



Limits to the estimation of species richness: The use of relative abundance distributions

Werner Ulrich* and Marcin Ollik

Nicolaus Copernicus University in Toruń,
Department of Animal Ecology, Gagarina 9,
87–100 Toruń; Poland; E-mail:
ulrichw@uni.torun.pl

*Correspondence: Werner Ulrich, Nicolaus
Copernicus University in Toruń, Department of
Animal Ecology, Gagarina 9, 87–100 Toruń;
Poland. Tel.: 0048-56-611-4469;
Fax.: 0048-56-6114-443;
E-mail: ulrichw@uni.torun.pl; Internet:
www.uni.torun.pl/~ulrichw

ABSTRACT

The present study demonstrates the possibility of estimating species numbers of animal or plant communities from samples using relative abundance distributions. We use log-abundance–species-rank order plots and derive two new estimators that are based on log-series and lognormal distributions. At small to moderate sample sizes these estimators appear to be more precise than previous parametric and non-parametric estimators. We test our estimators using samples from 171 published medium-sized to large animal and plant communities taken from the literature. By this we show that our new estimators define also limits of precision.

Keywords

Relative abundance distribution, species-rank order distribution, species richness, lognormal distribution, log-series distribution, RAD, jackknife estimator.

INTRODUCTION

The estimation of the number of species from a series of samples is an important problem in community ecology and conservation. It is no wonder that many different estimation methods were developed that find wide application in community ecology, biodiversity assessment and conservation. Most estimation techniques are either based on parametric extrapolation models using species accumulation curves (Stout & Vandermeer, 1975; De Caprariis *et al.*, 1976; Lauga & Joachim, 1987; Soberon & Llorente, 1993; Edwards, 1997; Winklehner *et al.*, 1997; Keating, 1998; Ulrich, 1999; Turner *et al.*, 2003) or on nonparametric capture–recapture models (Burnham & Overton, 1978, 1979; Chao, 1984, 1987; Chao & Lee, 1992; Chao *et al.*, 1992; Bunge & Fitzpatrick, 1993; Lee & Chao, 1994; Brose *et al.*, 2003; Shen *et al.*, 2003).

These methods give best results when applied to small communities where most of the species can be sampled with moderate sampling effort (Burnham & Overton, 1978; Ulrich, 1999). A series of tests showed that they are strongly negatively biased at small to moderate sample sizes (Palmer, 1990, 1991; Colwell & Coddington, 1994; Walther & Morand, 1998; Hellmann & Fowler, 1999; Ulrich, 1999; Gotelli & Colwell, 2001; Longino *et al.*, 2002; Brose *et al.*, 2003; Brose & Martinez, 2004). Further, capture–recapture models like the popular jackknife estimators (Burnham & Overton, 1978, 1979) that use single or double occurrences in the samples are strictly defined only when the sample contains already more than half of the total species number. However, in biodiversity studies it is often necessary to estimate species numbers of regional or even continental faunae

and floras. In these cases we have to deal with hundreds or even several thousands of species and the necessary sample sizes for achieving the required coverage (the proportion of species present in the sample) would become extraordinarily large.

It seems therefore worthwhile to seek methods that allow at least a rough estimate of species richness at low sample sizes, where the sample coverage is far below 50% (Ulrich, 2001a). There have been several attempts to use extrapolations from species-number–log₂-abundance (octave) plots for estimating species richness (Pielou, 1977; Magurran, 1988; Miller & Wiegert, 1989; May, 1990; Curtis *et al.*, 2002). Under the assumption that these can be fitted by a truncated symmetrical lognormal distribution, it is possible to extrapolate beyond the famous veil line (Preston, 1962). However the loss of information due to the clustering of species into octaves resulted in very imprecise estimates with high standard errors. Further, it appeared that most distributions are not symmetrical in having more rare species (Nee *et al.*, 1991; Tokeshi, 1999). Such skewed distributions make any reasonable extrapolation beyond the veil line impossible. Here we use log-abundance–species-rank order plots (rank abundance distributions, RADs) (Fig. 1) and show that they provide estimates of species richness at small sample sizes that are more precise and less biased than previous methods.

Notation

- 1 S_S, N_S : Numbers of species and individuals obtained by sampling a certain area A_S ;
- 2 S_A, N_A : Numbers of species and individuals of a community in the area A_A ;

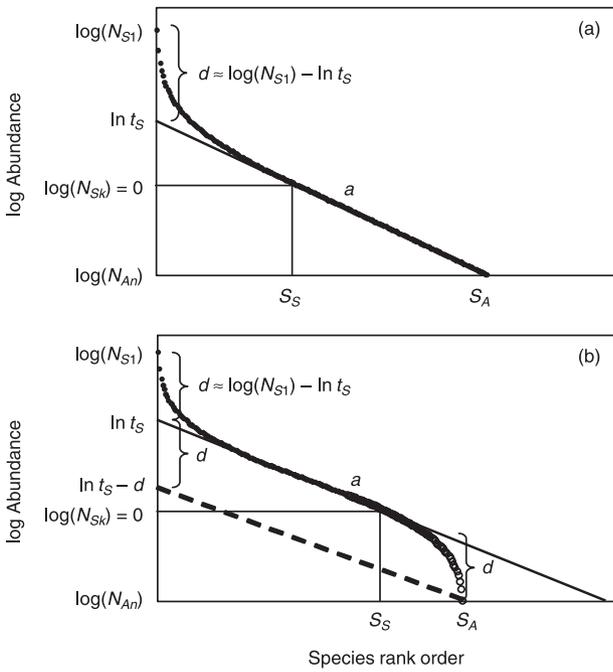


Figure 1 Inferring species numbers from log-abundance–species-rank order plots. a: log-series type distribution, b: lognormal type distribution. Abbreviations as in the text.

- 3 $N_{A1}, N_{Ai}, N_{S1}, N_{Sk}$: Numbers of individuals of the most (S_1) and the least (S_k) abundant species of the community and the sample;
- 4 $a, \ln t_S$: Slope and intercept of the linear regression in Fig. 1;
- S_E : Estimated species number regardless of model type.

RAD shapes and estimators

Many studies have shown that RADs have three basic shapes: Linear (log-series type RADs (Fig. 1a), S-shaped (lognormal type) (Fig. 1b), or linear in a double log plot (power function) (e.g. Sugihara, 1980; Magurran, 1988; Tokeshi, 1993, 1999; Ulrich, 2001b,c).

Assume first a RAD of a community in a certain area A_A follows at least in their lower part a log-series (Fig. 1a). The starting point is a sample of an area A_S containing S_S species and N_S individuals. From this sample we obtain four basic parameters of the RAD: N_{S1}, N_{Sk}, a , and $\ln t_S$ (Fig. 1a). To compute N_{A1} from N_{S1} we have to refer to the total area A_A . $N_{A1} = N_{S1} (A_A/A_S)$. Hence

$$\log(N_{A1}) = \log(N_{S1}) + \log(A_A) - \log(A_S) \tag{1}$$

In the same way we get $\ln t_A$ and N_{Ak} . We define A_A in such a way that $N_{An} = 1$. Hence

$$\begin{aligned} E_{\log\text{-series}} &\approx \frac{\ln t_A}{\text{slope}} = \frac{\ln t_S + (\log(A_A) - \log(A_S))}{\text{slope}} \\ &= \frac{\ln t_S + \log(N_{A1}) - \log(N_{S1})}{\text{slope}} \end{aligned} \tag{2}$$

If the RAD follows a power function the use of log transformed species numbers gives

$$E_{\text{Power function}} \approx e^{\frac{\ln t_S + (\log(N_{A1}) - \log(N_{S1}))}{\text{slope}}} \tag{3}$$

If the RAD follows an S-shaped lognormal type distribution an estimator of species numbers can be obtained from Fig. 1(b). Applying eq. 1 leads to

$$\begin{aligned} E_{\log\text{normal}} &\approx \frac{\ln t_A - d}{\text{slope}} = \frac{\ln t_S + (\log(N_{A1}) - \log(N_{S1})) - (N_{S1} - \ln t_S)}{\text{slope}} \\ &= \frac{2 \ln t_S + \log(N_{A1}) - 2 \log(N_{S1})}{\text{slope}} \end{aligned} \tag{4}$$

It should be noted that in this case a symmetrical lognormal shape is assumed. In fact unsymmetrical shapes with a higher number of rare species seem to be much more common (Tokeshi, 1999; Hubbell, 2001). However, the main aim of the present study is to define upper and lower boundaries of species richness. Eqs. 2 and 4 merely intend to define them. They tell that it is possible to estimate ranges of species numbers even from small samples if we can estimate (a) the total abundance (number of individuals) of the most abundant species of the focal community and (b) the slope and intercept of the log-abundance–species-rank order plot of the sample.

MATERIALS AND METHODS

To test whether eqs. 2–4 are able to predict species numbers we used random samples from real medium-sized to large animal and plant communities and compared the estimates of species richness with true species numbers. For this we searched the internet data base of relative abundance distributions (Ollik & Ulrich, 2003) that contains species abundance data obtained in quantitative studies about animal and plant communities from all sorts of organisms and habitats. From a total of 1345 data sets we extracted 171 data sets from 46 studies (Table 1) that met three basic criteria. First, the survey had to contain all species occurring in the given study area. Hence, samples that contained only a part of the species (the dominant ones) were left out. Second, the data set had to contain quantitative abundance data (not only abundance classes) of all species of the focal community. Third, communities had to contain more than 20 species. The largest community had 242 species (Condit *et al.*, 1996) (Table 1).

Abundances are frequently given as relative abundances. Because equations 2–4 operate with absolute values we adjusted all data by dividing through the abundance of the least abundant species. Hence, the least abundant species of each community was set to one individual. This procedure provided the abundance N_{A1} of the most abundant species in each community. By optical inspection we determined whether each RAD had significant upper ascending and lower descending curvatures. We classified communities with upper ascending and lower descending curvatures as being of the lognormal type. Communities without clear lower curvature were instead classified as following a log-series distribution. We used optical inspection instead of formal statistical tests because the latter tend to be conservative

Table 1 Data used in the present study to test the performance of species richness estimators. Data base codes refer to the respective entries in the internet data base of relative abundance distributions (Ollik & Ulrich, 2003)

Reference	Data base code	Species numbers	Adjusted community sizes	Continent	Taxon	Habitat
Anonymous (2002)	S 00443	81	3574–95,561	Europe	Pisces	Estuary
Arango (1990)	S 00363	39	630	South America	Chiroptera	Mountains
Biesiadka & Cichocka (1997)	S 00456	40	1304	Europe	Hydracarina	Lake
Bluzma (1995)	S 00253	32	0	Europe	Mammalia	Agricultural landscape
Chapman <i>et al.</i> (1997)	S 00225	88	2576	Africa	Magnoliophyta	Rain forest
Chapman <i>et al.</i> (1999)	S 00221	80	2609	Africa	Magnoliophyta	Rain forest
Cichocka (1996)	S 00469	80	10,338	Europe	Hydracarina	River
Clutton-Brock & Gillett (1979)	S 00309	50	2917	Africa	Magnoliophyta	Forest
Collado & Schmelz (2001)	S 00360	36	2233	Europe	Oligochaeta	Lake
Condit <i>et al.</i> (1996)	S 00260	228–242	20,613–62,480	North America	Magnoliophyta	Rain forest
Corbert (1964)	S 00245	22–26	2024–9836	Africa	Culicidae	Forest
Crichton (1964)	S 00246	42	24,508	Europe	Trichoptera	Agricultural landscape
Davis (1996)	S 00207	68–81	29,182–34,752	Africa	Scarabeidae	Savannah
Dewalt <i>et al.</i> (2000)	S 00218	75	1725	North America	Magnoliophyta	Rain forest
Dulčić <i>et al.</i> (1997)	S 00163	23–32	1183–2702	Europe	Pisces	Littoral
Falcon <i>et al.</i> (1996)	S 00162	75	150,000	Atlantic	Pisces	Littoral
Fawzi <i>et al.</i> (2002)	S 00341	31–75	761–890	Africa	Diatomidae	Stream
Felfili <i>et al.</i> (2000)	S 00250	37	1806	South America	Magnoliophyta	Woodland
Fox & Fox (1986)	S 00187	27–38	177–489	Australia	Telomophyta	Woodland
Gaston & Blackburn (2000)	S 00486	25–217	103–125,000,000	Europe	Aves	Large island
Henderson & Crampton (1997)	S 00283	23–25	77–85	South America	Pisces	River Halley Ponds
Horeau <i>et al.</i> (1998)	S 00216	39	683	South America	Pisces	River
Hydnes <i>et al.</i> (1999)	S 00155	38–55	1969–8220	Australia	Pisces	Shelf
Kartanas (1995)	S 00070	50	300	Europe	Aves	Forest
Klimczuk (1974)	S 00105	22–69	70–651	Europe	Nematoda	Waste land
Klitgaard-Kristensen and Buhl-Mortensen (1999)	S 00365	129	206,953	Europe	Foraminifera	Fjord
Lenzi-Grillini <i>et al.</i> (1996)	S 00345	32	1084	Africa	Magnoliophyta	Savannah
MacNally (1996)	S 00147	50	94,000	Australia	Aves	Forest
Morozov (1992)	S 00320	46	48,497	Europe	Aves	Forest
Mugohalu & Isichei (1991)	S 00301	34	1680	Africa	Magnoliophyta	Savannah
Nadkarni <i>et al.</i> (1995)	S 00258	46	2030	North America	Magnoliophyta	Rain forest
Pieczynski (1967)	S 00289	30	1456	Europe	Hydracarina	lake
Pinto & Punchihewa (1996)	S 00180	54	8153	Asia	Pisces	Mangroves
Poulsen (1996)	S 00273	92	6214	Asia	Magnoliophyta	Rain forest
Smith (1989)	S 00170	34–56	312–8153	Australia	Aves	Forest
Szczonowicz (1978)	S 00097	20–21	231–706	Europe	Nematoda	Field
Szkuclarek (1983)	S 00138	52	1014	Europe	Apoidea	Forest
Thiollay (1978)	S 00310	25–32	758–940	Africa	Falconiformes	Mixed landscape
Toriola <i>et al.</i> (1998)	S 00228	92	214	South America	Magnoliophyta	Forest
Valesini <i>et al.</i> (1997)	S 00177	42	43,339	Australia	Pisces	Estuary
van der Merwe <i>et al.</i> (1996)	S 00211	33	9360	Africa	Aranea	Forest
Williams <i>et al.</i> (1999)	S 00139	30	813	Australia	Magnoliophyta	Savannah
Wilson (2001)	S 00192	93–96	5631–14,802	North America	Pisces	Coral reef
Witkowski (1979)	S 00321	90–199	9905–14,291	Europe	Nematoda	Fields
Young <i>et al.</i> (1996)	S 00179	107	23,999	Australia	Pisces	Littoral
Zych (1977)	S 00096	49	2746	Europe	Nematoda	Field

while favouring even at low significance values the null hypothesis (a log-series distribution).

We sampled individuals at random from these communities until (if possible) 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%,

45%, 50%, 55% and 60% of S_A were found. The necessary sample sizes, N_S , varied between 0.00001% N_A and 77% N_A . Because estimates from low sample sizes are prone to Poisson errors we used only those samples for further analysis that contained more than

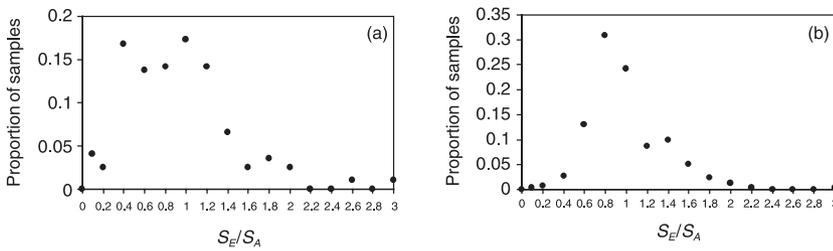


Figure 2 Proportion of data sets having certain S_E/S_A ratios in the case of log series (a) and lognormal type (b) data sets. In A, estimates were obtained using eq. 2; in B, using eq. 4.

20 individuals. Of course, real samples from medium-sized to large communities will nearly always be larger. However, because we were in particular interested in the behaviour of the above estimators at small sample sizes we used samples as small as only 20 individuals. The final data matrix used for analysis contained 469 independent samples from 171 data sets. For each sample we computed the slope, a , and the intercept, Int_s , of the linear regression through the lower 50%ile of species of the sample (Fig. 1).

For comparison we estimated species numbers using four estimators that have recently been critically tested and advocated as being the least biased at low sample sizes: The nonparametric second (J2) and fifth (J5) order jackknife estimators (Burnham & Overton, 1978; Brose *et al.*, 2003) and the parametric P5 (Turner *et al.*, 2003) and asymptotic linear estimators (AL) (Ulrich, 1999). The latter are based on species accumulation curves. AL is defined by

$$S_n = (\alpha n + \beta) \left(1 + \frac{\alpha n + \beta}{S_A} \right) \quad (5)$$

and P5 by

$$S_A = S_n^{(1-n^{-\alpha})} \quad (6)$$

n denotes the sample size and α and β are the parameters of the models. We fitted both parametric models with the nonlinear estimation module (Rosenbrocks pattern seeking) of Statistica 5 (Statsoft, 1997). Of course, many other estimators of species richness are advocated in the literature, for instance CHAO 1 and 2 (Chao & Lee, 1992), or rarefaction curves (Magurran, 2003). However, they appeared to be more strongly biased at low sample sizes than the estimators used here (Colwell & Coddington, 1994; Ulrich, 1999).

Following Brose *et al.* (2003) we calculated bias and precision of these estimators for single samples from

$$\text{bias} = \frac{S_E - S_A}{S_A} \quad (7)$$

and

$$\text{precision} = \left(1 - \frac{S_E}{S_A} \right)^2 \quad (8)$$

For each coverage class we computed mean bias and mean precision and compared class values by t -tests using the standard deviations of these measures.

RESULTS

Performance of the new estimators when the type of RAD is known

None of the 171 RADs followed a power function. Eq. 3 was therefore excluded as a possible estimator of species richness. A lognormal type RAD described 79 data sets best, a log-series 92. Hence, if the type of RAD is unknown the log-series estimate should mark an upper limit of species number. The estimate from the lognormal model should give the lower limit.

Eqs 2 and 4 were both able to estimate species numbers even from small samples if applied to *their* type of RAD (Fig. 2). $E_{\text{lognormal}}$ (eq. 4) had a mean S_E/S_A ratio of 0.90 ± 0.69 (mean \pm standard deviation) (Fig. 2a). Sixty-nine percent of all estimates ranged between $0.6S_A$ and $1.4S_A$, 95% between $0.2S_A$ and $2S_A$. The estimates were not negatively biased with respect to sample coverage (Fig. 3a). The estimator tended to overestimate species numbers at very low sample sizes (Fig. 3c).

$E_{\text{log-series}}$ had a mean S_E/S_A ratio of 1.10 ± 0.37 ; 87% of all estimates ranged between $0.6S_A$ and $1.4S_A$; and 98% of all estimates ranged between $0.2S_A$ and $2S_A$. The estimator was negatively biased at low coverage (Fig. 3b).

Performance without a priori knowledge of the type of RAD

Forty-six percent of all data sets had species numbers inside the range given by eqs 2 and 4 (Fig. 4). 63% of all data sets ranged between $0.9E_{\text{lognormal}}$ and $1.1E_{\text{log-series}}$. 96% of all data sets ranged between $0.67E_{\text{lognormal}}$ and $1.5E_{\text{log-series}}$. Hence, eqs 2 and 4 are indeed able to define upper and lower limits of species numbers.

We used the quotient $E_{\text{log-series}}/E_{\text{lognormal}}$ to infer how precisely we can estimate species numbers if neither the relative abundance distribution nor the sample coverage are known. Figure 5 shows that the ratio $E_{\text{log-series}}/E_{\text{lognormal}}$ was in 49.9% of all cases between 1 and 1.6. Sixty-seven percent of the ratios were between 1 and 2. Hence in 2/3 of the samples S_A was between $E_{\text{lognormal}}$ and two times $E_{\text{lognormal}}$. In 93.5% of all cases S_A was between 0.8 times $E_{\text{lognormal}}$ and six times $E_{\text{lognormal}}$. These ranges define the maximum precision any estimator of species numbers can achieve if the sample contains less than half of the total number of species.

In about 6.5% of samples the $E_{\text{log-series}}/E_{\text{lognormal}}$ ratio was above six. Ratios were independent of the type of RAD ($2 \times 2 \chi^2$: $P = 0.61$). A multiple regression with log-transformed $E_{\text{log-series}}/$

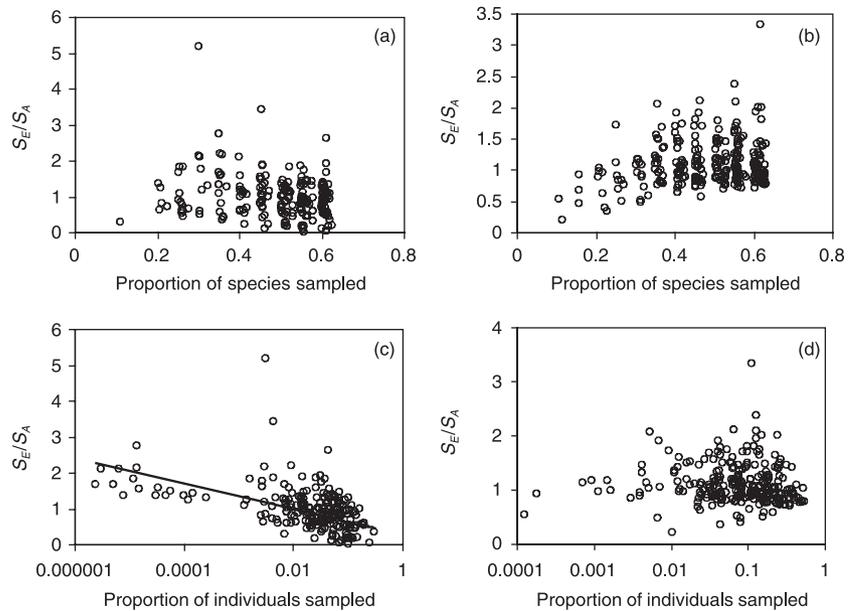


Figure 3 S_E/S_A ratios in relation to sample coverage (a and b) and sample sizes (c and d) and c: log-series type distributions (estimates obtained from eq. 2), b and d: lognormal type distributions (estimates obtained from eq. 4). In C S_E/S_A ratios were significantly correlated with sample size ($R^2 = 0.31$, $P(t) < 0.01$).

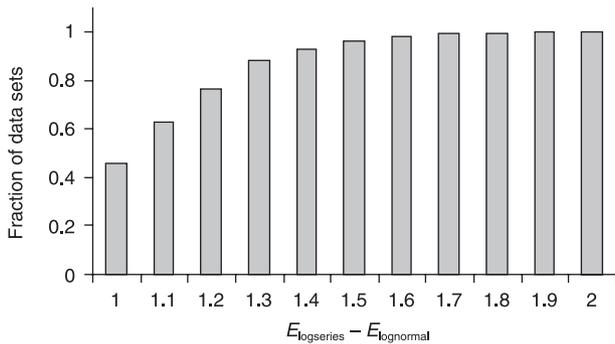


Figure 4 Proportions of data sets outside the range defined by the lognormal (lower limit) and log-series (upper limit) estimates. Hence 46% of the data set had species numbers inside this range, 63% of data sets ranged between $0.9E_{\lognormal}$ and $1.1E_{\log-series}$, and 96% between $0.67E_{\lognormal}$ and $1.5E_{\log-series}$. All data sets had species numbers above 0.5 times E_{\lognormal} and below two times $E_{\log-series}$.

E_{\lognormal} ratios as dependent and log-transformed sample size and coverage as independent variables instead showed that coverage and sample size correlated positively with the $E_{\log-series}/E_{\lognormal}$ ratios ($\beta_{\text{species coverage}} = 0.15$, $p(t) = 0.003$; $\beta_{\text{individual coverage}} = 0.17$,

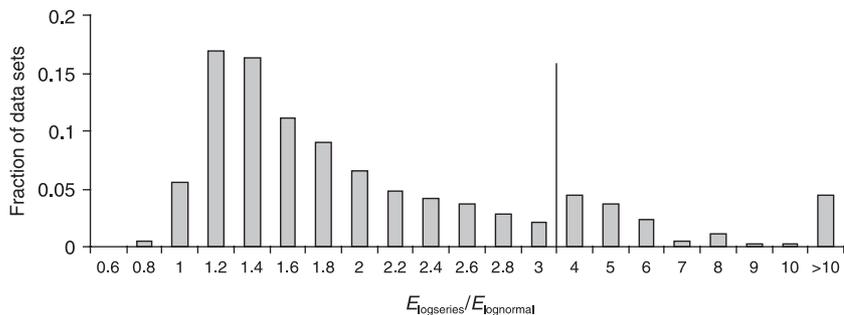
$p(t) = 0.001$; corrected multiple $R^2 = 0.68$, $P = 0.0001$). High $E_{\log-series}/E_{\lognormal}$ ratios above six occurred only when the sample contained already more than 35% of the total number of species to be estimated.

Comparison with existing estimators

All of the four standard estimators for species richness tested were negatively biased at small sample coverage (Fig. 6a). Their mean precision was also very similar (Fig. 6b). This contrasts with the much smaller bias and the higher precision of $E_{\log-series}$ and E_{\lognormal} when applied to 'their' RAD.

At a sample coverage of 30% J5 underestimated S_A in the mean by a factor of $f = 2.02 \pm 0.90$, J2 by a factor of 2.23 ± 0.26 , AL by a factor of 2.3 ± 0.57 , and P5 by a factor of 2.68 ± 0.46 . At 20% coverage these factors were 3.18 ± 2.21 , 3.22 ± 0.36 , 3.53 ± 0.40 , and 3.82 ± 0.86 , respectively. At 10% coverage J5 underestimated S_A by a factor of 5.82 ± 5.78 and J2 by a factor of 3.92 ± 0.66 , respectively. Hence if the sample coverage is unknown but above 10% the least biased estimator J5 cannot do better than predicting that at $\alpha = 95\%$ S_A will range between S_S and $S_S(f + 2\sigma) = S_S$ and $17.4S_S$. In this respect J2 performed best while giving a range of S_S and $5.24S_S$.

Figure 5 The quotient $E_{\log-series}/E_{\lognormal}$ is a measure of precision. In 65% of all cases the quotient was below two. Hence the true species number ranged between E_{\lognormal} and two times E_{\lognormal} . In 83% of all data sets this quotient was below three. However, in 6% of the data sets the method gave highly imprecise estimates with quotient values above 10. The vertical line indicates the change of scale of the X-axis.



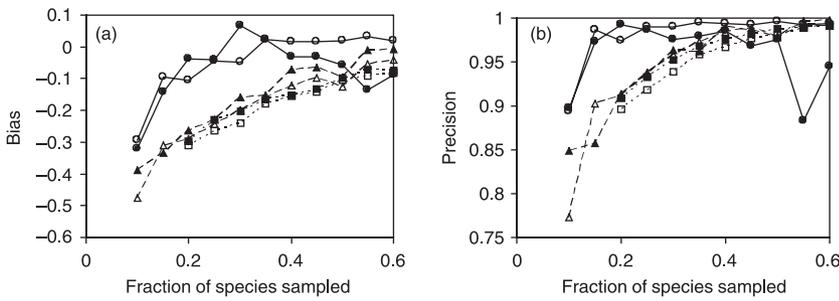


Figure 6 The four standard estimators tested had very similar negative bias (a) and precision (b) at low sample coverage. The estimators developed here for log-series (open circles) and for lognormal distributions (full circles) were much less biased. J5 (full triangles), J2 (open triangles), AL (full squares), P5 (open squares).

DISCUSSION

The present study shows that it is possible to estimate species richness from the study of RADs. A different graphical presentation of lognormal distributions (species number–octave plots) has already been used before to estimate species numbers (e.g. Miller & Wiegert, 1989; May, 1990; Curtis *et al.*, 2002; Longino *et al.*, 2002). However, this method proved to be highly imprecise if the mode of this distribution is unknown. Hence, the method needs a coverage of more than 50%, a magnitude at which nonparametric methods give much better results (Colwell & Coddington, 1994; Ulrich, 1999). The log-series distribution has only seldom been applied for estimating species numbers. This stems surely from the fact that the common way to present the log-series is in terms of sample parameters (Pielou, 1977; Magurran, 1988; Tokeshi, 1999). This does not allow S_A to be estimated (Pielou, 1977). However, when introducing a lower limit for the least abundant species (the inverse of the area required to find this species) it is of course easy to estimate the number of species from the model parameters (Coleman, 1981; Green & Ostling, 2003).

The main goal of our study was to define upper and lower limits of species richness even from small samples. Hence, our approach differs somewhat from what other estimators intend to do, to provide an estimate of species richness. Our estimators instead define limits of precision for any estimate of species richness. Both approaches should therefore be applied together to give an impression about the real species richness.

Naturally, the range defined by E_{\lognormal} and $E_{\log-series}$ is quite large. But it gives an impression about the magnitude of species richness. Estimates of such ranges are particularly welcome in studies of tropical diversity where the sample coverage is frequently well below 20%. Current parametric and nonparametric estimators of species richness are all strongly biased at low species representivity or sample size and give only lower limits of species numbers. Of the four estimators tested here only J2 was able to give similar ranges of species numbers.

However, there is strong evidence that regional large animal or plant communities follow in the majority of cases lognormal type distributions with a lower downward curvature (Sugihara, 1980; Tokeshi, 1993, 1999; Gregory, 1994; Gaston & Blackburn, 2000; Sugihara *et al.*, 2003). It seems therefore justified to use E_{\lognormal} for getting an impression about the probable species numbers. $E_{\log-series}$ would then provide an estimate of the upper boundary of species richness. On the other hand, Magurran &

Henderson (2003) showed that local communities with high rates of species turnover will be straight at least in the lower part of the RAD. Figures 2(b) and 3 showed that in these cases $E_{\log-series}$ will give precise estimates even at small sample sizes. Hence estimates of species numbers might be improved from knowledge about dispersion patterns.

However, Fig. 5 showed that in about 6% of all cases our method failed while resulting in highly imprecise estimates. Such failures were independent of the type of RAD but occurred when the sample contained already more than 35% of the total number of species to be estimated. Hence, sample coverage influenced the precision of the estimators but in a different way than for previous estimators. Our estimators did best at lower coverage. They should therefore rather be seen as a complementation than as an alternative.

Figure 6 showed that bias and precision of the conventional parametric and nonparametric estimators were very similar. The least biased estimator J5 performed only slightly better than the others. Recently, Brose *et al.* (2003) advocated J5 in the case of small sample sizes as being less biased than J2. However, in their model simulations with predefined relative abundance distributions J5 did only slightly better than J2. It might be that differences between estimators revealed in model situations vanish when applying them to real world communities as in the present study. Brose *et al.* (2003) also showed that community evenness influences bias and precision of most estimators. They tend to be less biased but more imprecise at higher evenness. Although the present study does not address this problem directly a preliminary analysis (Ulrich, in prep.) showed a similar behaviour of E_{\lognormal} and $E_{\log-series}$. As found by Brose *et al.* (2003), bias and precision of E_{\lognormal} and $E_{\log-series}$ are negatively correlated (Ulrich, in prep.). Further studies have to show whether it is possible to correct both estimators for their dependence on evenness as had been done by Ulrich (1999) for estimators based on species accumulation curves.

There is one possible pitfall in our method. It assumes that the type of distribution is scale independent. While this assumption seems justified when assessing species richness of local patches recent theoretical and observational studies (Tokeshi, 1999; Plotkin *et al.*, 2000; Hubbell, 2001; Longino *et al.*, 2002) pointed to changes in abundance distributions at regional scales. However, our method merely intends to set upper and lower limits of species richness. Hence a switch from a lognormal to a log-series distribution at regional scales (Hubbell, 2001) would not influence the estimate as long as the slope a in Fig. 1 remains

unchanged. We are not aware of any studies reporting parameter values of RADs at different scales. Such studies would be very interesting and would surely serve to improve available estimators of species richness.

ACKNOWLEDGEMENTS

We thank M. A. McGeoch and three anonymous referees for critical and valuable suggestions on the manuscript. Miss Hazel Pearson kindly improved our English. This work was supported by a grant from the Polish Science Committee (KBN, 3 F04F 034 22) to WU.

REFERENCES

- Anonymous (2002) Hinkley Point/Severn Estuary fish and crustacean monitoring — annual reports. <http://www.irhouse.demon.co.uk/HinkleyTo2002.xls>.
- Arango, J.M. (1990) Diversidad y habitats alimenticios de murciélagos en transectos altitudinales a través de la Cordillera Central de los Andes en Colombia. *Studies of Neotropical Fauna and Environment*, **25**, 11–17.
- Biesiadka, E. & Cichocka, M. (1997) Hydracarina fauna in Lobelia-type lakes near Bytów. *Fragmenta Faunistica*, **40**, 81–93.
- Bluzma, P. (1995) Abundance, biomass and biotopes structure of mammals in the landscape of hilly-morainic uplands of east Lithuania. *Ekologija*, **2**, 54–59.
- Brose, U. & Martinez, N.D. (2004) Estimating of richness of species with variable mobility. *Oikos*, **105**, 292–300.
- Brose, U., Martinez, N.D. & Williams, R.J. (2003) Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology*, **84**, 2364–2377.
- Bunge, J. & Fitzpatrick, M. (1993) Estimating the number of species: a review. *Journal of the American Statistical Association*, **88**, 364–373.
- Burnham, K.P. & Overton, W.S. (1978) Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika*, **65**, 623–633.
- Burnham, K.P. & Overton, W.S. (1979) Robust estimation of population size when capture probabilities vary among animals. *Ecology*, **60**, 927–936.
- Chao, A. (1984) Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, **11**, 265–270.
- Chao, L. (1987) Estimating the population size for capture–recapture data with unequal catchability. *Biometrics*, **43**, 783–791.
- Chao, A. & Lee, S.M. (1992) Estimating the number of classes via sample coverage. *Journal of the American Statistical Association*, **87**, 210–217.
- Chao, A., Lee, S.M. & Jeng, S.L. (1992) Estimation of population size for capture–recapture data when capture probabilities vary by time and individual animal. *Biometrics*, **48**, 201–216.
- Chapman, C.A., Chapman, L.J., Wrangham, R., Isabirye-Basuta, G. & Ben-David, K. (1997) Spatial and temporal variability in the structure of tropical forest. *African Journal of Ecology*, **35**, 287–302.
- Chapman, C.A., Wrangham, R.W., Chapman, L.J., Kennard, D.K. & Zanne, A.E. (1999) Fruit and flower phenology at two sites in Kibale National Park, Uganda. *Journal of Tropical Ecology*, **15**, 189–211.
- Cichocka, M. (1996) Wodopójki (Hydracarina) rzeki Pasłęki. *Fragmenta Faunistica*, **36**, 179–205.
- Clutton-Brock, T.H. & Gillett, J.B. (1979) A survey of forest composition in the Gombe National Park, Tanzania. *African Journal of Ecology*, **17**, 131–158.
- Coleman, B.D. (1981) On random placement and species–area relations. *Mathematical Bioscience*, **54**, 191–215.
- Collado, R. & Schmelz, R.M. (2001) Oligochaeta. Distribution patterns in two German hardwater lakes of different trophic State. *Limnologia*, **31**, 317–328.
- Colwell, R.K. & Coddington, J.A. (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions Royal Society London B*, **345**, 101–118.
- Condit, R., Hubbell, S.P. & Foster, R.B. (1996) Changes in tree species abundance in a Neotropical forest: impact of climate change. *Journal of Tropical Ecology*, **12**, 231–256.
- Corbert, P.S. (1964) Nocturnal flight activity of sylvan Culicidae and Tabanidae (Diptera) as indicated by light-traps: a further study. *Proceedings Royal Entomological Society London A*, **36**, 53–57.
- Crichton, M.I. (1964) Observations on captures of Trichoptera in suction and light traps near Reading, Berkshire. *Proceedings Royal Society London A*, **39**, 53–67.
- Curtis, T.P., Sloan, W.T. & Scannel, J.W. (2002) Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences*, **10.1073**, 1–6.
- Davis, A.L.V. (1996) Community organization of dung beetles (Coleoptera: Scarabeidae): differences in body size and functional group structure between habitats. *African Journal of Ecology*, **34**, 258–275.
- De Caprariis, P., Lindemann, R.H. & Collins, C. (1976) A method for determining optimum sample size in species diversity studies. *Mathematical Geology*, **8**, 575–581.
- Dewalt, S.J., Schnitzer, S.A. & Denslow, J.S. (2000) Density and diversity of lianas along a chronosequence in a central Panamanian lowland forest. *Journal of Tropical Ecology*, **16**, 1–19.
- Dulčić, J., Kraljević, M., Grbec, B. & Pallaoro, A. (1997) Composition and temporal fluctuations of inshore juvenile fish populations in the Kornati Archipelago, eastern middle Adriatic. *Marine Biology*, **129**, 267–277.
- Edwards, L.E. (1997) A useful procedure for estimating the species richness of spiders. *Journal of Arachnology*, **25**, 99–105.
- Falcon, J.M., Bortone, S.A., Brito, A. & Bundrick, C.M. (1996) Structure and relationships within and between the littoral, rock substrate fish communities off four islands in the Canarian Archipelago. *Marine Biology*, **125**, 215–231.
- Fawzi, I., Loudiki, M., Oubraim, S., Sabour, I. & Chlaida, M. (2002) Impact of wastewater effluent on the Diatom Assemblages Structure of a Brackish Small Stream: Oued Hassar (Morocco). *Limnologia*, **32**, 54–65.

- Felfili, J.M., Rezende Alba, V., da Silva Jr, M.C. & Silva, M.A. (2000) Changes in the floristic composition of cerrado sensu stricto in Brazil over a nine-year period. *Journal of Tropical Ecology*, **16**, 579–590.
- Fox, M.D. & Fox, B.J. (1986) The effect of fire frequency on the structure and floristic composition of a woodland understorey. *Australian Journal of Ecology*, **11**, 77–85.
- Gaston, K.J. & Blackburn, T.M. (2000) *Pattern and process in macroecology*. Blackwell Science Ltd, Oxford.
- Gotelli, N. & Colwell, R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.
- Green, J.L. & Ostling, A. (2003) Endemics–area relationships: the influence of species dominance and spatial aggregation. *Ecology*, **84**, 3090–3097.
- Gregory, R. (1994) Species abundance patterns of British birds. *Proceedings Royal Society London B*, **257**, 299–301.
- Hellmann, J.J. & Fowler, G.W. (1999) Bias, precision and accuracy of four measures of species richness. *Ecological Applications*, **9**, 824–834.
- Henderson, P.A. & Crampton, W.G.R. (1997) A comparison of fish diversity and abundance between nutrient-rich and nutrient-poor lakes in Upper Amazonian. *Journal of Tropical Ecology*, **13**, 175–198.
- Horeau, V., Cerdan, P., Champeau, A. & Richard, S. (1998) Importance of aquatic invertebrates in the diet of rapid-dwelling fish in the Sinnamary River, French Guiana. *Journal of Tropical Ecology*, **14**, 851–864.
- Hubbell, S.P. (2001) *The unified theory of biogeography and biodiversity*. Princeton University Press, Princeton.
- Hydnes, G.A., Platell, M.E., Potter, I.C. & Lenanton, R.C.J. (1999) Does the composition of the demersal fish assemblages in temperate coastal waters change with depth and undergo consistent seasonal changes? *Marine Biology*, **143**, 335–352.
- Kartanas, E. (1995) Zespół ptaków łęgowych rezerwatu ‘Las Piwnicki’ w latach (1985–87). *AUNC Biologia*, **49** (79), 100.
- Keating, K.A. (1998) Estimating species richness: the Michaelis-Menten model revisited. *Oikos*, **81**, 411–416.
- Klimczuk, G. (1974) *Nicienie występujące w glebie zraszanej solanką w pobliżu tężni w ciechocinku*. Unpublished diploma work, Department of Animal Ecology UMK, Toruń, 1974.
- Klitgaard-Kristensen, D. & Buhl-Mortensen, L. (1999) Benthic foraminifera along offshore gradient: a comparison with amphipods and molluscs. *Journal of Natural History*, **33**, 317–350.
- Lauga, J. & Joachim, J. (1987) L'échantillonnage des populations d'oiseaux par la méthode des E.F.P. intérêt d'une étude mathématique de la courbe de richesse cumulée. *Acta Oecologica*, **8**, 117–124.
- Lee, S.M. & Chao, A. (1994) Estimating population size via sample coverage for closed capture-recapture models. *Biometrics*, **50**, 88–97.
- Lenzi-Grillini, C.R., Viskanac, P. & Mapesa, M. (1996) Effects of 20 years of grazing exclusion in area of the Queen Elizabeth national park, Uganda. *African Journal of Ecology*, **34**, 333–341.
- Longino, J.T., Coddington, J. & Colwell, R.K. (2002) The ant fauna of a tropical rain forest: estimating species richness three different ways. *Ecology*, **83**, 689–702.
- MacNally, R. (1996) A winter's tale: among-year variation in bird community structure in a south-eastern Australian forest. *Australian Journal of Ecology*, **21**, 280–291.
- Magurran, A.E. (1988) *Ecological diversity and its measurement*. Princeton University Press, Princeton.
- Magurran, A.E. (2003) *Measuring biological diversity*. Blackwell Publishers, Malden.
- Magurran, A.E. & Henderson, P.A. (2003) Explaining the excess of rare species in natural species abundance distributions. *Nature*, **422**, 714–716.
- May, R.M. (1990) How many species? *Philosophical Transactions of the Royal Society London B*, **330**, 293–304.
- van der Merwe, M., Dippenaar-Schoeman, A.S. & Scholtz, C.H. (1996) Diversity of ground-living spiders at Ngome State Forest, Kwazulu-Natal: a comparative survey in indigenous forest and pine plantations. *African Journal of Ecology*, **34**, 342–350.
- Miller, R.I. & Wiegert, R.G. (1989) Documenting completeness, species–area relations, and the species abundance distribution of a regional flora. *Ecology*, **70**, 16–22.
- Morozov, N.S. (1992) Breeding forest birds in the Valdi Uplands, north-west Russia: assemblage composition, interspecific associations and habitat amplitudes. *Annales Zoologici Fennici*, **29**, 7–27.
- Mugohalu, J.I. & Isichei, A.O. (1991) Nigerian Man and Biosphere research plots in Guinea savanna: floristics and structure of the vegetation. *African Journal of Ecology*, **29**, 229–240.
- Nadkarni, N.M., Matelson, T.J. & Haber, W.A. (1995) Structural characteristics and floristic composition of a Neotropical cloud forest, Monteverde, Costa Rica. *Journal of Tropical Ecology*, **11**, 481–495.
- Nee, S., Harvey, P.H. & May, R.M. (1991) Lifting the veil on abundance patterns. *Proceedings of the Royal Society London B*, **243**, 161–163.
- Ollik, M. & Ulrich, W. (2003) The internet database of relative abundance distributions. <http://www.uni.torun.pl/~ulrichw/RADDatabase/RADDatabase.htm>.
- Palmer, M.W. (1990) The estimation of species richness by extrapolation. *Ecology*, **71**, 1195–1198.
- Palmer, M.W. (1991) The estimation of species richness: the second-order jackknife reconsidered. *Ecology*, **72**, 1512–1513.
- Pieczyński, E. (1967) The occurrence of water mites (Hydracarina) and some other invertebrates in the littoral and central part of lake Śniardwy. *Ecologia Polska*, **15**, no. 26.
- Pielou, E.C. (1977) *Mathematical ecology*. Wiley, New York.
- Pinto, L. & Punchihewa, N.N. (1996) Utilization of mangroves and seagrasses by fishes in the Negombo Estuary Sri Lanka. *Marine Biology*, **126**, 333–345.
- Plotkin, J.B., Potts, M.D., Yu, D.W., Bunyavejchewin, S., Condit, R., Foster, R., Hubbell, S., LaFranckie, J., Manokaran, N., Seng, L.H., Sukarmar, R., Nowak, M.A. & Ashton, P.S. (2000) Predicting species diversity in tropical forests. *Proceedings of the National Academy of Sciences USA*, **97**, 10850–10854.

- Poulsen, A.D. (1996) Species richness and density of ground herbs within a plot of lowland rainforest in north-west Borneo. *Journal of Tropical Ecology*, **12**, 177–190.
- Preston, F.W. (1962) The canonical distribution of commonness and rarity. Part I and II. *Ecology*, **43** (185–215), 410–432.
- Shen, T.J., Chao, A. & Lin, C.F. (2003) Predicting the number of new species in further taxonomic sampling. *Ecology*, **84**, 798–804.
- Smith, P. (1989) Changes in a forest bird community during a period of fire and drought near Bega, New South Wales. *Australian Journal of Ecology*, **14**, 41–54.
- Soberon, M.J. & Llorente, B.J. (1993) The use of species accumulation functions for the prediction of species richness. *Conservation Biology*, **7**, 480–488.
- Statsoft Inc. (1997) *Statistica for windows* (Computer program manuals). OK, Tulsa.
- Stout, J. & Vandermeer, J. (1975) Comparison of species richness for stream-inhabiting insects in tropical and midlatitude streams. *American Naturalist*, **109**, 263–280.
- Sugihara, G. (1980) Minimal community structure: an explanation of species abundance patterns. *American Naturalist*, **116**, 770–787.
- Sugihara, G., Bersier, L.-F., Southwood, T.R.E., Pimm, S.L. & May, R.L. (2003) Predicted correspondence between species abundances and dendrograms of niche similarities. *Proceedings of the National Academy of Science USA*, **100**, 5246–5251.
- Szczonowicz, M. (1978) *Fauna i ekologia nicieni w uprawie mieszanki jęczmieni — owies*. Unpublished diploma work, Department of Animal Ecology UMK, Toruń 1978.
- Szkudlarek, H. (1983) *Struktura zgrupowań dzikich pszczołowatych (Hymenoptera, Apoidea) rezerwatu leśnego 'Las Piwnicki' i otuliny*. Unpublished diploma work, Department of Animal Ecology UMK, Toruń 1983.
- Thiollay, J.M. (1978) Population structure and seasonal fluctuations of the Falconiformes in Uganda National Parks. *East African Wildlife Journal*, **16**, 145–151.
- Tokeshi, M. (1993) Species abundance patterns and community structure. *Advances in Ecological Research*, **24**, 111–186.
- Tokeshi, M. (1999) *Species coexistence*. Blackwell, Oxford.
- Toriola, D., Chareyre, P. & Buttler, A. (1998) Distribution of primary forest plant species in a 19-year old secondary forest in French Guiana. *Journal of Tropical Ecology*, **14**, 323–340.
- Turner, W., Leitner, W. & Rosenzweig, M. (2003) WS2M. User's manual. <http://eebweb.arizona.edu/diversity/>.
- Ulrich, W. (1999) Estimating species numbers by extrapolation I: Comparing the performance of various estimators using large model assemblages. *Polish Journal of Ecology*, **47**, 271–291.
- Ulrich, W. (2001a) Estimating species numbers by extrapolation: a cautionary note. *Polish Journal of Ecology*, **49**, 299–305.
- Ulrich, W. (2001b) Models of relative abundance distributions I: model fitting by stochastic models. *Polish Journal of Ecology*, **49**, 145–157.
- Ulrich, W. (2001c) Relative abundance distributions of species: The need to have a new look at them. *Polish Journal of Ecology*, **49**, 393–407.
- Valesini, F.J., Potter, I.C., Platell, M.E. & Hydnes, G.A. (1997) Ichthyofaunas of temperate estuary and adjacent marine embayment. Implications regarding choice of nursery area and influence of environmental changes. *Marine Biology*, **128**, 317–328.
- Walther, B.A. & Morand, S. (1998) Comparative performance of species richness estimation methods. *Parasitology*, **116**, 395–405.
- Williams, R.J., Cook, G.D., Gill, A.M. & Moore, P.H.R. (1999) Fire regime, fire intensity and tree survival in a tropical savanna in northern Australia. *Australian Journal of Ecology*, **24**, 50–59.
- Wilson, D.T. (2001) Patterns of replenishment of coral-reef fishes in the nearshore waters of the San Blas Archipelago, Caribbean Panama. *Marine Biology*, **139**, 735–753.
- Winklehner, R., Winkler, H. & Kampichler, C. (1997) Estimating local species richness of epigeic Collembola in temperate dry grassland. *Pedobiologia*, **41**, 154–158.
- Witkowski, T. (1979) Reakcja nicieni glebowych na zaprawę nasienną GT. *Studia Societatis Scientiarum Toruniensis*, **10**, 44–79.
- Young, J.W., Lamb, T.D. & Bradford, R.W. (1996) Distribution and community structure of midwater fish in relation to the subtropical convergence off eastern Tasmania, Australia. *Marine Biology*, **126**, 571–584.
- Zych, A. (1977) *Występowanie nicieni żyjących w glebie pod uprawą Jęczmienia*. Unpublished diplom work, Department of Animal Ecology UMK, Toruń 1977.