

A meta-analysis of species–abundance distributions

Werner Ulrich, Marcin Ollik and Karl Inne Ugland

W. Ulrich (*ulrichw@umk.pl*) and M. Ollik, Dept of Animal Ecology, Nicolaus Copernicus Univ. in Toruń, Gagarina 9, PL–87-100 Toruń, Poland. – K. I. Ugland, Dept of Biology, Univ. of Oslo, PO 1064, Blindern, NO-0316 Oslo, Norway.

The species–abundance distribution (SAD) describes the abundances of all species within a community. Many different models have been proposed to describe observed SADs. Best known are the logseries, the lognormal, and a variety of niche division models. They are most often visualized using either species richness–log abundance class (Preston) plots or abundance–species rank order (Whittaker) plots. Because many of the models predict very similar shapes, model distinction and testing become problematic. However, the variety of models can be classified into three basic types: one that predicts a double S-shape in Whittaker plots and a unimodal distribution in Preston plots (the lognormal type), a second that lacks the mode in Preston plots (the logseries type), and a third that predicts power functions in both plotting types (the power law type). Despite the interest of ecologists in SADs no formal meta-analysis of models and plotting types has been undertaken so far. Here we use a compilation of 558 species–abundance distributions from 306 published papers to infer the frequency of the three SAD shapes in dependence of environmental variables and type of plotting. Our results highlight the importance of distinguishing between fully censused and incompletely sampled communities in the study of SADs. We show that completely censused terrestrial or freshwater animal communities tend to follow lognormal type SADs more often than logseries or power law types irrespective of species richness, spatial scale, and geographic position. However, marine communities tend to follow the logseries type, while plant communities tend to follow the power law. In incomplete sets the power law fitted best in Whittaker plots, and the logseries in Preston plots. Finally our study favors the use of Whittaker over Preston plots.

The species–abundance distribution (SAD) describes the abundances of all species within a community. Since its formal introduction by Raunkiaer (1909) ecologists have used SADs as a tool to infer dominance orders, diversity and evenness (Magurran 2004, McGill et al. 2007). SADs are mostly plotted at semilog scales. Preston (1948) introduced frequency \log_2 abundance (octaves) plots similar to plots of probability distributions (hereafter Preston plots, Fig. 1). This method involves a binning of abundances and therefore a loss of information. There is a considerable debate about the bias introduced from this binning (reviewed by Nekola et al. 2008). Plots of log-abundance versus species rank order were introduced by Whittaker (1965) (hereafter Whittaker plots) in order to avoid binning and the associated loss of information. While this method seems intuitively superior most studies still rely on Preston plots because it makes analytical treatment easier and fits do not rely on species richness (McGill et al. 2007). Standard fitting routines and common test statistics are based on Preston plots.

The interest of ecologists in SADs is mirrored by the large number of models designed to mimic observed abundance distributions. The recent review of McGill et al. (2007) lists 27 different species–abundance models. They range from purely statistical sampling models like the log-series (Fisher et al. 1943), the lognormal (Preston 1948), or maximum

entropy (Harte et al. 2008) to elaborate mechanistic models of niche division (Tokeshi 1999), ecological drift (Hubbell 2001), or spatial distribution (Harte et al. 1999). The majority of them predict very similar shapes that prohibit any clear model identification (Ollik 2008). However, all of these models can be subsumed under three basic SAD shapes in Whittaker and Preston plots (Fig. 1) (May 1975, Hughes 1986, Magurran 2004, McGill et al. 2007, Ollik 2008), a lognormal, a logseries, and a power law shape.

In Whittaker plots the lognormal shape is characterized by an upper positive and a lower negative curvature made by the few very dominant and a few very rare species (Fig. 1A). In Preston plots the lognormal type has a single mode (Fig. 1B). It is typical for communities with lowered frequencies of very abundant and very rare species. Since Preston (1948) and May (1975) lognormal shapes have mainly been associated with communities stabilized by the influence of many ecological factors that together generate a lognormal abundance distribution according to the central limit theorem of statistics (Hill and Hamer 2003, Magurran 2004, Sizing et al. 2009). The resource partitioning models of Sugihara (1980), Tokeshi (1996) and MacArthur (1957) also predict lognormal shapes.

The logseries is the Poisson sampling of a gamma distribution and appears roughly as a skewed J-curve in Whittaker

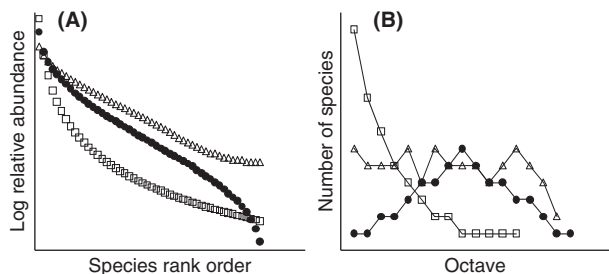


Figure 1. Basic shapes of species–abundance distributions in Whittaker (A) and Preston plots (B). Black dots: lognormal shape; open triangles: logseries; open squares: power law.

plots without a marked lower curvature and at least for medium abundance classes as an equal distribution in Preston plots (Fig. 1). Since Patrick (1963), Whittaker (1965) and Gray (1979) the logseries has been associated with disturbed, unstable, or early successional habitats. Neutrality predicts logseries at the metacommunity level (Hubbell 2001).

Lastly, power law SADs have been generated from fractal (Moulliot et al. 2000, Pueyo 2006) or Galton-Watson branching processes (Nee 2003). They have a constant positive curvature in Whittaker and in Preston plots (Fig. 1) and lack any mode in Preston plots. Chu and Adami (1999) applied power functions to fossil data and particularly argued for power laws at taxonomic levels above species.

Despite the proliferation of SAD models and the discussion around the shape of SADs, few studies were concerned with the frequency of these shapes in nature. Hughes (1986) compiled 222 communities and fitted lognormal, logseries and his dynamics model by eye, using the appearance of an upper and lower curvature in Whittaker plots and the existence of a single mode in Preston plots. Dewdney (2000) compared the fits of the logseries and his predation based model for 100 communities taken from the literature. Both authors reported a mix of shapes with a predominance of the logseries type. However, they did not use rigorous model testing and did not account for different numbers of free parameters in their models.

Many authors discuss the point that observed local SADs are samples from a regional metacommunity SAD and the respective sample distributions are now well understood (reviewed by May 1975, Etienne and Alonso 2005, Green and Plotkin 2007). Preston's famous 'veil line' (1948) and the associated problem of unseen (undersampled) species (reviewed by Magurran 2004, 2007) highlighted the need to use fairly completely censused communities for model testing. Indeed, any SAD model contains the species richness S as a free parameter. However, most comparative studies on SADs (Hughes 1986, Tokeshi 1996, Dewdney 2000, Ugland et al. 2007, Harte et al. 2008) treated the communities under study as being fully censused and compared model fits irrespective of the completeness of the data set. In our view, such a procedure may often be misleading because traditionally a SAD is defined for all species of a community. Therefore, we here intend to undertake a metaanalysis about SADs of fully censused communities. Rather than comparing the fits of different models we focus on the three mentioned basic shapes of SADs and relate them to basic taxonomic,

biogeographic and ecological variables. We will show that fully censused communities are predominantly fitted by models that predict lognormal shapes but that common ecological and biogeographic classifications largely fail to predict a certain shape. Obviously we need a deeper understanding of how abundance differences emerge.

Material and methods

We compiled 558 species–abundance distributions from 306 published papers, of which 504 were based on abundance (animals), or coverage (most plant data) and 54 on data on biomass. Species richness ranged between 10 and 460. The complete reference is given in Ulrich and Ollik (2003) and Ollik (2008). We used only data based on quantitative counting of individuals within a certain area. Coverage was only used when it was possible to recalculate to the individual level. This limitation is the reason why we considered only 53 plant data sets. Plant abundances are most often given as grouped coverage data within phytosociological classification schemes. These data do not allow for a direct comparison with individual based data.

Any fitting of abundance models is extremely sensitive to the completeness of the species list (Hughes 1986, Miller and Wiegert 1989). Therefore, we classified the abundance data into fully censused and incomplete sets. In order to do so we used a two step approach and first estimated the total number of species from the number of singletons in the data sets using the first order jackknife estimator (Burnham and Overton 1978). We classified all data sets with more than five singletons and for which the estimator indicated a coverage of less than 95% of the total fauna or flora as being incomplete. We further screened the remaining data sets for whether they had been obtained from spatially identifiable habitats like lakes, whole forests, meadows, or from a whole ecoregion and/or whether they can be judged as being fully censused from the knowledge of the species richness of the associated metacommunity (many vertebrate and plant data sets). All data that were possibly samples from a larger region were also classified as being incomplete. This procedure identified 119 data sets that were sufficiently censused to fit abundance models (termed complete sets below). 385 data sets were classified as being too incomplete for further detailed analysis and were only used for a general comparison of model fit in the case of incomplete censuring.

To relate distributions to environmental, ecological, and taxonomic variables we classified the 119 complete data sets into five major categories using the available information of the respective publications: a) continent (all continents except Asia for which only two complete data sets were available); b) biome (marine, freshwater, terrestrial); c) latitude (temperate, subtropical, tropical); d) taxonomy (vertebrates, invertebrates, plants); e) scale (metacommunity, smaller scale).

Our meta-analysis is based on the three basic shapes defined in the introduction and we fitted three models that generate these shapes (the lognormal, the logseries and the power law abundance distributions) to Whittaker and Preston plots of the respective empirical distributions. The present paper deals therefore basically with SAD shapes and

Table 1. Summary statistics of 558 species–abundance distributions each fitted by a lognormal, a logseries and a power law model. Given are numbers of ranks (best = 1 to worst = 3) of these fits for complete and incomplete data sets and for data sets based on estimates of biomass. Highest numbers of best and worst fits are given in bold type.

Model	n	Whittaker			Preston		
		Rank			Rank		
		1	2	3	1	2	3
Complete							
Lognormal	119	53	40	26	60	37	22
Logseries	119	30	37	52	45	61	13
Power law	119	36	42	41	14	21	84
Incomplete							
Lognormal	385	103	161	121	120	133	132
Logseries	385	68	127	190	143	189	53
Power law	385	214	97	74	122	63	200
Biomass							
Lognormal	54	42	11	1	38	13	3
Logseries	54	5	7	42	14	39	1
Power law	54	9	34	11	4	0	50

not with the fits of different models that predict these shapes. Fitting was done with the software RAD (Ulrich 2001). We used the sum of least squares of Euclidean distances of the data points from the fitted curve in Whittaker plots (Ulrich 2001, Ollik 2008) and the χ^2 -statistic in Preston plots as measures of goodness of fit. Because model fitting often gives equivocal results and different models frequently fit nearly equally well (Wilson 1991, Magurran 2004, Ollik 2008) we ranked the fits of the three models for each data set and counted how often a model fitted best (1), intermediate (2) and worst (3). Thus we did not assess statistical significance of a particular model or shape. We rather infer whether a certain SAD is nearer to one or another shape.

Significant differences in model fit between our grouping variables were obtained from Yates corrected χ^2 -tests within ordinary contingency analysis (Everitt 1980) using the 3×3 matrices of ranked model fits (3 models \times 3 ranks; Table 1). For the complete data sets we also used logistic regression to infer the possible influence of species richness, number of singletons, log-transformed number of individuals and evenness (measured by the slope of the Whittaker plot) on model fits (Whittaker plots: best vs worst). All probability levels were Bonferroni corrected to account for multiple testing.

Results

For the complete data sets a contingency table analysis pointed to a significant difference in the frequency of fits ($p(\chi^2) = 0.003$) between the three models (Table 1). The lognormal model most often fitted the complete sets (Whittaker plots: 44.5%, Preston plots: 50.4%) best, the logseries most often fitted worst in Whittaker plots (43.7%) and the power law most often worst in Preston plots (70.6%). The power law fitted the incomplete sets best (55.6%) in Whittaker plots (Table 1) and the logseries in Preston plots

Table 2. B-values and standard errors (SE) of logistic regressions to link fit to important community variables (estimated species richness, number of singletons, ln-transformed number of individuals, evenness). The best and the worst fits for 81 complete, 281 incomplete, and 24 biomass data sets for which the number of individuals was given served as dependent variable. Bonferroni corrected probability levels ($p < 0.01$) in bold type. Note that negative B-values point to better fits at higher variable values.

Complete data sets						
Variable	Lognormal		Logseries		Power function	
	B	SE	B	SE	B	SE
Species	0.01	0.01	0.02	0.01	−0.03	0.02
Singletons	−0.69	0.4	−0.27	0.28	0.65	0.35
ln N	−0.54	0.22	−0.25	0.17	0.49	0.22
Evenness	5.28	4.49	−3.75	3.1	−0.69	3.58
Incomplete data sets						
Variable	Lognormal		Logseries		Power function	
	B	SE	B	SE	B	SE
Species	−0.08	0.02	0.001	0.007	0.03	0.02
Singletons	0.57	0.13	0.1	0.04	−0.42	0.08
ln N	−0.086	0.27	−0.21	0.13	0.36	0.16
Evenness	30.32	7.63	−9.63	2.55	1.1	2.34
Biomass data sets						
Variable	Lognormal		Logseries		Power function	
	B	SE	B	SE	B	SE
Species	0.01	0.03	0.02	0.17	−0.11	0.19
Singletons	0.03	0.38	1.04	1.99	−5.14	12.06
ln N	0.05	0.37	−1.14	1.12	1.03	1.44
Evenness	−5.34	6.22	−16.68	56.83	44.12	66.35

(49.4%). Irrespective of whether the data sets were complete or incomplete the lognormal fitted the biomass data sets for both plotting methods best (Table 1). Again in the Whittaker plots the logseries did worse than in Preston plots. For both plotting methods we also found highly significant differences in the pattern of fitting between the complete, the incomplete, and the biomass sets (all $p(\chi^2) < 0.0001$). The lognormal model fitted the biomass data sets significantly more often than the complete and particularly than the incomplete sets (both $p(\chi^2) < 0.001$). Species richness, numbers of singletons, abundance, and evenness did not significantly influence model fits (Table 2) of the complete and biomass data sets. In turn, fits of the lognormal to the incomplete sets improved with rising species richness and got worse ($p(\text{no change}) < 0.01$) at higher numbers of singletons and higher evenness (Table 2). Accordingly, the logseries fitted better at higher evenness and the power law fitted better at higher numbers of singletons ($p(\text{no change}) < 0.01$).

The type of plotting had significant influence on model fits of the power law and the logseries (Table 1). For complete, incomplete and biomass data sets mean ranks of logseries fits from Preston plots were significantly lower (better fits: $p(\chi^2) < 0.001$) and mean ranks of power law fits significantly higher ($p(\chi^2) < 0.001$) than from Whittaker plots. Fits of the lognormal were not significantly influenced by the plotting type ($p(\chi^2) > 0.1$). However only 44 of the 69 communities that fitted in one of the plotting types best

Table 3. Summary statistics for 106 (continent) and 118 (latitude) complete species–abundance distributions. Ranks and fits as in Table 1. Some of the 119 complete distributions could not be classified either being from Asia (2; best fits: logseries and power law) or marine (1; best fit: lognormal).

Model	n	Whittaker			Preston		
		Rank			Rank		
		1	2	3	1	2	3
Africa							
Lognormal	25	14	8	3	14	8	3
Logseries	25	5	8	12	10	15	0
Power law	25	6	9	10	1	2	22
Europe							
Lognormal	65	29	26	10	36	21	8
Logseries	65	18	15	32	22	33	10
Power law	65	18	24	23	7	11	47
North America							
Lognormal	7	4	1	2	2	2	3
Logseries	7	0	4	3	2	5	0
Power law	7	3	2	2	3	0	4
South America							
Lognormal	9	3	2	4	3	2	4
Logseries	9	2	6	1	4	4	1
Power law	9	4	1	4	2	3	4
Australia							
Lognormal	10	2	2	6	4	3	3
Logseries	10	4	3	3	5	3	2
Power law	10	4	5	1	1	4	5
Tropical							
Lognormal	42	20	11	11	19	13	10
Logseries	42	7	18	17	16	24	2
Power law	42	15	13	14	7	5	30
Subtropical							
Lognormal	13	5	4	4	7	3	3
Logseries	13	5	4	4	5	7	1
Power law	13	3	5	5	1	3	9
Temperate							
Lognormal	63	27	25	11	32	22	9
Logseries	63	18	15	30	25	28	10
Power law	63	18	23	22	6	13	44

(63.8%) were jointly best fitted by the lognormal in Preston and in Whittaker plots.

Continent membership did not significantly influence the pattern of model fitting (Table 3) except possibly for South America, where the contingency analysis pointed to worse ($p(\chi^2) < 0.05$) fits of the lognormal model in Whittaker plots despite the low number of data sets. However, neither for Preston nor for Whittaker plots did the pattern of model fitting (Table 3) differ between temperate, tropical, and subtropical data sets (all pair wise $p(\chi^2) > 0.1$). In turn, the logseries fitted in the marine communities significantly better than in terrestrial and freshwater communities (both $p(\chi^2) < 0.05$) (Table 4) while terrestrial and fresh water data sets did not differ ($p(\chi^2) > 0.1$) with respect to the pattern of model fit.

Vertebrate and invertebrate communities were best fitted by the lognormal model (Table 5) irrespective of the type of

plotting while the logseries fitted worst in Whittaker and the power law in Preston plots. In turn, the plant data deviated from this pattern and logseries (Preston plots) and power laws (Whittaker plots) fitted best, although this was statistically significant only for the invertebrate – plant comparison ($p(\chi^2) < 0.01$). The spatial scale did not influence this pattern (Table 6) because we did not find any significant difference ($p(\chi^2) > 0.1$) between data sets from metacommunities and lower (mostly local) scales. At both scales the lognormal model fitted best (Table 6) although the differences in model fit between the three models were weak at lower scales and statistically insignificant ($p(\chi^2) > 0.05$).

Discussion

The major result of our study is that completely censused terrestrial or freshwater animal communities tend to follow lognormal type species–abundance distributions more often than logseries or power law types irrespective of species richness, spatial scale (Table 6), and geographic position (Table 3). However, plant and marine communities tend to deviate from this terrestrial pattern by more often following logseries (marine communities; Table 4) or power laws (plants; Table 5).

On the other hand, our results do not point unequivocally to a specific shape that should apply to a certain type of community and might therefore serve as a general null model for studies of species abundances. In only a few cases did one of the three models in Table 1 and 3–6 fit more than 50% of the communities best. Also our results do not allow for a rejection of a certain shape and the associated models that predict them. Therefore we strongly call for a pluralistic way of dealing with species abundances (McGill et al. 2007) and rather do not corroborate single models based on simple processes of niche division (Tokeshi 1993, 1996, 1999, Mouillot et al. 2000), fractal spatial geometry (Harte et al. 1999), statistical mechanics (Harte et al. 2008), or the central limit

Table 4. Summary statistics for 119 complete species–abundance distributions sorted according to biome (marine, terrestrial, fresh water). Ranks and fits as in Table 1.

		Whittaker			Preston		
		Rank			Rank		
Model	n	1	2	3	1	2	3
Marine							
Lognormal	10	1	5	4	3	5	2
Logseries	10	6	2	2	7	3	0
Power law	10	3	3	4	0	2	8
Freshwater							
Lognormal	17	9	5	3	8	7	2
Logseries	17	4	7	6	7	9	1
Power law	17	4	5	8	2	1	14
Terrestrial							
Lognormal	92	43	30	19	48	26	18
Logseries	92	20	28	44	32	48	12
Power law	92	29	34	29	12	18	62

Table 5. Summary statistics for 117 complete species–abundance distributions sorted according to taxon (vertebrate, invertebrate, plant). Ranks and fits as in Table 1. Two data sets stem from protists and were best fitted by a logseries (Diatomeea) and the power law (Foraminifera).

		Whittaker			Preston		
		Rank			Rank		
Model	n	1	2	3	1	2	3
Invertebrates							
Lognormal	59	29	21	9	31	22	6
Logseries	59	17	14	28	22	29	8
Power law	59	13	24	22	6	8	45
Vertebrates							
Lognormal	40	19	12	9	23	10	7
Logseries	40	9	14	17	13	22	5
Power law	40	12	14	14	4	8	28
Plants							
Lognormal	18	5	5	8	6	3	9
Logseries	18	3	9	6	8	10	0
Power law	18	10	4	4	4	5	9

theorem (May 1975, Šizling et al. 2009) that predict a single type of shape.

Our results highlight the importance of distinguishing between fully censused and sampled communities in the study of SADs (Tables 1, 2). Fully and incompletely censused communities significantly differed with regard to the best and worst fitting models. Additionally we found fits to the incomplete data sets to be affected by aggregate community properties like total species richness and evenness (Table 2). This dependence might severely bias model comparisons across communities. Particularly low coverage of the sample and higher evenness (similar abundances across species) proved to work against the lognormal. Because there were no such dependences in the fully censused communities good fits of logseries and power law SADs might simply be a statistical artifact. A similar conclusion has been drawn by Connolly et al. (2005) who found local coral and reef fish SADs to be fitted best by a truncated lognormal while fits by

Table 6. Summary statistics for 119 complete species–abundance distributions sorted according to spatial scale (metacommunity, lower scale). Ranks and fits as in Table 1.

		Whittaker			Preston		
		Rank			Rank		
Model	n	1	2	3	1	2	3
Metacommunity							
Lognormal	27	14	8	5	17	6	4
Logseries	27	4	9	14	6	17	4
Power law	27	9	10	8	4	4	19
Lower scale							
Lognormal	92	39	32	21	43	31	18
Logseries	92	26	28	38	39	44	9
Power law	92	27	32	33	10	17	65

the logseries probably resulted from undersampling (the ‘veil effect’). Previous compilations (Hughes 1986, Dewdney 2000) overlooked the distinction between complete and incomplete data sets and fitted models irrespective of the completeness of the data set. This resulted in a bias towards the logseries or similar shaped distributions (Hughes 1986, Dewdney 2000) due to a sampling artifact. Interestingly many theoretical treatises on SADs are discussing the question of sample completeness in the context of Preston’s veil line (Preston 1960, reviewed in Magurran 2007) but did not mention the need to treat complete and incomplete data sets differentially with respect to model fits. It should be noted that a similar distinction exists within communities. Magurran and Henderson (2003), Ulrich and Ollik (2004) and Ulrich and Zalewski (2006) reported that SADs of the core (persistent) species of a community were best fitted by a lognormal while the satellite (transient) species that can be seen as a sample from a larger species pool rather followed logseries.

Green and Plotkin (2007) studied the sample distributions of metacommunity SADs and showed aggregated spatial species distributions result in sample distributions that are skewed towards rare and common species. Similarly, models of species abundance based on ecological drift (Hubbell 2001, Chave 2004) predict logseries SADs of metacommunities at the regional scale and skewed lognormal SADs at local scales if dispersal limitation leads to spatial autocorrelation (Chave 2004). Our metaanalysis does not directly address sample distributions. However, if local communities are viewed as samples from the metacommunity we expect differences in shape between local and regional scales. Under a neutral framework we expect a trend towards lognormal distributions at local scales if spatial aggregation were a common phenomenon. Our results do not corroborate this prediction (Table 6). We could not find significant differences in the frequency of SAD shape between local and metacommunity scales (all $p(\chi^2) > 0.1$). Our results rather imply that local SADs are downscaled versions of the respective metacommunity SADs as reported for corals and reef fishes by Connolly et al. (2005). Such shape preserving downscaling results under random sampling (Green and Plotkin (2007). Because individuals of most species occur aggregated the similar pattern for local and metacommunity SADs with its prevalence of the lognormal shape might indicate that the local communities are not simple samples. We rather advocate a view that sees SADs at different spatial scales as being influenced by comparable ecological processes that result in similar shaped SADs. In fact, we agree with Connolly et al. (2005) that the preponderance of the lognormal irrespective of spatial scale, and taxon indicates that the central limit theorem might be of greater importance than currently assumed by neutral theorists.

Recently, the maximum entropy principle of statistical mechanics (Jaynes 1957) has become popular among ecologists as a tool for modeling species abundances and diversity (Pueyo et al. 2007, Dewar and Porté 2008; but see Haegeman and Loreau 2008). Maximum entropy proved to be particularly prone to predict exponential distributions (Haegeman and Loreau 2008) and Harte et al. (2008) accordingly predicted a prevalence of logseries SADs at local and regional scales. Again, our results rather point to

lognormal SADs at both scales (Table 6) and therefore do not back up the use of maximum entropy to abundance distributions.

Nekola et al. (2008) argued that the log-transformation used in Preston plots always results in modes of the histogram. This would imply a bias towards lognormal fits to SADs irrespective of the true abundance distribution (Nekola and Brown 2007) and might even argue against the existence of the lognormal at all (Harte pers. comm.). Again our results do not corroborate this view. Indeed the plotting type strongly influenced model fit (Table 1). However, this affected mainly logseries and power laws. Preston plots favored fits of the logseries and tended to discard power laws (Table 1). The lognormal fitted Preston plots more often than Whittaker plots (Table 1). But this was statistically not significant. Nevertheless, Nekola et al. (2007) are surely right in arguing that any binning of data might potentially influence the shape of the SAD and therefore model fit. From our results we strongly advocate the use of raw data plots (like Whittaker plots) to fit SAD models and not the use of histograms of probability distributions that are based on some sort of grouping. This view has strong implications on model choice. Important models of relative abundance are based on statistical distributions, hence on binned data. Particularly the negative binomial (Brian 1953), the logseries (Fisher et al. 1943), and the gamma (Ugland et al. 2007) distribution have been used to describe SADs but do not have simple translates into abundance-rank order or other raw data plots.

Our results corroborate recent findings by Connolly et al. (2005) and Morlon et al. (2009) that SADs obtained from biomass and/or resource use data are more lognormal than SADs obtained from abundance data (Table 1). In the Whittaker plots 78% of the biomass data sets but only 45% of the complete abundance data sets were best fitted by the lognormal. In turn, only five biomass data set was best fitted by the logseries. These findings contradict previous views that abundance and resource use are equivalent with respect to the shape of SAD distributions (Sugihara 1980, Taper and Marquet 1996). Such equivalence is expected if population biomass would scale linearly with population size (the energy equivalence rule, Damuth 1981). Any non-linear scaling as advocated by recent compilations (White et al. 2007) should influence the SAD shape if we switch from abundance to biomass and vice versa (Morlon et al. 2009). The fact that we observed a shift towards lognormal SADs when using biomass irrespective of taxon and habitat (Table 1) strongly points to a common mechanism that deserves further attention.

Our meta-analysis was mainly concerned with the basic shapes of SADs. Therefore we did not try to identify a certain model that fits best. The SAD literature is full of different models (reviewed by Magurran 2004, McGill et al. 2007) that predict similar or even identical shapes and therefore prohibit any true model identification. In our view the identification and analysis of basic SAD shapes should be a first step towards a mechanistic approach to SADs. We need models that are based on basic ecological processes and that have parameters with clear ecological interpretation. Most models lack such an interpretation. Exceptions are neutral models (Hubbell 2001) where the parameters Θ and m are linked to dispersal and metacommunity size. Recently

Ugland et al. (2007) showed that the gamma distribution as a general description of the central limit theorem fits many observed SADs of different shape but needs only two ecological meaningful parameters, the species richness and the number of niche dimensions that influence abundance. Further studies have to identify these dimensions and to link them to environmental variables (Harte 2003).

Acknowledgements – Hazel Pearson kindly improved our English. MO received a university scholarship. This work was in part supported by grants from the Polish Ministry of Science (KBN, 3 P04F 034 22, 2 P04F 039 29).

References

- Brian, M. V. 1953. Species frequencies in random samples from animal populations. – *J. Anim. Ecol.* 22: 57–64.
- Burnham, K. P. and Overton, W. S. 1978. Estimation of the size of a closed population when capture probabilities vary among animals. – *Biometrika* 65: 623–633.
- Connolly, S. R. et al. 2005. Community structure of corals and reef fish at multiple scales. – *Science* 309: 1363–1365.
- Chave, J. 2004. Neutral theory and community ecology. – *Ecol. Lett.* 7: 241–253.
- Chu, J. and Adami, C. 1999. A simple explanation for taxon abundance patterns. – *Proc. Natl Acad. Sci. USA* 96: 15017–15019.
- Damuth, J. 1981. Population density and body size in mammals. – *Nature* 290: 699–700.
- Dewdney, A. K. 2000. A dynamical model of communities and a new species–abundance distribution. – *Biol. Bull.* 198: 152–165.
- Dewar, R. C. and Porté, A. 2008. Statistical mechanics unifies ecological patterns. – *J. Theor. Biol.* 251: 389–403.
- Etienne, R. S. and Alonso, D. 2005. A dispersal-limited sampling theory for species and alleles. – *Ecol. Lett.* 8: 1147–1156.
- Everitt, B. S. 1980. The analysis of contingency tables (2nd ed.). *Monographs on Statistical and Applied Probability* 45. – CRC Press.
- Fisher, R. A. et al. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. – *J. Anim. Ecol.* 12: 42–58.
- Gray, J. S. 1979. Pollution-induced changes in populations. – *Philos. Trans. R. Soc. Lond. B* 286: 545–561.
- Green, J. and Plotkin, J. B. 2007. A statistical theory for sampling species abundances. – *Ecol. Lett.* 10: 1037–1045.
- Haegeman, B. and Loreau, M. 2008. Limitations of entropy maximization in ecology. – *Oikos* 117: 1700–1710.
- Harte, J. 2003. Ecology: tail of death and resurrection. – *Nature* 424: 1006–1007.
- Harte, J. et al. 1999. Self-similarity in the distribution and abundance of species. – *Science* 284: 334–336.
- Harte, J. et al. 2008. Maximum entropy and the state-variable approach to macroecology. – *Ecol. Monogr.* 89: 2700–2711.
- Hill, J. K. and Hamer, K. C. 2003. Using species abundance models as indicators of habitat disturbance in tropical forests. – *J. Appl. Ecol.* 35: 458–460.
- Hubbell, S. P. 2001. The unified theory of biogeography and biodiversity. – Princeton Univ. Press.
- Hughes, R. G. 1986. Theories and models of species abundance. – *Am. Nat.* 128: 879–899.
- Jaynes, E. T. 1957. Information theory and statistical mechanics. – *Phys. Rev.* 106: 620–630.
- MacArthur, R. 1957. On the relative abundance of bird species. – *Proc. Natl Acad. Sci.* 43: 293–295.

- Magurran, A. E. 2004. Measuring biological diversity. – Blackwell.
- Magurran, E. A. 2007. Species abundance distribution over time. – *Ecol. Lett.* 10: 347–357.
- Magurran, A. E. and Henderson, P. A. 2003. Explaining the excess of rare species in natural species abundance distributions. – *Nature* 422: 714–716.
- May, R. M. 1975. Patterns of species abundance and diversity. – In: Cody, M. L. and Diamond, J. M. (eds.), *Ecology and evolution of communities*. Cambridge Univ. Press, pp. 81–120.
- McGill, B. J. et al. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. – *Ecol. Lett.* 10: 995–1015.
- Miller, R. I. and Wiegert, R. G. 1989. Documenting completeness, species-area relations, and the species-abundance distribution of the regional flora. – *Ecology* 70: 16–22.
- Morlon, H. et al. 2009. Taking species abundance distributions beyond individuals. – *Ecol. Lett.* 12: 488–501.
- Moulliot, D. et al. 2000. The fractal model: a new model to describe the species accumulation process and relative abundance distribution (RAD). – *Oikos* 90: 333–342.
- Nee, S. 2003. The unified phenomenological theory of biodiversity. – In: Blackburn, T. M. and Gaston, K. J. (eds), *Macroecology: concepts and consequences*. Blackwell, pp. 31–44.
- Nekola, J. C. and Brown, J. H. 2007. The wealth of species: ecological communities, complex systems, and the legacy of Frank Preston. – *Ecol. Lett.* 10: 188–196.
- Nekola, J. C. et al. 2008. Artifacts in the log-transformation of species abundance distributions. – *Folia Geobot.* 43: 259–268.
- Ollik, M. 2008. The shape of species abundance distributions: a meta-analysis. PhD thesis. – Univ. Toruń.
- Patrick, R. 1963. The structure of diatom communities under varying ecological conditions. – *Ann. N.Y. Acad. Sci.* 108: 359–365.
- Preston, F. W. 1948. The commonness and rarity of species. – *Ecology* 29: 254–283.
- Preston, F. W. 1960. Time and space and the variation of species. – *Ecology* 41: 611–627.
- Pueyo, S. 2006. Self-similarity in species–area relationship and in species–abundance distribution. – *Oikos* 112: 156–162.
- Pueyo, S. et al. 2007. The maximum entropy formalism and the idiosyncratic theory of biodiversity. – *Ecol. Lett.* 11: 1017–1028.
- Raunkiaer, C. 1909. Formationsundersøgelse og formationsstatistik. – *Bot. Tidsskr.* 30: 20–132.
- Šizling, A. L. et al. 2009. Species abundance distribution results from a spatial analogy of central limit theorem. – *Proc. Natl Acad. Sci. USA* 106: 6691–6695.
- Sugihara, G. 1980. Minimal community structure: an explanation of species–abundance patterns. – *Am. Nat.* 116: 770–787.
- Taper, M. L. and Marquet, P. A. 1996. How do species really divide resources? – *Am. Nat.* 147: 1072–1086.
- Tokeshi, M. 1993. Species abundance patterns and community structure. – *Adv. Ecol. Res.* 24: 111–186.
- Tokeshi, M. 1996. Power fraction: a new explanation of relative abundance patterns in species-rich assemblages. – *Oikos* 75: 543–550.
- Tokeshi, M. 1999. Species coexistence. – Blackwell.
- Ugland, K. I. et al. 2007. Modelling dimensionality in species abundance distributions: description and evaluation of the Gambin model. – *Evol. Ecol. Res.* 9: 313–424.
- Ulrich, W. 2001. RAD - a FORTRAN program for the study of relative abundance distributions. – <www.uni.torun.pl/~ulrichw>.
- Ulrich, W. and Ollik, M. 2003. The internet database of relative abundance distributions. – <www.uni.torun.pl/~ulrichw>.
- Ulrich, W. and Ollik, M. 2004. Frequent and occasional species and the shape of relative abundance distributions. – *Div. Distr.* 10: 263–269.
- Ulrich, W. and Zalewski, M. 2006. Abundance and co-occurrence patterns of core and satellite species of ground beetles on small lake islands. – *Oikos* 114: 338–348.
- White, E. P. et al. 2007. Relationships between body size and abundance in ecology. – *Trends Ecol. Evol.* 22: 323–330.
- Wilson, J. B. 1991. Methods for fitting dominance/diversity curves. – *J. Veg. Sci.* 2: 35–46.
- Whittaker, R. H. 1965. Dominance and diversity in land plant communities. – *Science* 147: 250–260.