

POLISH JOURNAL OF ECOLOGY (Pol. J. Ecol.)	48	1	3-20	2000
---	-----------	----------	-------------	-------------

Werner ULRICH

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

ON SPECIES-AREA RELATIONSHIPS I: FIT OF POWER FUNCTION AND EXPONENTIAL MODEL

ABSTRACT: A FORTRAN program is developed that generates model assemblages on the basis of three basic features of animal communities: the species-weight relationship, the density-weight relationship, and the minimal density. Samplings from random placed individuals of such assemblages revealed the influence of the sampling method (sequential adding, nested and non-nested), the scale, and the underlying relative abundance distribution on resulting species-area relationships (SPARs). It is concluded that the type of the species-area relationship is not an intrinsic factor of an assemblage but depends especially on the sampling method and the unit of area. The fraction of species found only once in the sample (singletons) was the major factor influencing the model that fitted the SPAR best (at low fractions the exponential, at higher fractions the power function model). All sampling and structural factors that influence the fraction of singletons also influence the fit of the SPAR model. A mathematical derivation showed that at a certain fraction of singletons in the sample a shift from the power function to the exponential model is expected independent of assemblage type. This shift will occur between 20 and 30% singletons.

KEY WORDS: species-area relationship, model species assemblages, diversity, evenness, relative abundance distributions, random sampling

1. INTRODUCTION

The notion that species number and area are related is one of the oldest concepts in ecology and probably was first described by von Humboldt and Bonpland (1807).

The question which type of functions best describes the relationship between number of species and area (SPAR) in animals and plants and how to interpret the resulting parameters have long been discussed by ecologists. In principal, every constantly rising or asymptotic function may serve to describe this relationship and it is often more a matter of convenience or mathematical simplicity (for instance in interpreting the resulting parameters) to choose one or another model. Indeed various models have been proposed (Arrhenius 1921, Gleason 1922, Kylin 1926, Archibald 1949, de Caprariis *et al.* 1976, Coleman 1981, Coleman *et al.* 1982, Williams 1995, Harte and Kinzig 1997, Leitner and Rosenzweig 1997, Ney-Nifle and Mangel 1999, Ulrich 1999a, b).

Gleason (1922) postulated an exponential function:

$$S = s + b \ln(\text{area}) \quad (1)$$

and Fisher *et al.* (1943) and Williams (1943) showed that such a SPAR results directly from a log series relative abundance distribution.

On the other hand, Arrhenius (1921) and following him Preston (1962), Kilburn (1966), May (1975) and recently

Wissel and Maier (1992), Hanski and Gyllenberg (1997) and Harte *et al.* (1999) favored a power function of the type

$$S = s \text{ area}^z \quad (2)$$

which is often given in the double log form (to include zero counts and to normalize the distribution of S)

$$\ln(S) = \ln(s) + z \ln(\text{area}) \quad (3)$$

and gave theoretical justifications.

These two models are by far the most often used in ecology (because of mathematical simplicity and ease of interpretation) and therefore the following study will deal only with them. Other models will be dealt with in a forthcoming paper.

In a lot of cases both models, the exponential and the power function, fit nearly equally well (Connor and McCoy 1979). One reason for this is that double log transformed data are more easy to fit than untransformed or semitransformed ones resulting in the fact that 75 out of 100 SPARs studied by Connor and McCoy (1979) were fitted by a double log transformed power function model but only 38 by the exponential model. The power function is now the most widely used model to describe SPARs.

Numerous attempts have been made to derive the type of model either from a given community structure (Fisher *et al.* 1943, Preston 1962, McArthur and Wilson 1967, May 1975, Sugihara 1980) or from habitat properties (Wissel and Maier 1992, He and Legendre 1996, Harte *et al.* 1999). However, all of this derivations are based on certain assumptions and a general solution seems hardly to be derivable. As Connor and McCoy (1979) rightly pointed out, species–area relationships are a feature of their own right and need to be determined empirically. Additionally, it turned out that the choice of model and the parameter values not only depend on the intrinsic structure of a given assemblage but also on the type of sampling and the scale. Recently, especially the

latter dependence deserved much attention (Fahrig 1992, Palmer and White 1994, He and Legendre 1996, Wilson *et al.* 1998, Jonsson and Moen 1998, Rosenzweig 1999). He and Legendre (1996) showed that both models can be derived by truncating the high power terms of a general Taylor series and argued that the exponential model will fit best in small sample areas whereas the power function will be more appropriate at intermediate scales. Such a pattern had earlier been proposed by Williams (1943) and Kilburn (1963). Palmer and White (1994) found that especially the grain (sampling unit) has a profound influence on the shape of SPARs but that distance of samples, and sample number also shape the resulting curves.

It has long been noticed that the method of sampling may influence the actual form of the species–area relationship (and therefore the choice of the model to describe it) (Pielou 1977, Palmer and White 1994, Ney-Nifle and Mangel 1999) although a detailed study of this influence is still missing. In this paper the three most often applied methods to construct SPARs will be compared: 1. sequentially adding independent quadrates, 2. sequentially enlarging the area sampled resulting in a nested design, and 3. sampling independent areas of different size (non-nested design). It is often assumed that nested sampling results in lower slopes than non-nested samplings due to spatial autocorrelation of nearby quadrates (Palmer and White 1994, Leitner and Rosenzweig 1997, Jonsson and Moen 1998) caused by habitat variability (Williamson 1988). Palmer and White (1994) also found a strong scale dependence of the SPARs in case one. Increasing grain resulted in more log-linear curves (a swift towards the exponential model) indicating a gradual change in slope.

The purpose of this article and the following parts II and III (Ulrich 2000a, b) is to study the fit and the parameter values of the power function and the exponential model in detail. In doing so it analyses random sam-

plings out of model areas in which species of a variety of model assemblages were placed at random. The paper focuses then on three main questions: Which model describes the resulting SPARs best under various sampling conditions? Can the form of a SPAR be derived from underlying community structuring parameters (area heterogeneity and density fluctuations? How are type of SPAR and diversity, evenness and community structure related?

2. MATERIAL AND METHODS

Generally, model species assemblages are generated using underlying relative abundance relationships (often log-normal or log-series models) (May 1975, Leitner and Rosenzweig 1997, Wilson *et al.* 1998). However, to get a wide range of 'realistic' model assemblages another approach was undertaken: real species assemblages are characterized by three basic features. They have a certain species-weight distribution (SWD) [nearly always normal or log-normal (Novotny and Kindlmann 1996)] and a density-weight distribution (DWD) with an upper boundary line (most often described by a power function) which has frequently – but not always – a triangular form (Currie and Fritz 1993, Currie 1993, Ulrich 1999c) (Fig. 1). Thirdly, species do not have infinitely low densities, there is a lower density limit below which a species goes locally extinct.

Therefore, the model assemblages used in this study were generated using the FORTRAN program *Community Model* (Ulrich 1999b) first by assigning a specific SWD (16 weight classes, mean = 8, variance = 2), a lower and upper density limit per unit of area (a cell) (0.0001 and 0.00002 individuals per unit area) and an upper density boundary of the DWD (slope of -0.25 to -3) (Table 1). Species of assemblages with these features were then allowed to fluctuate inside the above density boundaries. In real species very high and very low densities are less of-

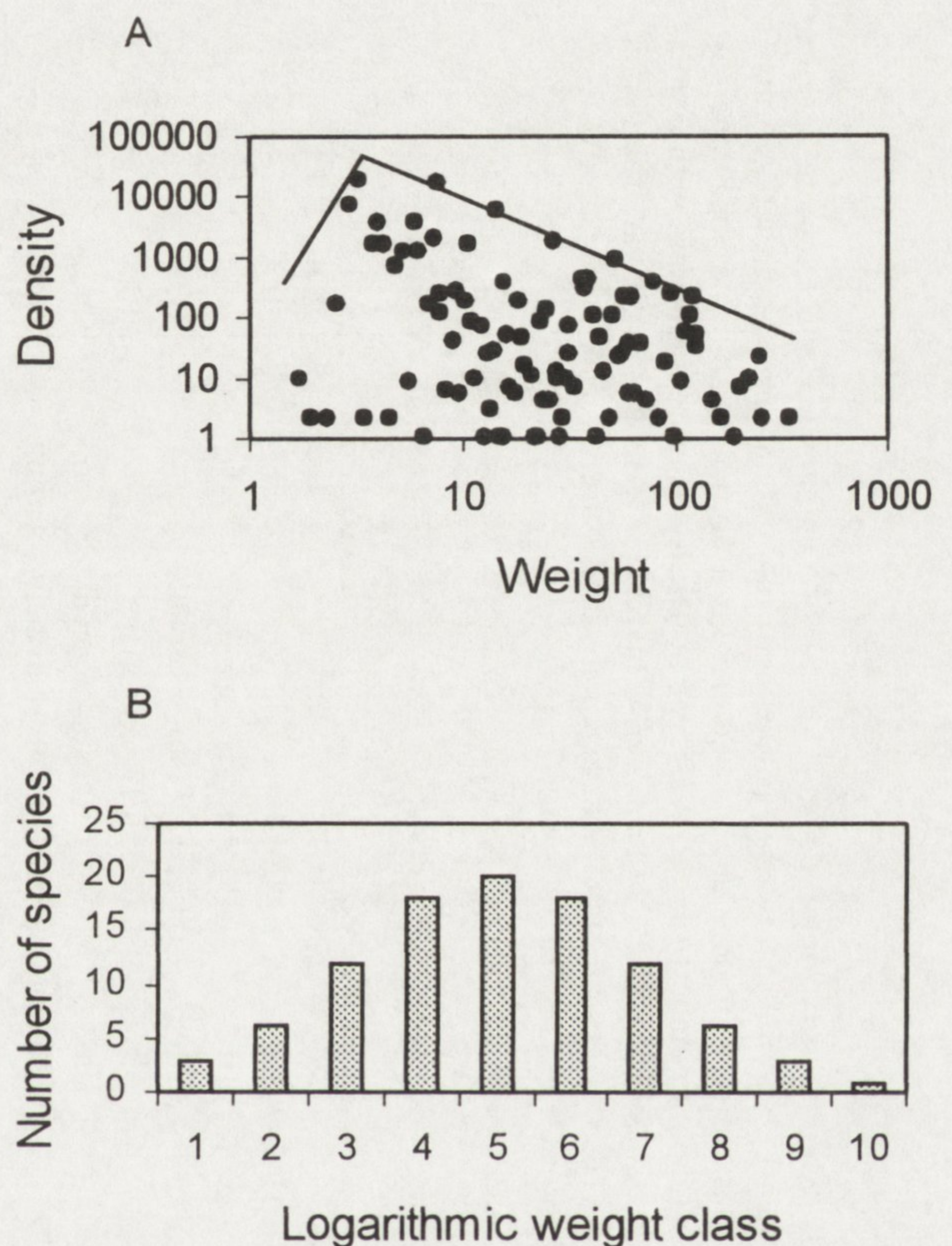


Fig. 1. A) Density-weight distribution of a model assemblage with 100 species, a normal species weight distribution with 10 weight classes, a mean of 5 and a variance of 3 after log-random density fluctuation of each species and a slope of the upper density boundary of -1.5 (marked by the line). The DWD has also the triangular form which is often found in real assemblages. B) Species – weight distribution of the same assemblage

ten found than intermediate ones, or low densities are more probable than high ones (Ulrich in prep.). To simulate this pattern the density fluctuations were either simulated by random fluctuations with log-transformed densities (resulting in a higher number of low density species) or with a normal random generator (resulting in a higher number of intermediate densities). Such a generation procedure resulted in a wide variety of community structures of which Figure 2 gives four examples. This Fig. also shows that the assemblages have typical density-weight distributions (Currie 1993).

The individuals of species of a total of 768 model assemblages (with 12 to 250 spe-

Table 1. Parameter design of Community Model to generate 768 assemblages used in this study. Max. density: upper density boundary of the smallest species; min. density: minimum allowed density per cell; type of density fluctuation: densities set randomly between the max. and min. allowed density either using untransformed densities (normal), or log-transformed densities (log-normal), or a random mixture of both (mixed). The random assessment of densities was done with a normal random number generator producing random numbers with a mean = Flucmean and a variance = Flucvariance, or with linear random numbers (non). DWD: type of density-weight relationship with a slope of the upper density boundary = DWDslope; SWD: species-weight relationship with maximum 16 binary size classes, a mean = SWDmy, and a variance = SWDvar; The degree of aggregation of the species (Aggr) was set to 0 (resulting in a value of Lloyd's index around 1) The degrees of habitat heterogeneity ranged from 0 (no heterogeneity) to 20 (high heterogeneity). The heterogeneity was either fixed for all species (Hetfix = yes) or different for each species (Hetfix = no)

No.	No. Assemblages	No. cells sampled	Method of sampling	No. traps	Unit of area	No species	Max. density	Min. density	Type of density fluctuation	FlucMean	Fluc Variance
1	120	300	seq. adding	24	1	14 to 222	10	0.0001	mixed	0.5	0.4
2	120	300	nested	24 and 70	1	13 to 220	10	0.0001	mixed	0.5	0.4
3	120	300	non-nested	24	1	15 to 214	10	0.0001	mixed	0.5	0.4
4	120	300	seq. adding	24	1	15 to 229	10	0.00002	log-normal	0.3	0.4
5	120	300	seq. adding	24	1	20 to 250	10	0.0001	normal	0.5	0.4
6	36	300	seq. adding	24	1	13 to 203	100	0.0001	log-normal	non	non
7	36	300	nested	24	1	13 to 190	100	0.0001	log-normal	non	non
8	36	300	non-nested	24	1	12 to 193	100	0.0001	log-normal	non	non
9	30	300	seq. adding	10, 20, 30,...100	1; 4; 9;...100	89 to 100	10	0.0002	mixed	0.5	0.4
10	30	300	nested	10, 20, 30,...100	1; 4; 9;...100	89 to 98	10	0.0002	mixed	0.5	0.4

No.	DWD	SWD	Classes	SWDmy	SWDvar	DWDslope	Aggr	Het	Hetfix
1	power	normal	16	8	2		0		no
2	power	normal	16	8	2	-0.25; -0.5;	0		no
3	power	normal	16	8	2	-0.75; -1;	0		no
4	power	normal	16	8	2	-1.5	0	0; 5; 10;	no
5	power	normal	16	8	2		0	20	no
6	power	normal	16	8	2	-0.5; -1;	0		yes
7	power	normal	16	8	2	-1.5; -2;	0		yes
8	power	normal	16	8	2	-2.5; -3	0		yes
9	power	normal	16	8	2	-0.25; -1;	0	0	no
10	power	normal	16	8	2	-1.5	0	0	no

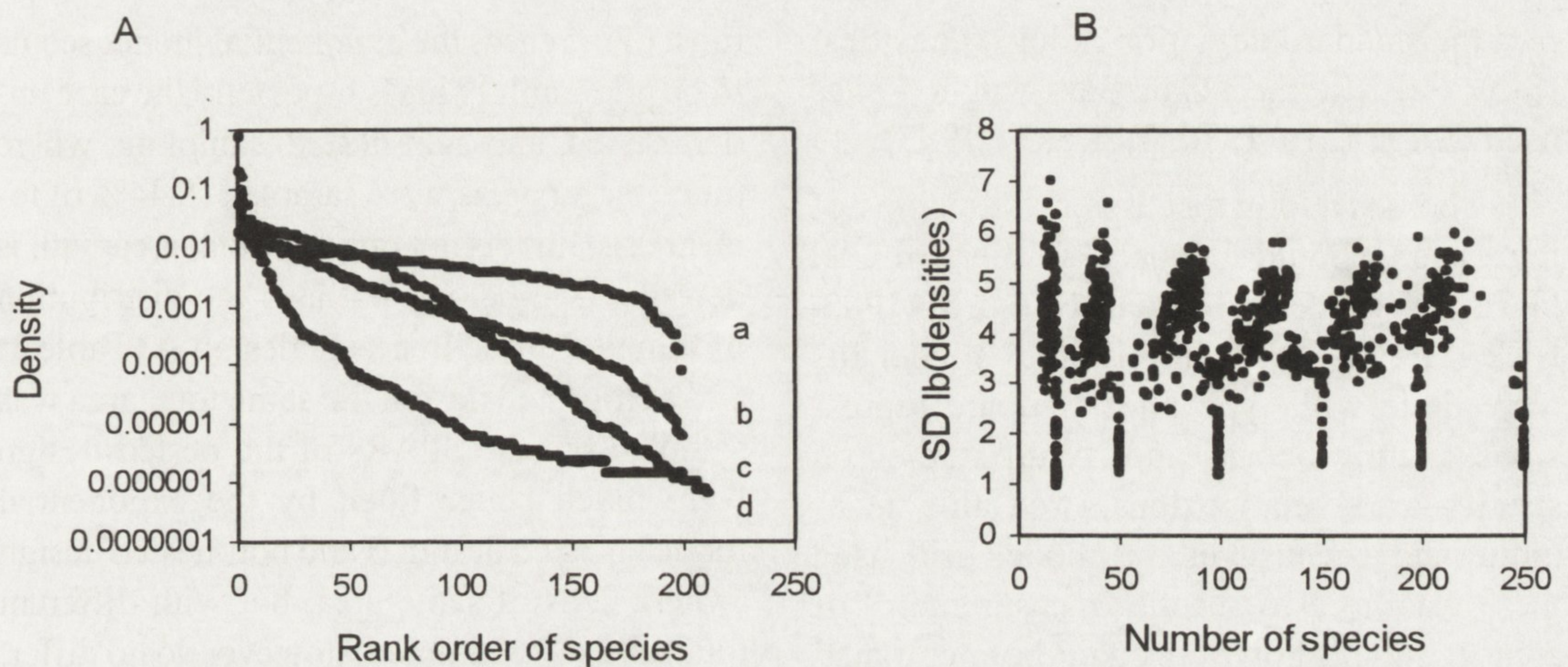


Fig. 2. A. 4 different species-abundance distributions generated with *Community Model*. a) random density fluctuations, upper boundary slope of DWD = -0.25 ; b) random density fluctuations, upper boundary slope of DWD = -1.5 ; c) mix of random and log-random density fluctuation, DWD-slope = -0.25 ; d) log-random density fluctuation, upper boundary slope of DWD = 1.25 .

B. SD of \log_2 (densities) of all assemblages of Table 1 in relation to the number of species

cies) were then placed at random into the cells of a grid of 300×300 cells. In 708 cases habitat heterogeneity was simulated giving the cells an unequal chance to be occupied by individuals [either the same heterogeneity pattern for all species (108 assemblages) or a different pattern for each species (660 assemblages)]. As a probability distribution function to generate the heterogeneity factors the formula

$$N_{r|i} = \text{het} [\sin(a_i \text{ rnd1}) + \cos(a_i \text{ rnd2})] + a_i \text{ het} \quad (4)$$

was used where $N_{r|i}$ describes the number of individuals of species i to be placed into a cell, het is the factor of habitat heterogeneity, a_i is a species specific heterogeneity factor (from 0 to 1) and rnd1 , 2 random numbers (from 1 to 300) which refer to the cell. Factors of heterogeneity of 0, 10, and 20 then resulted in individuals per cell from 0 to 90 for each placement procedure. The program however does not assume an upper limit of the number of species per cells (which would introduce a concept of competition into the model).

From these 300×300 cell grids samples were taken (1) either by sequentially adding independent cells (10 to 100 samples), (2) by

a nested design (sequentially adding larger areas: 3 to 100 samples of 1, 4, 9, 16,...4900 cells), and (3) by sampling of independent areas of different largeness (24 samples of 1, 4, 9, 16,...576 cells). Because in case 1 the sampling order has great influence on the resulting functions when computing SPARs sampling order was first reshuffled 20 times and the resulting mean values were used (following the procedure of Colwell and Coddington 1994). Table 1 gives all of the parameter values of the program, the number of assemblages generated, the resulting species numbers and the samples taken. Sampling unit (the grain) was either a cell or 4, 9, 16,...100 cells.

For each of the 768 assemblages the following parameters were computed: number of species in the assemblage S_a , species density S_{unit} (number of species per unit area), relative species density $\text{rel. } S_{\text{unit}} (= S_{\text{unit}}/S_a)$, number of species sampled S_s and fraction of total species number sampled (S_s/S_a), number of species found in only one sample (singletons), relative number of singletons (singletons/ S_s), and Shannon-diversity and evenness (whole assemblage and sample separately). As an descriptor of the relative abundance distributions of the assemblages and the sam-

ples the standard deviation of \log_2 (densities) (SD) – also termed Gaussian weight – was used (Sugihara 1980, Tokeshi 1993).

There is a long discussion which type of regression should be used to fit a model (Ricker 1973, Seim and Saether 1983, LaBarbera 1985). In this study model fits were done with type I least square regressions, mainly because nearly all studies on species–area relationships used this technique and because in the model grid used there was no error in the ‘measurement’ of area. It should however again be mentioned that this type of regression consistently yields higher slope values than other regression methods (LaBarbera 1985)

3. RESULTS

In a first step the sampling efficacy of the nested and the non-nested design was determined. With both methods a total of 4900 cells were sampled (5.44% of total area) (assemblages 2 and 3 of Table 1). Table 2 shows that there are no differences in total sampling efficacy. As expected, the degree of habitat heterogeneity influenced the proportion of species sampled. At a high degree of heterogeneity with both methods about 20% less species were sampled than in a homogeneous area.

Nearly all of the SPARs were fitted well by both models (Table 3). By optical inspection of the resulting curves however it became apparent that R^2 – values below 0.9 can hardly be accepted as good fits, therefore in

most of the cases the exponential model seems to be preferable. This is especially the case under nested and non-nested sampling where much larger areas were sampled (5.44% of total area). But it is not only the total area with is decisive (see below) but also the distribution of sampled areas. In nested design A (Table 3) and non-nested design the same total area was sampled but the SPARs of the nested design were much better fitted by the exponential model. Nested design B and non-nested design (where cells of same area, but with different total area were sampled) however do no differ. Under a sequential adding design with a low total area sampled (0.03%) both models fitted equally good.

To determine which factors influence the model fit, it’s dependence of the following variables (simple and combined) was studied: S_a , species density (S_{unit}), relative species density (S_{unit}/S_a), number of species found (S_s), S_{unit}/S_s , numbers of singletons, fraction of species found (S_s/S_a), standard deviation, diversity, and evenness of assemblage, slope of DWD, and heterogeneity. Table 4 shows that these variables are not independent but can be included into 4 main factors: a first factor which combines variables more or less associated with the sampling procedure and the habitat heterogeneity; a second factor including structural parameters of the assemblages; a third factor which loads high with S_a ; and a fourth factors loading high only which the DWD-slope. Together these factors explain 87% of total variance.

Table 2. Differences in sampling efficiency between a nested (type 2) and a non-nested sampling design. Results from 240 model assemblages placed in a 300 x 300 cell grid. In the non-nested design 24 samples were taken with areas of 1, 4, 9, 16,...576 cells (total area of 4900 cells), in the nested design 24 samples with 1, 9, 36,... 4900 cells. Low heterogeneity: heterogeneity-factor: 5; Medium heterogeneity: heterogeneity-factor: 10, High heterogeneity: heterogeneity-factor: 20.

Sampling method	Total	Without area heterogeneity	Low heterogeneity	Medium heterogeneity	High heterogeneity
Percent species found					
Nested design	82.1%	93.7%	84.1%	77.1%	73.3%
Non-nested design	81.5%	93.9%	82.9%	76.6%	72.7%

Table 3. Fits of the power function and the logarithmic model of the species–area relationship. Number of best fits, mean variance explanation (R^2) and their standard deviation derived from 636 model assemblages described in the text. In the non-nested design (ass. No. 3) 24 samples were taken with areas of 1, 4, 9, 16,...576 cells (total area of 4900 cells), in the nested design (ass. No. 2) 24 samples with 1, 9, 36,... 4900 cells (nested design A) or 24 samples with 1, 4, 9, 16,...576 cells (nested design B), in the sequential adding design 24 samples of 1 cell. Sequential adding: assemblages No. 1 and 4 to 6.

Method	Power function				Logarithmic function		
	No. Assemblages	No. best fits	Mean R^2	StDev. of R^2	No. best fits	Mean R^2	StDev. of R^2
Sequential adding	396	179	0.955	0.062	217	0.958	0.055
Nested design A	120	18	0.872	0.067	102	0.951	0.09
Nested design B	120	33	0.893	0.069	87	0.946	0.068
Non-nested design	120	30	0.869	0.07	90	0.918	0.08

Table 4. Principal component analysis (Equimax normalized rotation) to extract factors influencing the fit of the exponential or the power function model of SPAR. Variables with high loading (> 0.75) are marked in bold type. The analysis was done with all assemblages of Table 1.

Variable	Factor 1	Factor 2	Factor 3	Factor 4
Species density	−0.63	0.43	−0.46	0.08
Total number of species in the assemblage	0.11	0.15	−0.95	0.00
Fraction of species found only once	0.76	−0.18	−0.17	−0.31
Relative species density	−0.80	0.47	0.07	0.09
Species density/species found	−0.92	−0.05	0.08	0.17
Fraction of species found	−0.51	0.80	0.11	0.14
SD of assemblage	0.00	−0.91	−0.05	0.25
Diversity of assemblage	0.00	0.80	−0.53	0.12
Evenness of assemblage	−0.07	0.93	−0.10	0.15
Slope of DWD	0.38	−0.12	0.06	−0.86
Heterogeneity	0.75	0.09	0.10	0.47
Eigenvalue	4.84	2.52	1.20	1.08
Proportion of total variance	0.31	0.32	0.13	0.11

3.1. INFLUENCE OF SAMPLING

Under all three sampling regimes the proportion of species sampled had a clear effect on model fit (Fig. 3). The higher the proportion of species sampled the better was the fit of the exponential model. But Figure 3 also shows that the correlations are far from being perfect. Only 67% (A) to 7% (E) of the total variance is explained by the proportion of species sampled. The relationship was better with sequential adding sample and least pronounced under a nested sampling design. And Figures 3A, B, and C indicate that the

type of density fluctuation had a major influence on the relation. With random fluctuations on log transformed densities the above relation is much less pronounced. This latter type of density fluctuation results in a higher number of rare species and more log-linear relative abundance distributions (Fig. 2).

Indeed, the number of rare species (estimated by the number of singletons in the sample) had a major influence on model fit (Fig. 4). With sequential adding sampling both variables were highly correlated and the switch from the exponential to the power

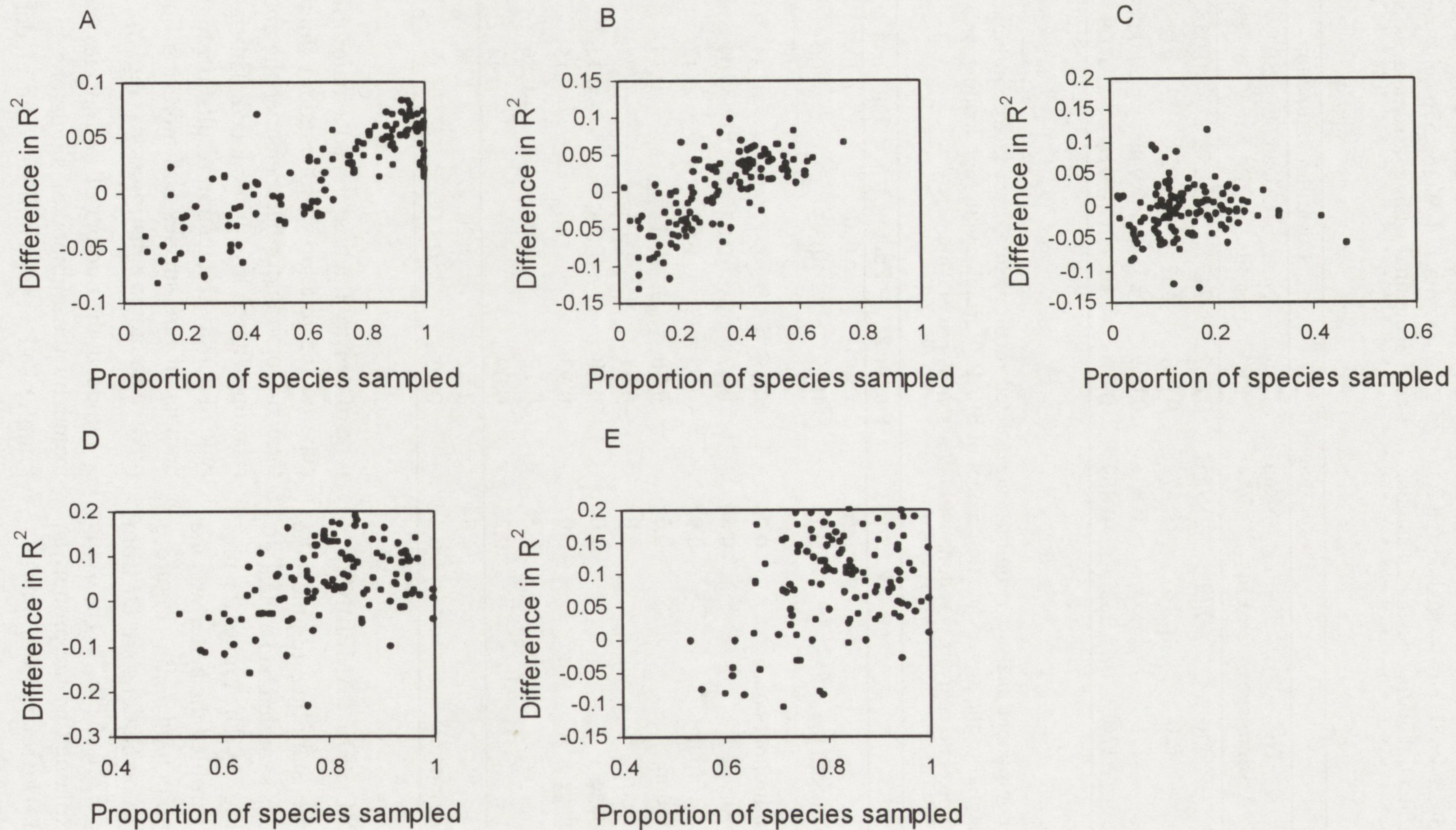


Fig. 3. Difference in variance explanation of the exponential and the power function model ($R^2_{\text{exponential}} - R^2_{\text{power}}$) of SPAR (as a measure of model fit) in relation to the proportion of species sampled. A: sequential adding, random density fluctuations, assemblages No. 5 in Table 1; B: sequential adding, mixed random log-random fluctuations, ass. No. 1; C: sequential adding, log-random fluctuations, ass. No. 4; D: non-nested design, ass. No. 3; E: nested design, ass. No. 2. Positive differences in R^2 denote a better fit of the exponential model

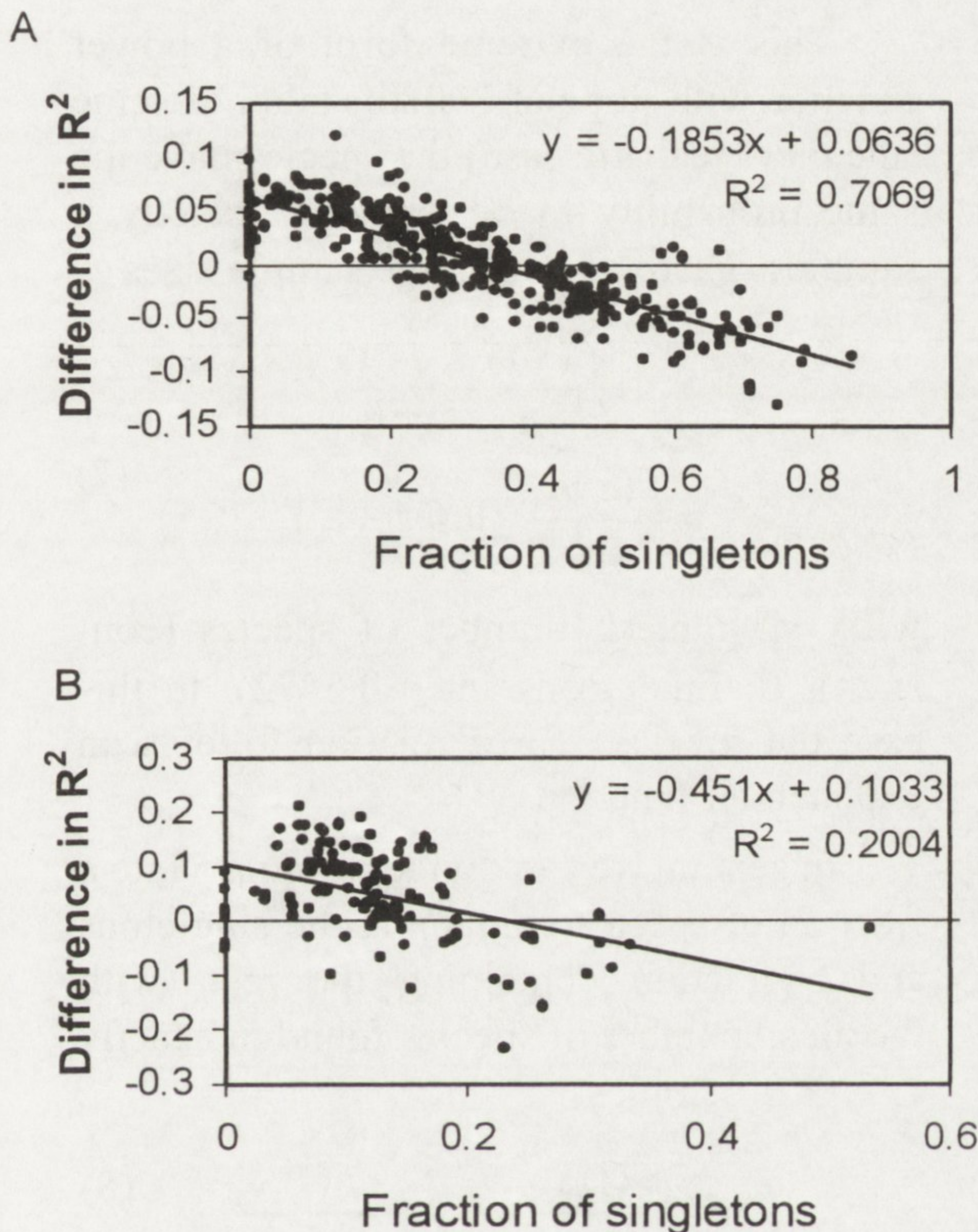


Fig. 4. Difference in variance explanation of the exponential and the power function model ($R^2_{\text{exponential}} - R^2_{\text{power}}$) of SPAR in relation to the fraction of singletons (species found only once in the sample). A: sequential adding, ass. No. 1, 4–6; B: non-nested design, ass. No. 3

function model occurred above 1/3 singletons. If less than 20% singletons occurred only 1 out of 77 assemblages (1.3%) was better fitted by the power function model but if

more than 40% of the species were singletons 120 out of 129 (93%) did so. In the middle range both model fit nearly equally well. Visual inspection of the plots showed that good fits all have values of variance explanation above 95%, below 90% the fit can hardly be called well. Plots of R^2 -values against fraction of singletons showed that in the middle range (between 10 and 50% singletons) both models have R^2 -values above 95% (power function: 82%, exponential model: 89%). The performance of the power functions dropped steeply below 7.5% singletons with 15 out of 21 R^2 -values below 90%. The exponential model generally performed slightly better: R^2 -values below 90% occurred only below 5% and above 70% singletons.

The above relationship appeared to be independent of sample size. With sample sizes from 10 to 100 (assemblages No. 9 in Table 1) the same correlation came up (data not shown) and the model fit was not correlated with sample size (Fig. 5B). There was, however, a dependence on unit of area (Fig. 5A). The larger the unit the less pronounced was the above correlation. A units above 3×3 cells the relation vanished because the fraction of singletons dropped below 20% and nearly all of the assemblages were better fitted by the exponential model (Fig. 5A) (the relation between unit of area and percentage of singletons followed a logarithmic function: % singletons = $-0.06 \ln(\text{unit}) + 0.21$, $r = 0.55$,

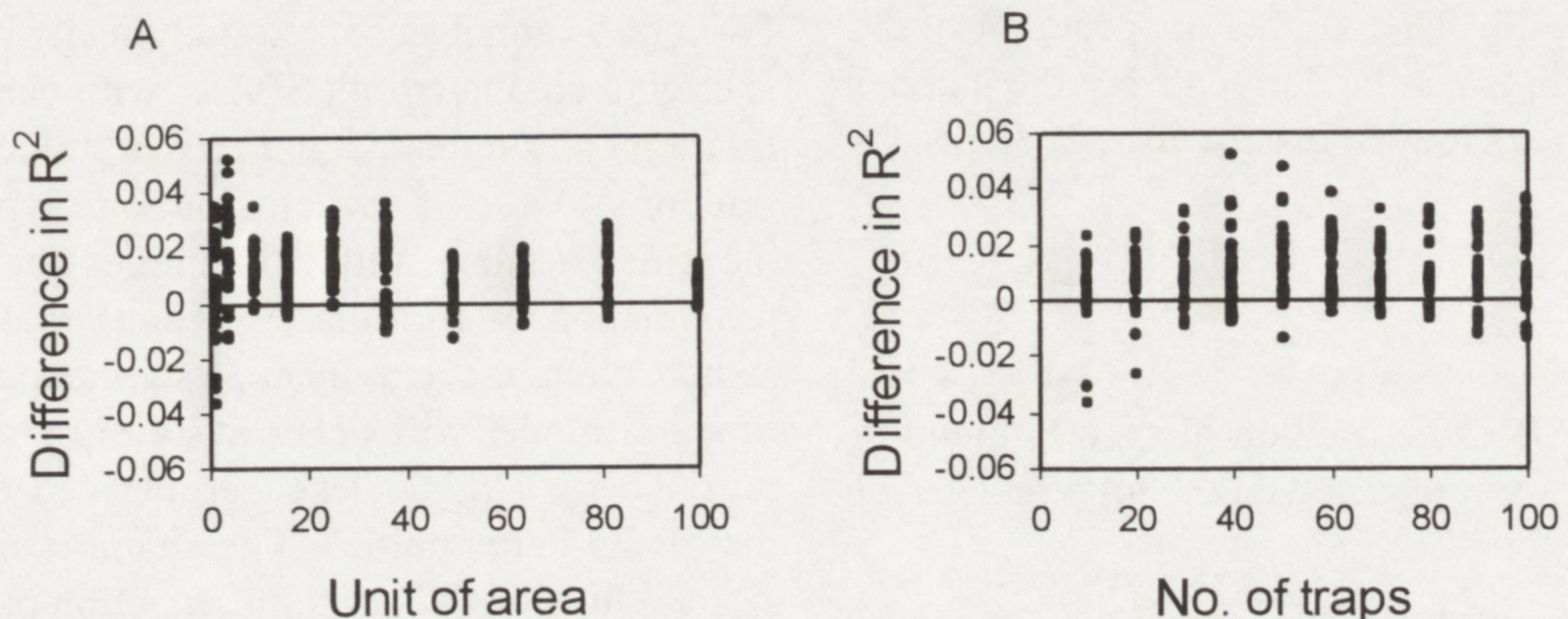


Fig. 5. Difference in variance explanation of the exponential and the power function model ($R^2_{\text{exponential}} - R^2_{\text{power}}$) of SPAR in relation to unit of area (A) and number of samples (B). Ass. No. 9 in Table 1

$p(t) < 0.01$). In the case of sample size, on the other hand, no such relation could be detected. Both variables appeared to be not correlated.

On what underlying parameters depend the fraction of singletons in the sample? A multiple correlation revealed a complicated pattern, with no single main influencing factor. Significantly positively correlated with the fraction of singletons were the species density (β -weight = 0.19), the diversity ($\beta = 0.84$), the DWD slope ($\beta = 0.21$), and the heterogeneity ($\beta = 0.19$), negatively correlated were the number of species found ($\beta = -0.5$), the relative species density ($\beta = -0.39$), SD ($\beta = -0.25$), and the evenness ($\beta = -0.67$) (total multiple r of the model: $r = 0.76$). The low variance explanation (58%), however, indicates that it is not possible to derive unambiguously the fraction of singletons from the underlying assemblage structure. Stochastic effects of the sampling procedure result in a comparatively high variance in the number of singletons.

Figure 4 also shows that a non-nested sampling design (the nested design is not plotted due to the low number of singletons: in each of the assemblages less than 10%) leads to a different result. Although the general trend is the same as in the case of sequential adding the intercept appears to be lower (at 22%) and the variance is much higher.

The above result can also be obtained analytically. The number of species S_x of the x -th sample can be derived from the number of species sampled exactly in 1, 2, 3... n samples.

$$S_x = \sum_{i=1}^x \frac{f(i)S_n}{ix} \quad (5)$$

where $f(i)$ is the fraction of species found in exactly i samples and S_n the total number of species found.

If all of the species found only once this formula reduces to:

$$S_x = x (f(1) S_n/n) \quad (6)$$

This is the extreme form of a power function with slope of 1. In the other extreme (the more realistic case) the species have the same probability to be found in exactly x samples. Therefore, for large sample sizes

$$S_x = \frac{f(\bar{x})S_n}{n} \sum_{i=1}^x \frac{1}{x} \approx \frac{f(\bar{x})S_n}{n} (\ln(x) + C) \quad (7)$$

with $f(\bar{x})$: mean number of species (constant), C : Euler constant (≈ 0.5572). In this case the species – area relation follows an exponential function.

In general we can divide formula 5 into a fraction of species containing the singletons and a fraction containing the rest (with the mean number of species found in exactly 1, 2, 3... n samples:

$$S_x = \frac{f(1)S_n}{n}x + \sum_{i=2}^x \frac{f(i)S_n}{(n-1)x} \quad (8)$$

$$\approx \frac{f(1)S_n}{n}x + \frac{f(\bar{i})S_n}{(n-1)} \sum_{i=2}^x \frac{1}{x}$$

$$S_x \approx \frac{f(1)S_n}{n}x + \frac{f(\bar{i})S_n}{(n-1)} (\ln(x) + (C-1)) \quad (9)$$

Because $f(1) = S_n - (n-1) f(\bar{i}) S_n$ the SPAR model (power function or exponential depends on the fraction of singletons $f(1)$: at a certain fraction $f(1)$ there will be a switch from the power to the exponential model. Formula 9 cannot be solved analytically but it is easy to compute both SPARs with various fractions of singletons and to determine the fraction at which a switch from one type to the other occurs. With 10 samples the approximation of formula 9 predicts that the switch from the exponential to the power function model will occur at 24.8% singletons. This is slightly less than derived from the model communities. The shape is not a straight line but a logarithmic function due to the assumption of equal distribution of non-singletons. With a constant number of species sampled the model also depends on the

number of samples: the higher this number the less will be the point of switch (with 20 samples the model predicts a switch at 18% singletons).

One factor that influences the percentage of singletons is the relative species density (*rel. S_{unit}*): the higher *rel. S_{unit}* the lower the percentage of singletons (*No. Singletons* = $-0.13 \times \ln(\text{rel. } S_{\text{unit}}) + 0.03$; $r = 0.82$, $p < 0.0001$). However, there is no simple linear relationship between relative species density and model fit (Fig. 6). The relation is more bell shaped, a feature which is especially pronounced under sequential adding sampling. In this case, at low relative species densities (below 10%), the power function gave the best fits. The higher *rel. S_{unit}* the better was the fit of the exponential model. At *rel. S_{unit}* above 50% both models fitted nearly equally well. Under a nested and a non-nested sampling design this relation still can be detected but at low *rel. S_{rel}* there is no clear trend towards one of both models. A plot of species density/species number sampled (S_{unit}/S_s) resulted in plots which look very similar to the ones of Figure 6 (of course due to the high correlation of S_s and S_a : $r = 0.78$, $p(t) < 0.0001$; data not shown).

Is it possible to predict the model of best fit by a combination of parameter values. For this purpose a stepwise multiple regression was run with the difference in variance explanation of both SPAR models as dependent and the high loading variables of the first factor (Table 4) as independent variables (sequential adding sampling only). Table 5 shows the resulting model with the highest total variance explanation. In 85.7% of the cases this model predicted the correct SPAR-model (power or exponential).

What are the properties of the 50 assemblages of which no SPAR-model could be predicted with the regression of Table 5. Two factors could be identified which differ between the correctly and non-correctly predicted group: the relative species density and the diversity. All 50 assemblages had compa-

rably low relative species densities (mean of 0.058 but 0.12 in the correctly predicted group) and low diversities (1.9 and 2.7, respectively).

3.2. INFLUENCE OF ASSEMBLAGE STRUCTURE

Figure 7 shows that there is no simple relationship between SPAR model and assemblage structure. In the case of a nested and a non-nested sampling design model fit and SD were positively correlated and at SD values above 4 a switch from a power function SPAR to the exponential model occurred. The same was observed under a sequential adding design however with a larger variance. At low SD values the above correlation reversed. SD-values below 2 again favored the exponential model. This behavior coincided with the relation between fraction of singletons and SD (Fig. 8). This relation was strongly bell shaped with low fractions (below 20%) at the high (above 5) and the low end (below 2) of the SD-scala.

The results of a multiple regression based on the variables describing the structure of the assemblages gives Table 6. The higher the diversity and the habitat heterogeneity, the lower the evenness and the steeper the slope of the upper boundary of the DWD, the better was the fit of the power function. However, although being highly significant the variance explanation of the regression model was low and the predictive power was only 51.4% what means that in nearly half of the cases it proved to be impossible to predict the SPAR model from underlying assemblage parameters.

This result is strengthened by the fact that of the partial correlations of all of the above variables and the model fit none of the correlations of the structural variables remained significant (Table 7). The fraction of singletons dominated, a lesser influence had the relative species density (estimated from the sample) and the fraction of species found. A small influence had also the heterogeneity.

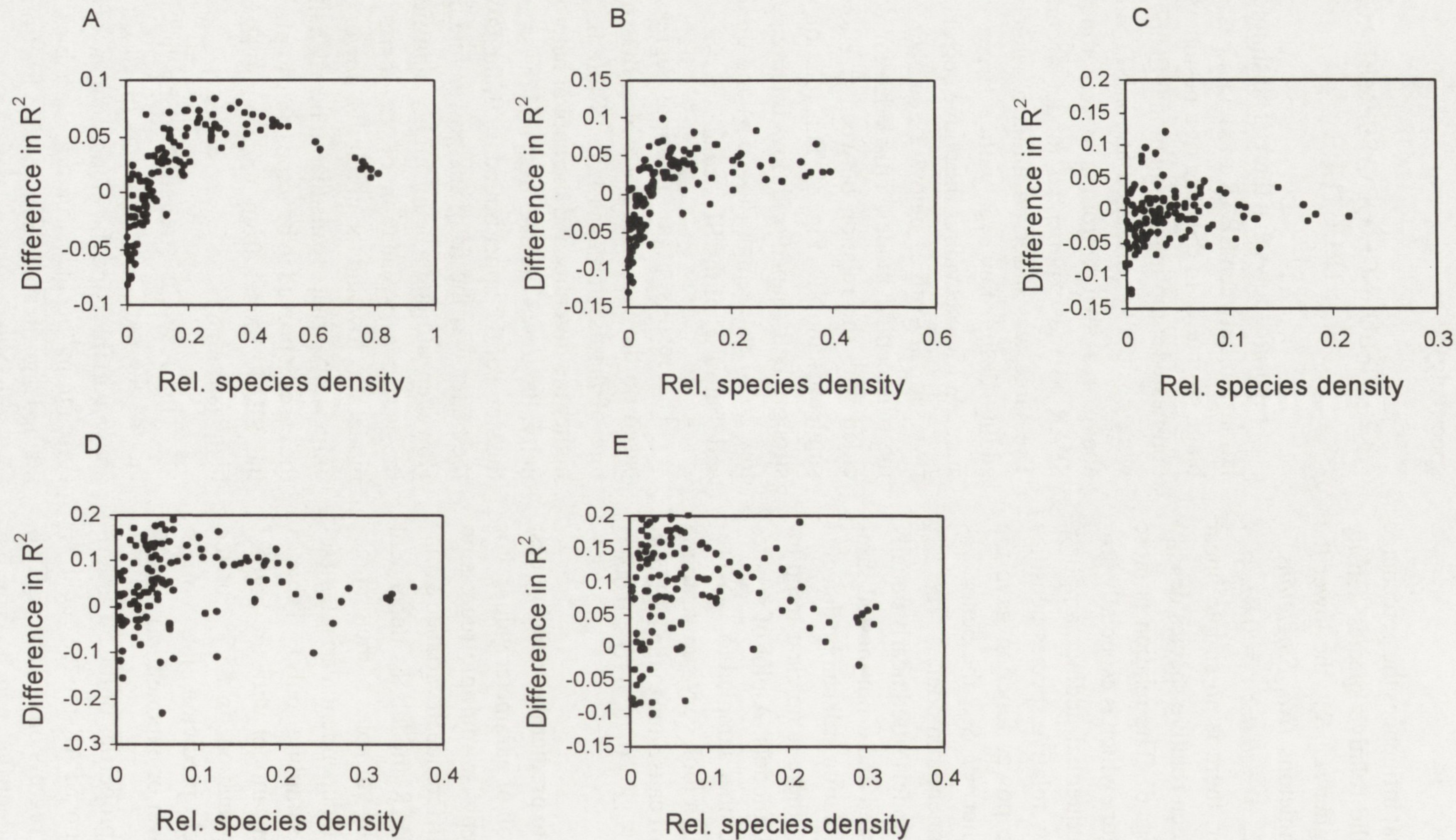


Fig. 6. Difference in variance explanation of the exponential and the power function model ($R^2_{\text{exponential}} - R^2_{\text{power}}$) of SPAR in relation to the relative species density. A: sequential adding, random density fluctuations, assemblages No. 5 in Table 1; B: sequential adding, mixed random log-random fluctuations, ass. No. 1; C: sequential adding, log-random fluctuations, ass. No. 4; D: non-nested design, ass. No. 3; E: nested design, ass. No. 2

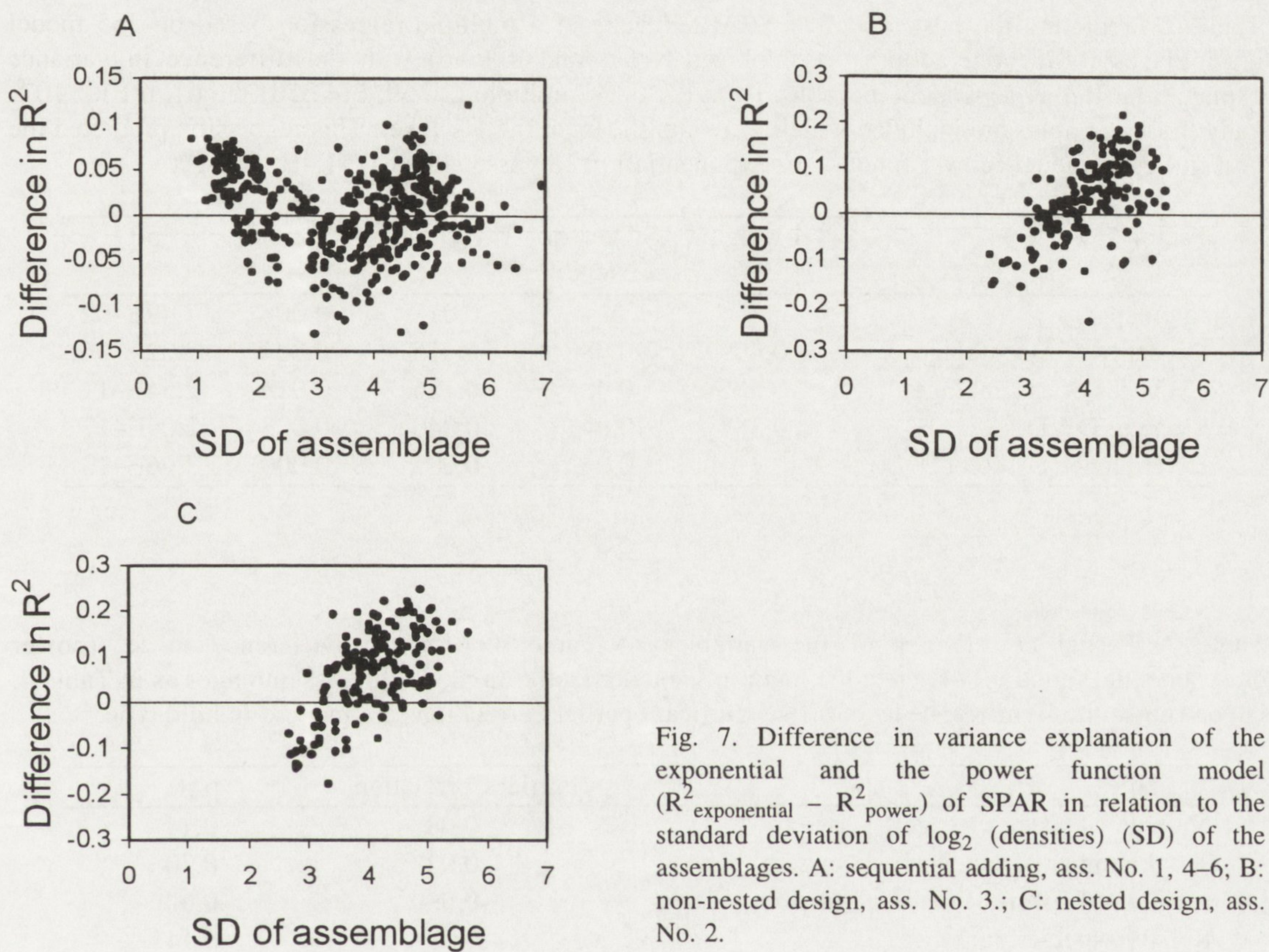


Fig. 7. Difference in variance explanation of the exponential and the power function model ($R^2_{\text{exponential}} - R^2_{\text{power}}$) of SPAR in relation to the standard deviation of \log_2 (densities) (SD) of the assemblages. A: sequential adding, ass. No. 1, 4–6; B: non-nested design, ass. No. 3.; C: nested design, ass. No. 2.

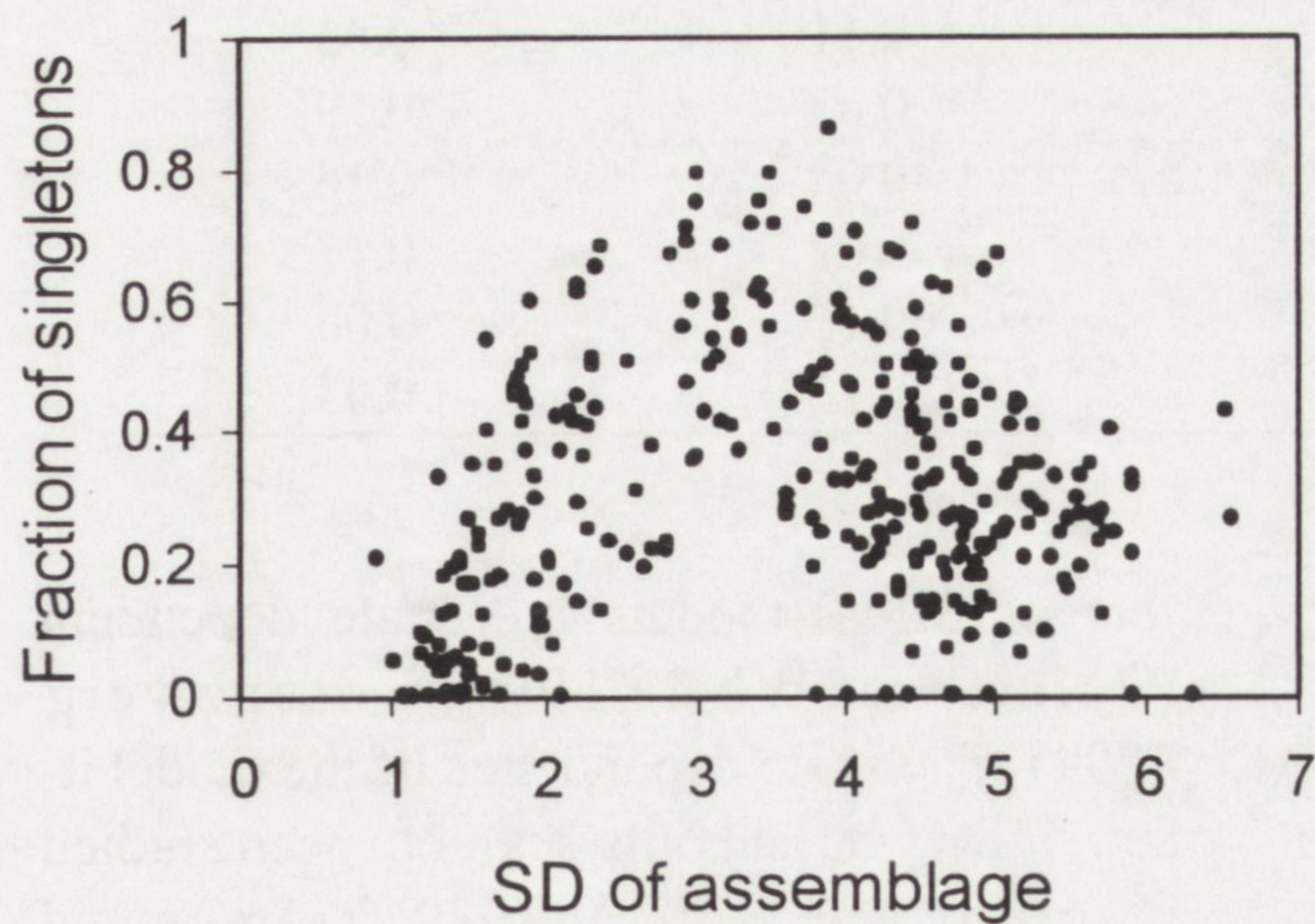


Fig. 8. Fraction of singletons in relation to the standard deviation of \log_2 (densities) (SD) of the assemblages (sequential adding sampling) (No. 1, 4–6)

Table 5. Predicting the best model of SPAR. Results of a multiple regression based on 355 model assemblages (sequential adding, assemblages 1 and 4 to 6) with the difference in variance explanation (R^2) as dependent variable. Total R^2 : 0.69; multiple r : 0.83; $F(3, 351)$: 254.55, $p(F) < 10^{-6}$. Only those variables are included which were significant at the 1% level. This regression predicted the best fit SPAR model (power function or exponential) in 305 assemblages (85.7%) correct.

Factor	β -weight	StDev. of β -weight	B	StDev. of B	p(t)
Intercept			0.07	0.01	2.334E-23
Proportion species sampled	0.17	0.04	0.03	0.01	6.6791E-06
Species density/No. species sampled	-0.25	0.04	-0.07	0.01	7.3708E-09
Proportion species found only in one sample	-0.87	0.05	-0.18	0.01	0

Table 6. Predicting the best model of SPAR. Results of a multiple regression based on 355 model assemblages (sequential adding, assemblages No. 1 and 4 to 6) with the difference in variance explanation (R^2) as dependent variable. Total R^2 : 0.34; multiple r : 0.59; $F(4,350)$: 56.61, $p(F) < 10^{-6}$. Only those variables are included which were significant at the 1% level. This regression predicted the best fit SPAR model (power function or exponential) in 183 assemblages (51.4%) correct.

Factor	β -weight	StDev. of β -weight	B	StDev. of B	p(t)
Intercept			0.015	0.008	7.49E-02
Diversity of assemblage	-0.409	0.086	-0.015	0.003	3.22E-06
Evenness of assemblage	0.602	0.087	0.126	0.018	2.54E-11
Slope of DWD	0.387	0.043	0.040	0.005	2.09E-17
Heterogeneity	-0.314	0.042	-0.002	0.000	3.96E-13

Table 7. Partial correlations of the variables of Table 4 with the Difference in R^2 (power or exponential function) to detect the main influencing factor on model fit. Assemblages as in Table 4. Given is also the significance level p(t). Significant partial correlations are marked in bold type.

Variables	Partial Correlation	p (t)
Number of species per cell	0.08	0.14
Total number of species in the assemblage	0.02	0.70
Fraction of species found only once	-0.65	0.00
Relative species density	-0.08	0.14
Species density/species found	-0.16	0.00
Fraction of species found	0.11	0.03
SD of assemblage	0.03	0.59
Diversity of assemblage	-0.08	0.16
Evenness of assemblage	0.07	0.17
Slope of DWD	0.10	0.06
Heterogeneity	-0.12	0.03

4. DISCUSSION

Ecologists have long thought that the species–area relationship is an intrinsic property of a certain plant or animal community and that the actual form of this relation can be derived by analyzing the structure of a community (or vice versa deriving structural properties by studying SPARs) (Fisher *et al.* 1943, Preston 1962, May 1975, Schoener 1976, Engen 1977, Martin 1981, McGuinness 1984, Wright 1988, Tokeshi 1993, Harte and Kinzig 1997, Leitner and Rosenzweig 1997, Harte *et al.* 1999). The reborn interest in species–area relations have shown that the actual parame-

ters of SPAR-models are scale dependent (Palmer and White 1994, Rosenzweig 1999). The scale dependence of the model itself, however, had only scarcely been studied (Connor and McCoy 1979, Palmer and White 1994, He and Legendre 1996) and the power function is generally accepted as a fundamental property of most communities (Hanski and Gyllenberg 1997, Harte *et al.* 1999) although the exponential model frequently gives better fits – especially to plant data (He and Legendre 1996).

The aim of the present study was to analyze the influence of three basic factors on the exponential and the power function model of SPAR: the type of sampling, the scale and the

underlying community and habitat structure. It was shown that all these factors influence the model fit.

Methods of sampling are difficult to compare because changing the method inevitably changes several variables simultaneously. Compared with a sequential adding technique, a nested sampling design frequently enhances the proportion of species sampled (due to the larger area sampled) and favors therefore an exponential SPAR model. But the comparison between the nested and the non-nested design showed that it is not only the area that made these difference (Table 3). The most pronounced effect of nested sampling (and to a lesser degree in the case of a non-nested design) is the reduction of the number of singletons in the sampling, which turned out to be of major importance for model fit. Nested and non-nested sampling resulted also in a higher variance in plots of sampling or structural parameters against model fit (Figs 3, 4, 6 and 7). Both sampling methods lowered therefore the possibility to predict the model of best fit.

The scale of sampling can be subdivided into three major variables: the unit of area, the number of samples, and the area sampled (Addicot *et al.* 1987). In line with the findings of Palmer and White (1994) the unit of area (the grain) had a profound influence on model fit, this again mainly due to the reduction of the number of singletons. Doubling the grain resulted in 1.2 to 2 times the number of species per unit area and reduced the number of singletons by the factor of 1.3 to 1.8. At units above 0.01% of total area the SPARs were nearly always better fitted by the exponential model. The number of equal sized samples, on the other hand, had only a minor influence on the fit.

Palmer and White (1994) also concluded pronounced influence of the extent, the area sampled, on the model fit. The above results indicate that this is only the case if fairly small fractions of the total area are sampled. Extensive sampling enhances the pro-

portion of species sampled and reduces the fraction of singletons, which both influences the shape of the SPAR.

He and Legendre (1996) developed the power function and the exponential model by truncating the terms of a Taylor series and argued that the power function will fit better at larger areas sampled, the exponential model at smaller areas. My results do not confirm these predictions. Instead, larger sampling areas resulted nearly always in a better fit of the exponential model due to the higher proportion of species sampled (Fig. 3).

Harte *et al.* (1999) assumed that area has a fractal nature. The concept of self-similarity then results directly in a power function SPAR. The above results cast doubts on this concept. A random placement of individuals into a grid with no heterogeneity results in similar structured cells, but do not inevitable result in power function SPARs (even when ignoring the effects of the sampling method). Self similarity would also have profound influence on the sampling itself. It would rule out really independent sampling – which is according to Preston (1962) necessary for a power function SPAR in the case of log-normal relative abundance distributions – but would result in a form of pseudoreplication (Hurlbert 1984) similar to a sequential adding design.

The most interesting question is how assemblage structure effects the species–area relationship. For all who seek a direct relationship between SPAR and underlying community structure (especially the form of the relative abundance distribution) the above results are disappointing. The low predictive power of the linear regression model of Table 6, the high variance in the relation between SD and model fit and the low partial correlations between diversity, evenness, SD and total species number and model fit show that there are no simple relations between these variables. Most important is the proportion of rare species (species with densities at or below $1/n$) in the community and the relation

between sample size (n) and fraction of rare species in the sample. In the model communities used here the latter relation was rather constant resulting in an independence of model fit and sample size but more detailed studies in natural communities are necessary to decide whether this is a common feature.

Of course the relative abundance distribution of an assemblage influences the number of rare species. The bell shaped relation between SD-value and model fit indicates however that this relation may be rather complicated. Figure 9 gives the SD-values for all major theoretical relative abundance distributions in relation to the number of species (for detailed descriptions of these models see Pielou 1977, Frontier 1985, Tokeshi 1993, 1996). Figure 8 shows the dependence of the SD-value on the fraction of singletons in the sample. It appears that at very high and at very low SD values the fraction of singletons is low which results in a better fit of the exponential model (Figs 4 and 7) and causes the bell shaped plot in Figure 6. Therefore, it is concluded that a dominance preemption or a logarithmic model (the former only being the stochastic equivalent of the latter) will mostly being fitted better by an exponential model (under appropriate sample conditions). The same will be the case with the relative abundance distributions with SD-values below 2 (random assortment, broken stick, dominance decay and overlapping niche). In the case of a Zipf-Mandelbrot

distribution the SPAR model will depend on the parameter values. For the broken stick model May (1975) expected a power function SPAR but the data presented here point to an exponential model. Random fraction, power fraction distributions and the log-normal should generally be better fitted by a power function SPAR. In the case of the log-normal this is of course the expected result (Preston 1962, May 1975). Log-series and geometric model are related and it has long been known that a log-series distributions leads to an exponential model of SPAR. The same result can be obtained from Figure 8.

Sugihara (1980) and following him Tokeshi (1996) showed that most larger animal assemblages have SD values in the range between 2 and 6. This is exactly the range in which a better fit of the power function model is expected and probably the cause why this model is by far the most widely used. The above results however show that such a relation is far from being perfect and depends highly on the sampling method and the grain. The only really good predictor of SPAR model proofed to be the fraction of singletons in the sample. While this variable depends on sampling method and assemblage structure it explains why and how both factors influence the fit of a certain SPAR model. Each change in sampling which enhances the fraction of singletons and relative abundance distributions which gen-

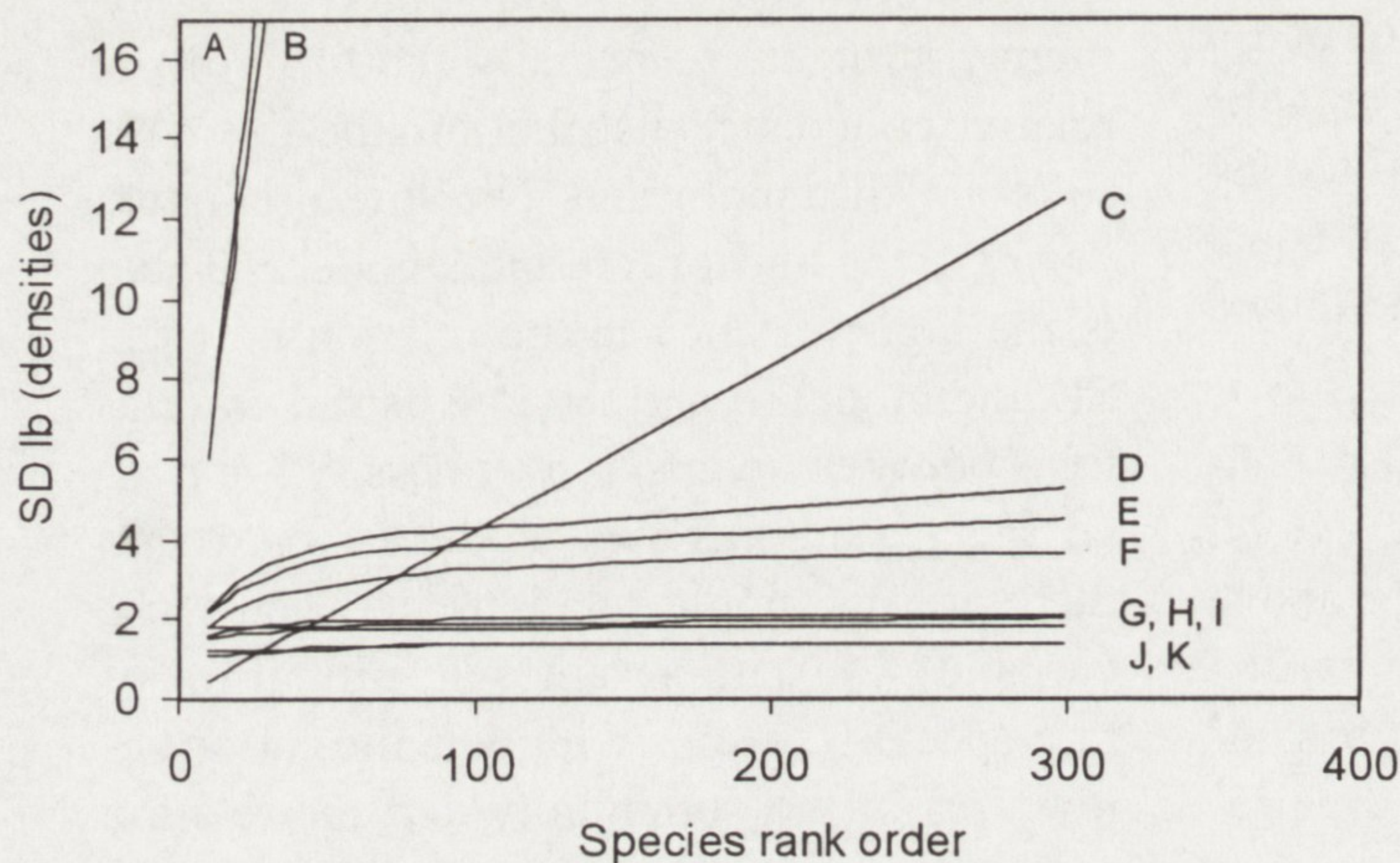


Fig. 9. SD \log_2 (densities) of various theoretical relative abundance distributions in relation to number of species. Models: A: dominance preemption, B: geometrical ($k = 0.75$), C: log-series ($\alpha = 10$), D: random fraction, E: power fraction ($k = 0.05$), F: canonical log-normal, G: Zipf-Mandelbrot ($z = 1.5$, $X = 0$), H: random assortment, I: broken stick, J: dominance decay, K: overlapping niche. The plots were computed with the program *Niche*, with computes and fits theoretical distributions (Ulrich in prep.)

erate high proportions of rare species will favor power function SPARs.

ACKNOWLEDGMENTS: I thank Prof. J. Buszko and Dr. Kartanas for critical and valuable suggestions on the manuscript. Miss H. Pearson kindly improved my English.

5. SUMMARY

A FORTRAN program is developed that generates model assemblages on the basis of three basic features of animal communities: the species-weight-relationship (SPAR), the density-weight relationship, and the minimal density (Table 1, Figs 1 and 2). Samplings from random placed individuals of such assemblages revealed the influence of the sampling method (sequential adding, nested and non-nested), the scale, and the underlying relative abundance distribution (Fig. 8) on resulting species-area relationships (Figs 3, 4, 5, 6 and 7, Tables 2, 3, 4, 5, 6 and 7). It is concluded that the type species-area relationship is not an intrinsic factor of an assemblage but depends especially on the sampling method and the unit of area. The fraction of species found only once in the sample (singletons) was the major factor influencing the model that fitted the SPAR best (at low fractions the exponential, at higher fractions the power function model) (Figs 4 and 9). All sampling and structural factors that influence the fraction of singletons also influence the fit of a SPAR model. A mathematical derivation showed that at a certain fraction of singletons in the sample a shift from the power function to the exponential model is expected independent of assemblage type. This shift will occur between 20 and 30% singletons.

6. REFERENCES

- Addicot J. M., Ado J. M., Antolin M. F., Padilla D. K., Richardson J. S., Soluk D. A. 1987 – Ecological neighborhoods: scaling environmental patterns – *Oikos*, 49: 340–346.
- Archibald E. E. A. 1949 – The specific character of plant communities. II. A quantitative approach – *J. Ecol.* 37: 260 – 274.
- Arrhenius O. 1921 – Species and area – *J. Ecol.* 9: 95–99.
- Coleman B. D. 1981 – On random placement and species – area relations – *Math. Bioscience*, 54: 191–215.
- Coleman B. D., Mares M. A., Willig M. R., Hsieh Y. H. 1982 – Randomness, area, and species richness – *Ecology*, 63: 1121–1133.
- Colwell R. K., Coddington J. A. 1994 – Estimating terrestrial biodiversity through extrapolation – *Phil. Trans. R. Soc. Lond. B*: 345: 101–118.
- Connor E. F., McCoy E. D. 1979 – The statistics and biology of the species-area relationship – *Am. Nat.* 113: 791–833.
- Currie D. J. 1993 – What shape is the relationship between body mass and population density – *Oikos*, 66: 353.
- Currie D. J., Fritz J. T. 1993 – Global patterns of animal abundance and species energy use – *Oikos*, 67: 56–68.
- De Caprariis P., Lindemann R. H., Collins C. 1976 – A method for determining optimum sample size in species diversity studies – *Math. Geol.* 8: 575–581.
- Engen S. 1977 – Exponential and logarithmic species area curves – *Am. Nat.* 111: 591–594.
- Fahrig L. 1992 – Relative importance of spatial and temporal scales in a patchy environment – *Theor. Pop. Biol.* 41: 300–314.
- Fisher A. G., Corbet S. A., Williams S. A. 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.
- Frontier S. 1985 – Diversity and structure in aquatic ecosystems (in: *Oceanography and Marine Biology – An Annual Review* Ed. M. Barnes) – Aberdeen, pp. 253–312.
- Gleason H. A. 1922 – On the relation between species and area – *Ecology*, 3: 158–162.
- Hanski I., Gyllenberg M. 1997 – Uniting two general patterns in the distribution of species – *Science*, 275: 397–400.
- Harte J., Kinzig A. P. 1997 – On the implications of species – area relationships for endemism, spatial turnover, and food web patterns – *Oikos*, 80: 417–427.
- Harte J., Kinzig A. P., Green J. 1999 – Self-similarity in the distribution and abundance of species – *Science*, 284: 334–336.
- He F., Legendre P. 1996 – On species area relations – *Am. Nat.* 148: 719–737.
- Humboldt A. V., Bonpland A. 1807 – *Essai sur la Géographie des Plantes* – Paris, Librairie Lebrault Schoell.
- Hurlbert S. H. 1984 – Pseudoreplication and the design of ecological field experiments – *Ecol. Monogr.* 54: 187–211.
- Jonsson B. G., Moen J. 1998 – Patterns in species association in plant communities: the importance of scale – *J. Veg. Sci.* 9: 327–332.
- Kilburn P. D. 1963 – Exponential values for the species-area relation – *Science*, 141: 1276.
- Kilburn P. D. 1966 – Analysis of the species-area relation – *Ecology*, 47: 831–843.
- Kylin H. 1926 – Über Begriffsbildung und Statistik in der Pflanzensoziologie – *Bot. Not.* 1926: 81–180.

- LaBarbera M. 1985 – The Evolution and ecology of body size (In: Patterns and processes in the History of Life Eds. D. M. Raup, D. Jabłoński) – Springer (Berlin), pp. 69–98.
- Leitner W. A., Rosenzweig M. L. 1997 – Nested species–area curves and stochastic sampling: a new theory – *Oikos*, 79: 503–512.
- MacArthur R. H., Wilson E. O. 1967 – The theory of island biogeography – Princeton Univ. Press, 203 pp.
- Martin T. E. 1981 – Species–area slopes and coefficients: a caution on their interpretation – *Am. Nat.* 118: 823–837.
- May R. M. 1975 – Patterns of species abundance and diversity (In: Ecology and evolution of communities Eds.: M. L. Cody, J. M. Diamond) – Belknap (Cambridge), pp. 81–120.
- McGuinness K. A. 1984 – Equations and explanations in the study of species–area curves – *Biol. Rev.* 59: 423–440.
- Ney-Nifle M., Mangel M. 1999 – Species–area curves based on geographic range and occupancy – *J. Theor. Biol.* 196: 327–342.
- Novotny V., Kindlmann P. 1996 – Distribution of body sizes in arthropod taxa and communities – *Oikos*, 75: 75–82.
- Palmer M. W., White P. S. 1994 – Scale dependence and the species–area relationship – *Am. Nat.* 144: 717–740.
- Pielou E. C. 1977 – Mathematical Ecology – John Wiley & Sons, New York, 385 pp.
- Preston F. W. 1962 – The canonical distribution of commonness and rarity. Part I and II – *Ecology*, 43: 185–215, 410–432.
- Ricker W. E. 1973 – Linear regression in fishery research – *J. Fish. Res. Board Can.* 30: 409–434.
- Rosenzweig M. L. 1999 – Heeding the warning in biodiversity's basic law – *Science*, 284: 276–277.
- Schoener T. W. 1976 – The species–area relation within archipelagos: models and evidence from island land birds (In: Proc. 16th Int. Ornith. Congr. 1976 Eds. H. J. Frith, J. H. Calaby) – Canberra, pp. 629–642.
- Seim E., Saether B. E. 1983 – On rethinking allometry: which regression model to use – *L. Theor. Biol.* 104: 161–168.
- Sugihara G. 1980 – Minimal community structure: an explanation of species abundance patterns – *Am. Nat.* 116: 770–787.
- 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.
- Ulrich W. 1999a – The numbers of species of Hymenoptera in Europe and assessment of the total number of Hymenoptera in the world – *Pol. J. Entomol.* 68: 151–164.
- Ulrich W. 1999b – Estimating species numbers by extrapolation I: Comparing the performance of various estimators using large model communities – *Pol. J. Ecol.* 47: 271–291.
- Ulrich W. 1999c – Abundance, biomass and density boundaries in the Hymenoptera: analysis of the abundance–body size relationship and differences between forest and open landscape habitats – *Pol. J. Ecol.* 47: 73–86.
- Ulrich W. 2000a – On species–area relationships II: Parameter values of the power and the exponential model and their dependence on sampling method, habitat and assemblage structure – *Pol. J. Ecol.* 48: 21–35.
- Ulrich W. 2000b – On species–area relationships III: The intercept of the power function and the exponential model – *Pol. J. Ecol.* 48: 37–48.
- Williams C. B. 1943 – Area and the number of species – *Nature*, 152: 264–267.
- Williams M. R. 1995 – An extreme – value function model of the species incidence and species – area relations – *Ecology*, 76: 2607–2616.
- Williamson M. 1988 – Relationship of species number to area, distance and other variables (In: Analytical biogeography, Eds. A. A. Myers, P. S. Giller) – Chapman & Hall (New York), pp. 91–115.
- Wilson W. B., Gitay H., Steel J. B., King W. McG. 1998 – Relative abundance distributions in plant communities: effects of species richness and spatial scale – *J. Veg. Sci.* 9: 213–220.
- Wissel Ch., Maier B. 1992 – A stochastic model for the species–area relationship – *J. Biogeogr.* 19: 355–362.
- Wright S. J. 1988 – Patterns of abundance and the form of the species–area relation – *Am. Nat.* 131: 401–411.

(Received after revising October 1999)