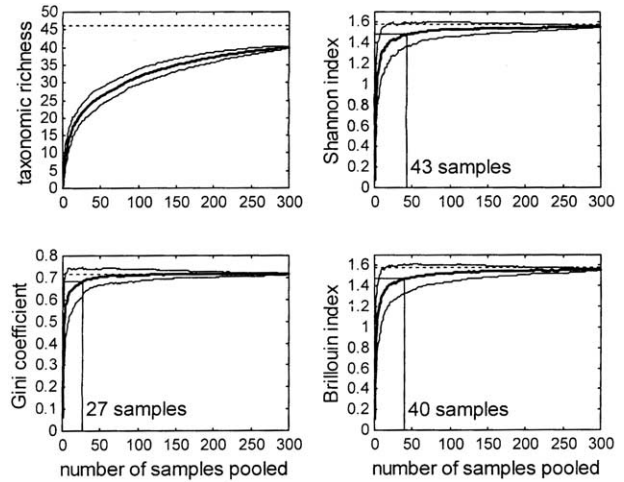


Figure 3. Diversity-sample curves for calanoid copepods in the Bay of Biscay for the period July-August (see *figure 1*). Each curve in bold represents an average of 100 other ones generated by randomising the sample pooling order. Each finer curve shows the confidence interval (\pm one standard deviation). For each case, a dashed line denotes the estimated value by using Pielou's pooled quadrat method. The number of samples needed to reach an average slope of 0.002 is given for each case (see Materials and Methods). The taxonomic richness did not reach the average slope of 0.002.

3. RESULTS

3.1. Case of the North Sea and the Bay of Biscay in summer

Figure 2 shows diversity-sample curves for all selected indices, for the North Sea (low diversity) and the period July-August. An asymptote is clearly reached for the Shannon and Brillouin diversity indices and the Gini coefficient with values calculated from Pielou's pooled quadrat method of 1.347, 1.346, 0.670, respectively. Diversity-sample curve for the taxonomic richness does not reach an asymptote even for 300 pooled samples. Its value was estimated to fifteen species. In this case, the Gini coefficient was the least sensitive to the sampling size with 21 samples at the average slope of 0.002. The Brillouin index (25) seems to be less influenced than the Shannon index (27) by sample size but the difference was weak. A large number of samples (283) is needed to estimate the taxonomic richness.

In the Bay of Biscay in July-August (*figure 3*), the asymptotic diversity of all estimators is higher than it was in the North Sea for the same months. For example, the

Shannon index had an asymptotic value of 1.347 in the North Sea (*figure 2*) whereas this value reached 1.61 in the Bay of Biscay (*figure 3*). Diversity indices also reach an asymptotic value whereas the taxonomic richness continues to increase regularly. The Gini coefficient, which needed 27 samples to reach an average slope of 0.002, is again the least sensitive to sample size, followed by the Brillouin index (40) and the Shannon index (43). In this higher diversity region, estimators require more samples to reach an average slope of 0.002 than in the North Sea (*figures 2, 3*).

Figure 4 displays changes in the average slope of diversity indices for the first thirty pooled samples in the North Sea (*figure 4a*) and the Bay of Biscay (*figure 4b*). It shows that comparison of the sensitivity of each estimator at each point of the diversity-sample curves is still valid until a slope of 0.001. Thus, it indicates that we can be confident, by using this method and by selecting the average slope, of determining the sensitivity of diversity estimators. The Gini coefficient rapidly reaches a low average slope for both situations. This is slower for the Brillouin and Shannon indices which are very close. In the North Sea (lower diversity), differences between

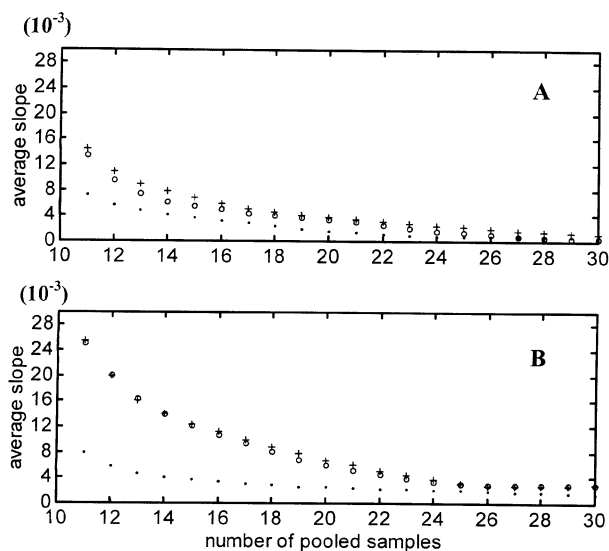


Figure 4. Changes in the average slope of diversity-sample curves with the pooling of samples for the Shannon index (+), the Gini coefficient (.) and the Brillouin index (°). **A.** The North Sea in July-August. **B.** The Bay of Biscay in July-August.

estimators are less pronounced than in the Bay of Biscay where diversity is higher.

Based on only one CPR sample, all estimators underestimate asymptotic diversity for both examples (table I). The taxonomic richness only estimates 24.47 % of the ‘actual’ value in the North Sea and 7.03 % in the Bay of Biscay. The Gini coefficient reaches 53.84 % of the

‘actual’ diversity in one sample in the North Sea and 43.95 % in the Bay of Biscay. With ten pooled samples, underestimation is reduced significantly except for the taxonomic richness (table I). In the Bay of Biscay, the best estimator was still the Gini coefficient but in the North Sea the Brillouin index seemed to perform better than other estimators.

3.2. Comparison between estimated and asymptotic diversity for all grid squares and every 2-month period

From both examples and for all estimators, it was clear that the use of a restricted number of CPR samples may imply an important underestimation of diversity of the pelagic community. Table I shows that the Gini coefficient performs better on average than the other indices. Based on one CPR sample, it estimated on average 47.46 % of the asymptotic diversity although the Shannon index only reached 36.80 %, the Brillouin index 33.23 % and the taxonomic richness 9.83 % (table I). This last estimator highly underestimated the asymptotic diversity. However, this does not inevitably mean that it is not possible to use these estimators for comparative studies. Figure 5 shows the diagram of dispersion of asymptotic diversity calculated from Pielou’s method versus the estimated diversity from only one CPR sample for all cases. The taxonomic richness appears to be the best estimator, allowing comparison between sites

Table I. Values of diversity indices expressed as percentage of the asymptotic diversity, based on one and ten pooled CPR samples. First column: special case of the North Sea in July-August. Second column: special case of the Bay of Biscay in July-August. Third column: Mean percentage for all cases corresponding to the squares of all six grids (see figure 1) in which the number of samples was sufficient to estimate asymptotic diversity (60 on a total of 144 cases). The last column indicates the number of times where the performance of an estimator was ranked at the first and second position (into brackets).

	North Sea (July-August)	Bay of Biscay (July-August)	Mean percentage (60 squares)	Performance of estimators (first and second rank)
Values based on one sample				
Taxonomic richness	24.47 %	7.03 %	9.83 %	0 (0)
Shannon index	50.75 %	31.33 %	36.80 %	2 (44)
Gini coefficient	53.84 %	43.95 %	47.46 %	54 (6)
Brillouin index	49.91 %	36.09 %	33.23 %	4 (10)
Values based on ten pooled samples				
Taxonomic richness	63.55 %	32.22 %	32.99 %	0 (0)
Shannon index	90.39 %	83.34 %	81.67 %	5 (31)
Gini coefficient	89.56 %	89.12 %	90.47 %	53 (5)
Brillouin index	92.15 %	83.34 %	80.26 %	2 (24)

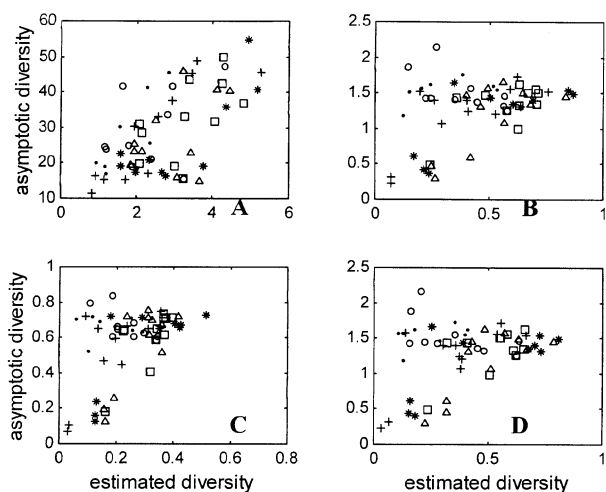


Figure 5. Diagrams of dispersion of asymptotic diversity against estimated diversity for one CPR sample. **A.** Taxonomic richness (Spearman rank correlation coefficient $r_s = 0.56$, probability $P < 0.001$). **B.** The Shannon index ($r_s = 0.22$, $P = 0.15$). **C.** The Gini coefficient ($r_s = 0.33$, $P = 0.03$). **D.** The Brillouin index ($r_s = 0.14$, $P = 0.36$). Each symbol denotes a particular square of grids built (see figure 1) for a particular period: January-February (.); March-April (+); May-June (*); July-August (triangle); September-October (square); November-December (circle). A total of 10 000 permutations of one of the original vector were realised to estimate the probability that the simulated correlation coefficient exceeds the original value.

or for different months to be made (Spearman correlation $r_s = 0.56$, probability $P = 0.0002$, degree of freedom $df = 60$). Gini coefficient was also correlated although its value was much weaker ($r_s = 0.33$, $P = 0.0347$, $df = 60$). Shannon and Brillouin diversity indices estimated from one sample are not significantly correlated with their asymptotic value. For these two indices, winter months contribute to a decrease in the correlation. However, if these months are eliminated from the calculation, the Spearman correlation coefficients all become highly significant (table II).

Table II. Spearman rank correlation coefficient for one and ten pooled samples without the months from November to February. The degree of freedom was equal to 42. A total of 10 000 permutations of one of the original vector were realised to estimate the probability that the simulated correlation coefficient exceeds the original value.

	Diversity based on one sample		Diversity based on ten pooled samples	
	Correlation	Probability ($\times 10^{-4}$)	Correlation	Probability ($\times 10^{-4}$)
Taxonomic richness	0.66	1	0.75	0
Shannon index	0.54	27	0.85	0
Gini coefficient	0.59	12	0.87	0
Brillouin index	0.49	41	0.82	0

If estimation is based on ten pooled samples, all estimators are highly significantly related to the asymptotic diversity (figure 6). The relationship between estimated and asymptotic diversity is less strong during winter months and the suppression of these months allowed the correlation to be improved (table II). The Shannon and Brillouin indices always underestimated 'actual' diversity although the Gini coefficient overestimated the asymptotic diversity in some cases (figure 6).

3.3. Sensitivity of estimators

Of a total of 144 possible situations, it was only possible to compare the sensitivity of diversity indices to the sample size in sixty squares. Table III summarises the results for different average slopes taken between 0.005 and 0.001. The Gini coefficient is by far the least sensitive diversity estimator with, for example, 20.75 samples needed to reach an average slope of 0.002 (table III). In a minimum of 85 % of all cases, this estimator reached a given average slope more rapidly on the diversity-sample curve. Taxonomic richness was the most influenced by sample size and it was possible to estimate the number of samples needed to reach a given average slope in few cases (a maximum of eight squares). The Shannon index seemed to be the least sensitive to sampling size when compared with the Brillouin index but differences were small (table III). Considering only diversity indices, the coefficient of variation is the lowest for the Gini coefficient and then for the Shannon index.

4. DISCUSSION AND CONCLUSION

Pielou's pooled quadrat method was applied to all estimators and situations. Although for diversity indices,

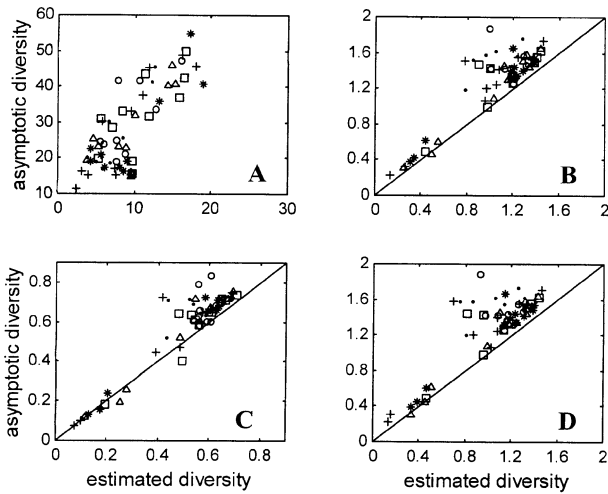


Figure 6. Diagrams of dispersion of asymptotic diversity against estimated diversity for ten CPR sample. Symbols that fall above the diagonal line are underestimates and those below are overestimates. **A.** Taxonomic richness (Spearman rank correlation coefficient $r_s = 0.69$, probability $P < 0.001$). **B.** The Shannon index ($r_s = 0.67$, $P < 0.001$). **C.** The Gini coefficient ($r_s = 0.72$, $P < 0.001$). **D.** The Brillouin index ($r_s = 0.61$, $P < 0.01$). Each symbol denotes a particular square of grids built (see figure 1) for a particular period: January-February (·); March-April (+); May-June (*); July-August (triangle); September-October (square); November-December (circle). A total of 10 000 permutations of one of the original vector were realised to estimate the probability that the simulated correlation coefficient exceeds the original value.

300 pooled samples were large enough to reach asymptotic diversity and thus achieve a satisfactory estimation of the diversity of the environment, this was not the case for taxonomic richness. In addition, in most cases, values of this latter estimator were serially correlated, which may have biased the estimation of asymptotic diversity (Magurran, 1988). Figures 2 and 3 show that the taxonomic richness is significantly above its diversity-sample curve. Other methods exist which could have also been used to estimate the asymptotic value of the taxonomic richness. For example, the Jack-knife procedure is thought to be a more robust procedure for clumped distributions and in some situations more accurate (Magurran, 1988). However, Pielou’s method was improved by generating 100 simulated curves, which reduced the effects of patchy distribution of calanoid copepods as well as improved its precision. Pielou’s method was also kept for two other reasons. (1) For this first product of the evaluation of the estimation of diversity with the CPR data, it was important to keep a graphical control of each estimation. (2) This method also

allowed variation in the efficiency of diversity indices with the sampling size to be investigated.

The size of squares was large (400 nautical miles) and the integration of 40 years of CPR sampling does not allow these squares to truly represent homogeneous habitats (α diversity). Each asymptotic value of diversity might have been influenced by the γ diversity (landscape scale). This is a tenuous problem in pelagic ecology and is difficult to resolve even in the terrestrial realm where investigation is easier (Palmer and White, 1994). Thus, the percentage of diversity estimated from a given sample size might have been underestimated. However, despite this constraint, it is quite satisfactory that from ten pooled CPR samples the estimated diversity reached 90 % with the Gini coefficient, and approximately 80 % with the Shannon and Brillouin indices. This phenomenon would have been less strong for diversity indices and more important for the taxonomic richness.

The reproducibility of results for different slopes (table III, figure 4) demonstrated that the average slope calculated for each diversity-sample curve could be applied in this study. Generally, when only a few species accumulation curves are considered, fitting by an appropriate model is more efficient, but in this study, it was not reliable to fit a particular model for all sixty cases. Thus when the use of such models is not workable, this method can represent an alternative solution to test the sensitivity of diversity estimators with the sampling size.

It is clear from this study that all estimators underestimated diversity if a limited number of samples (especially one) was used. It is obviously not restricted to the CPR sampling but it has been shown for both theoretical communities (Lande, 1996), and field communities from the terrestrial (Magurran, 1988; Palmer, 1990) to the marine realm (Smith and Grassle, 1977; Soetaert and Heip, 1990). Despite this limitation, comparisons are still possible even for one CPR sample for some estimators (figure 5). In addition, both underestimation and sensitivity of diversity indices rapidly decreased with a slight increase in the number of pooled CPR samples enabling satisfactory comparisons between sites or at different time periods.

Taxonomic richness is by far the most sensitive index estimating on average from one CPR sample only 10 % of the ‘actual’ value. Even if the sample size increased, underestimation was still important (32.99 % in ten pooled samples) in contrast to diversity indices (between

Table III. Mean number of samples needed to reach a given average slope of the diversity-sample curves for all squares of the six grids where the number of samples was sufficient to perform analyses. The coefficient of variation (in %), the number of values on which the average was based are also indicated. The number of time where a diversity index was the least sensitive is shown in the row 'first position' and the second least sensitive in the row 'second position'. All these parameters were calculated for values of the average slope from 0.005 to 0.001.

Slope	Parameters	Species richness	Shannon index	Gini coefficient	Brillouin index
0.005	Samples mean	191.12	22.43	14.85	23.95
	Variation coefficient (%)	23.46	39.36	25.79	41.87
	Number of values	8	60	60	60
	First position	0	9	51	0
	Second position	0	37	9	14
0.004	Samples mean	214.87	25.43	16.07	26.65
	Variation coefficient (%)	20.83	46.04	30.43	46.3
	Number of values	8	60	60	60
	First position	0	8	52	0
	Second position	0	33	8	19
0.003	Samples mean	211.33	29.1	17.7	31.02
	Variation coefficient (%)	20.6	49	33.05	50.58
	Number of values	6	60	60	60
	First position	0	5	55	0
	Second position	0	35	5	20
0.002	Samples mean	251	36.5	20.75	37.38
	Variation coefficient (%)	9.35	54.63	36.48	49.78
	Number of values	5	60	60	60
	First position	0	3	57	0
	Second position	0	32	3	25
0.001	Samples mean	–	50.48	27.65	51.18
	Variation coefficient (%)	–	48.69	52.33	47.62
	Number of values	0	60	60	60
	First position	–	2	55	3
	Second position	–	32	5	23

80 and 90 %). The Gini coefficient tends to be on average the least sensitive estimator. Of a total of sixty cases, it was 54 and 53 times the least biased of estimators from one and ten pooled samples, respectively. The Shannon index seems to be slightly less sensitive than the Brillouin index. This was an unexpected result. Maguran (1988) and Legendre and Legendre (1998) presented it as a more robust estimator than the Shannon information index to a non-random sampling. For all estimators tested here, the more diverse a community (e.g. Bay of Biscay), the more CPR samples needed to reach the asymptotic diversity. Taxonomic richness was still very sensitive to this characteristic although the Gini coefficient was the least affected index.

Despite systematic underestimations of all estimators, comparison between sites remained possible. If only one

CPR sample was used, taxonomic richness was the best index related to the estimated asymptotic diversity. The Gini coefficient was found significantly correlated with its asymptotic diversity but the relationship was weaker ($P = 0.035$). The Shannon and Brillouin diversity was not significantly related with their reference value. However, eliminating winter months (November to February) from correlations, all estimators were found highly correlated with their asymptotic diversity (*table II*). This increase, especially for diversity indices suggests that the performance of diversity indices is not constant throughout the months, particularly at the end of autumn and in winter when production and diversity is at a minimum. This influence is especially strong on the Shannon and Brillouin indices and the Gini coefficient which also integrate evenness in their calculation. Estimating diversity from one sample for winter months might lead to (1) an

underestimation of the species richness involving a higher bias on this estimator and (2) a rough estimation of the evenness because of the very diluted environment. Thus, for these months sampling effort should be increased. In only ten pooled samples, results show that this feature was considerably reduced (*figure 6*).

This study has clearly demonstrated that the performance of diversity estimators may vary in time (at a seasonal scale) and space. This effect was much more pronounced in higher diversity regions than in low diversity regions and when the sampling size was small. A similar conclusion was found by Soetaert and Heip (1990). Thus, diversity indices should be employed with extreme caution in a diluted environment like the pelagic realm.

Estimating diversity in the pelagic realm is particularly relevant when examining relationships between climate change, hydrography and the pelagic biota. It may also enable the evaluation of anthropogenic impact by over-fishing, introduction of exotic species and pollution in coastal regions. The CPR survey has monitored over 400 planktonic taxa since 1948 over the North Atlantic ocean and the North Sea at monthly intervals and is of great value in interpreting long-term changes in the marine environment. However, as the sampling procedure is unique to the CPR survey, diversity indices need to be evaluated to optimise the assessment of diversity. This study has demonstrated that the application of diversity indices on such a data set is reliable. However, to improve the quality of estimation, we recommend that when feasible, the use of more than one CPR sample before calculating diversity indices. It appears that ten samples give a very satisfactory result. When more than one CPR sample is used (e.g. ten samples), the Gini coefficient, the complement of the calculation of Simpson concentration index for a finite community, would be preferable as an assessment of diversity. This index emerged as the least sensitive estimator and gave (e.g. ten CPR samples) very satisfactory results for comparative studies in space and time. The Jack-knife procedure should also be used to improve estimates and to give a confidence interval. Using this protocol, comparison between sites or different times can be attempted by the analysis of variance. However, when it is not workable to use more than one CPR sample and/or when analysis is needed to be performed for all months, the taxonomic richness may be an alternative solution in a comparative study using CPR data.

Acknowledgements

The authors are grateful to all present members and supporters of the Sir Alister Hardy Foundation for Ocean Science whose continuous efforts have allowed the long-term establishment and maintenance of the CPR data-set. We are particularly grateful to F. Ibañez, P. Christopher Reid, Serge Dallot, J. Alistair Lindley, Thierry Comtet and one anonymous referee for advice and comments on the manuscript. This research was supported by the European Community Research Project No. MAS3-CT98-5058 and by the Programme National de l'Environnement Côtier.

REFERENCES

- Beaugrand, G., Ibañez, F., Reid, P.C., 2000. Spatial, seasonal and long-term fluctuations of plankton in relation to hydroclimatic features in the English Channel, Celtic Sea and Bay of Biscay. *Mar. Ecol. Prog. Ser.* 200, 93–102.
- Brillouin, L., 1956. *Science and Information Theory*. Academic Press, New York.
- Bunge, J., Fitzpatrick, M., 1993. Estimating the number of species: a review. *J. Am. Stat. Assoc.* 88, 364–373.
- Colwell, R.K., Coddington, J.A., 1994. Estimating terrestrial biodiversity through extrapolation. *Phil. Trans. R. Soc. Lond. B* 345, 101–118.
- Fromentin, J.M., Planque, B., 1996. *Calanus* and environment in the eastern North Atlantic. II. Influence of the North Atlantic oscillation on *C. finmarchicus* and *C. helgolandicus*. *Mar. Ecol. Prog. Ser.* 134, 111–118.
- Harper, J.L., Hawksworth, D.L., 1994. Biodiversity: measurement and estimation, Preface. *Phil. Trans. R. Soc. Lond. B* 345, 5–12.
- Hays, G.C., 1994. Mesh selection and filtration efficiency of the Continuous Plankton Recorder. *J. Plankton Res.* 16, 403–412.
- He, F., Legendre, P., 1996. On species-area relations. *Am. Nat.* 148, 719–737.
- Heip, C., Warwick, R.M., Carr, M.R., Clarke, R., Herman, P.M.J., Huys, R., Smol, N., Van Holsbeke, K., 1988. Analyses of community attributes: the benthic meiofauna of Langesund and Frierfjord, Norway. *Mar. Ecol. Prog. Ser.* 46, 171–180.
- Keating, K.A., 1998. Estimating species richness: the Michaelis-Menten model revisited. *Oikos* 81, 411–416.
- Lande, R., 1996. Statistics and partitioning of species diversity, and similarity among communities. *Oikos* 76, 5–13.
- Legendre, P., Legendre, L., 1998. *Numerical Ecology*. Elsevier Science, The Netherlands.
- Lovejoy, T.E., 1994. The quantification of biodiversity: an esoteric quest or a vital component of sustainable development? *Phil. Trans. R. Soc. Lond. B* 345, 81–87.
- Magurran, A.E., 1958. *Ecological Diversity and its Measurement*. University Press, Cambridge, 1988.
- Margalef, R., 1958. Information theory in ecology. *Gen. Syst.* 3, 36–71.

- Palmer, M.W., 1990. The estimation of species richness by extrapolation. *Ecology* 71, 1195–1198.
- Palmer, M.W., 1991. Estimating species richness: the second-order jackknife reconsidered. *Ecology* 72, 1512–1513.
- Palmer, M.W., White, P.S., 1994. Scale dependence and the species-area relationship. *Am. Nat.* 144, 717–740.
- Pielou, E.C., 1966. Species diversity and pattern diversity in the study of ecological succession. *J. Theor. Biol.* 10, 370–383.
- Pielou, E.C., 1975. *Ecological Diversity*. Wiley-Interscience, New York.
- Pielou, E.C., 1977. *Mathematical Ecology*. Wiley-Interscience, New York.
- Reid, P.C., Edwards, M., Hunt, H.G., Warner, A.J., 1998. Phytoplankton change in the North Atlantic. *Nature* 391, 546.
- Shannon, C.E., Weaver, W., 1962. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana.
- Simpson, E.H., 1949. Measurement of diversity. *Nature* 163, 688.
- Smith, W., Grassle, J.F., 1977. Sampling properties of a family of diversity measures. *Biometrics* 33, 283–292.
- Soetaert, K., Heip, C., 1990. Sample-size dependence of diversity indices and the determination of sufficient sample size in a high-diversity deep-sea environment. *Mar. Ecol. Prog. Ser.* 59, 305–307.
- Tong, Y.L., 1983. Some distribution properties of the sample species-diversity indices and their applications. *Biometrics* 39, 999–1008.
- Warner, A.J., Hays, G.C., 1994. Sampling by the Continuous Plankton Recorder survey. *Prog. Oceanogr.* 34, 237–256.
- Warwick, R.M., Clarke, K.R., 1991. A comparison of some methods for analysing changes in benthic community structure. *J. Mar. Biol. Assoc. UK* 71, 225–244.
- Williamson, M., 1998. Marine biodiversity in its global context. In: Ormond, R.F.G., Gage, J.D., Angel, M.V. (Eds.), *Marine Biodiversity: Patterns and Processes*. Cambridge University Press, Cambridge, pp. 1–17.