Journal of Biogeography (J. Biogeogr.) (2010) 37, 1698–1705

ORIGINAL ARTICLE

Effects of sampling protocol on the shapes of species richness curves

Jürgen Dengler* and Jens Oldeland

Biodiversity, Evolution and Ecology of Plants, Biocentre Klein Flottbek and Botanical Garden, University of Hamburg, Hamburg, Germany

ABSTRACT

Aim Scheiner (*Journal of Biogeography*, 2009, **36**, 2005–2008) criticized several issues regarding the typology and analysis of species richness curves that were brought forward by Dengler (*Journal of Biogeography*, 2009, **36**, 728–744). In order to test these two sets of views in greater detail, we used a simulation model of ecological communities to demonstrate the effects of different sampling schemes on the shapes of species richness curves and their extrapolation capability.

Methods We simulated five random communities with 100 species on a 64×64 grid using random fields. Then we sampled species–area relationships (SARs, contiguous plots) as well as species–sampling relationships (SSRs, non-contiguous plots) from these communities, both for the full extent and the central quarter of the grid. Finally, we fitted different functions (power, quadratic power, logarithmic, Michaelis–Menten, Lomolino) to the obtained data and assessed their goodness-of-fit (Akaike weights) and their extrapolation capability (deviation of the predicted value from the true value).

Results We found that power functions gave the best fit for SARs, while for SSRs saturation functions performed better. Curves constructed from data of 32^2 grid cells gave reasonable extrapolations for 64^2 grid cells for SARs, irrespective of whether samples were gathered from the full extent or the centre only. By contrast, SSRs worked well for extrapolation only in the latter case.

Main conclusions SARs and SSRs have fundamentally different curve shapes. Both sampling strategies can be used for extrapolation of species richness to a target area, but only SARs allow for extrapolation to a larger area than that sampled. These results confirm a fundamental difference between SARs and area-based SSRs and thus support their typological differentiation.

Keywords

Curve fitting, goodness-of-fit, macroecology, power function, random field, rarefaction, saturation function, simulation, species–area relationship, species–sampling relationship.

*Correspondence: Jürgen Dengler, Biodiversity, Evolution and Ecology of Plants, Biocentre Klein Flottbek and Botanical Garden, University of Hamburg, Ohnhorststrasse 18, 22609 Hamburg, Germany.
E-mail: dengler@botanik.uni-hamburg.de

INTRODUCTION

Scheiner (2003) and Dengler (2009) proposed contrasting typologies for the classification of species richness curves. Based on different mathematical properties of the curves, Dengler (2009) suggested that the most fundamental distinction should be made between species–area relationships (SARs in the strict sense, constructed from data points, each of which corresponds to a contiguous area) and species–sampling relationships (SSRs). By contrast, Scheiner (2009) denies such a fundamental difference in curve shapes. He therefore rejects Dengler's typology and subordinates area-based SSRs to SARs. Moreover, he claims that SSRs are more suitable for extrapolation of species richness than SARs, while Dengler (2009) put forward the opposite argument.

In this study, we use a simulation approach in order to analyse whether the predictions of Dengler (2009) or those of Scheiner (2009) are met. For this purpose, we applied both sampling strategies, SARs and SSRs, to a virtual landscape in which species were distributed according to realistic rules, and tested the resulting curves for statistical differences. In particular, we aimed to address two questions raised by

www.blackwellpublishing.com/jbi
doi:10.1111/j.1365-2699.2010.02322.x

Journal of Biogeography

Dengler (2009) and Scheiner (2009). (1) Are there consistent differences in the shapes of SARs and SSRs? (2) Which of these two curve types is more suitable for extrapolation?

METHODS

Simulation of species richness patterns

To demonstrate and analyse the effect of different sampling schemes on curve shapes and extrapolation capability, we simulated five ecological communities with a total richness of 100 species each on a 64×64 grid (4096 grid cells). We randomly created these communities, assuming only two simple, yet realistic, conditions held constant for all random runs: (1) the frequency of species follows an exponentially decreasing distribution, and (2) each species inhabits one or several coherent patches.

The distribution pattern of each species was generated randomly and independently of the other with the freely available package 'RandomFields' (Schlather, 2001) for the statistical software R (R Development Core Team, 2008). We used a Gaussian spherical model (settings: mean = 25; variance = 5; scale = 10; alpha = 2) to reflect that species normally inhabit coherent patches. In order to simulate a typical frequency distribution of species (i.e. few frequent and many rare species), we fixed the number of grid cells (n_i) inhabited by each of the species numbered from $i = 1$ to $i = 100$ with an exponential function [n_i = INTEGER $(4096 \cdot (i/100)^{3.17} + 0.999)$]. As 'RandomFields' produces continuous values around the set mean, we assigned presences to the fields with the n_i highest values among all 4096 grid cells. Appendix S1 in Supporting Information shows two typical distribution patterns of species of different frequency (c, d) and the resulting overall richness pattern at a grain of one grid cell for two simulated communities (a, b). The distributions patterns of species and richness that resulted from the initial settings corresponded well to our intuitively assumed picture of real communities. Moreover, slight modifications (e.g. scale = 5 or scale = 20) did not alter the results qualitatively. Thus, we refrained from varying the community model systematically in order to comprehensively assess its effect on curve properties because this would have exceeded the framework of this short contribution.

For each species, we generated rasters in the ASCII format, which were then imported to a geographic information system (ArcGIS 9.3; ESRI, 2008). Richness values of different area sizes were calculated in two different ways to reflect the two different sampling approaches in question, both for the full extent of 4096 grid cells (further referred to as 'Full') and for the central square of 1024 grid cells (further referred to as 'Centre'). The sizes of the analysed areas were 1, 4, 16, 64, 256, 1024 and, in the case of full extent, also 4096 grid cells.

The first sampling approach corresponds to curve type IIa of Scheiner (2003) and to SAR subtype A.b.i of Dengler (2009). This approach, hereafter referred to as 'SAR' is exemplified in the left part of Fig. 1. It was carried out with the 'BlockStatistic' function of the 'spatial analyst' extension for ArcGIS using quadratic cells of sizes 1, 4, 16, 64, 256, 1024 and 4096, always covering the entire area. The statistic option was set to 'MAXIMUM', which yields the maximum value in the sampled quadrat. This means that the returned value was 1 if the species occurred in one or several of the cells included (i.e. had a value of 1 there). This function was repeated for all 100 species rasters using a batch version of the 'BlockStatistic' feature. These hundred block-statistic rasters were summarized using the 'CellStatistic' feature of ArcGIS with the statistic option 'SUM'. For an area size of one grid cell, the resulting diversity pattern is exactly the richness distribution of the whole statistical population of 4096 grid cells as shown in Appendices S1(a) and (b).

The second sampling approach corresponds to curve type IIb of Scheiner (2003) and to SSR subtype B.b of Dengler (2009). This approach, further referred to as 'SSR', is exemplified in the right part of Fig. 1. This random sampling using non-contiguous quadrats was simulated by generating shape files representing five random sets for each area size. For example, for an area size of four grid cells, four cells were randomly selected from the 1×1 grid (see Fig. 1d). Using the batch version of the feature 'ZonalStatistic' in the Spatial Analyst extension, it was calculated whether at least one of the four selected cells captured an occurrence of species *i*. If yes, all other cells, whether empty or not, were assigned a value of 1, meaning presence. The hundred 'ZonalStatistic' rasters were summarized using 'CellStatistic' with the statistic option 'SUM'. For each area size, richness values for five random draws were calculated and used for further statistical analysis. We restricted ourselves to five random draws because the construction of a rarefaction curve based on all possible combinations of one, four, sixteen cells and so forth would have been extremely time-consuming due to the large number of cells with richness counts. Nevertheless, five draws already allowed us to establish unbiased mean values with only small standard errors (see Results).

Species richness models

For the statistical analyses of the richness data of the four sampling schemes (SAR full, SSR full, SAR centre, SSR centre), we used the mean richness values for the different areas. This means that in the case of the full extent we fitted curves to seven pairs of area–richness data and in the case of the reduced extent ('centre') six pairs of area–richness data. We applied the nonlinear regression module (user-defined regression, loss function) of STATISTICA 8.0 (StatSoft, Inc., 2008) to fit different functional models to these data. We used the default settings [loss function = $(\text{OBS} - \text{PRED})^2$; estimation method = quasi-Newton; convergence criterion = 0.0001; step width for all parameters = 0.5; starting values for all parameters = 0.1]. In those rare cases when the algorithm did not converge with these uniform starting values we changed them individually to approximate values we expected for the respective parameters, which always led to convergence.

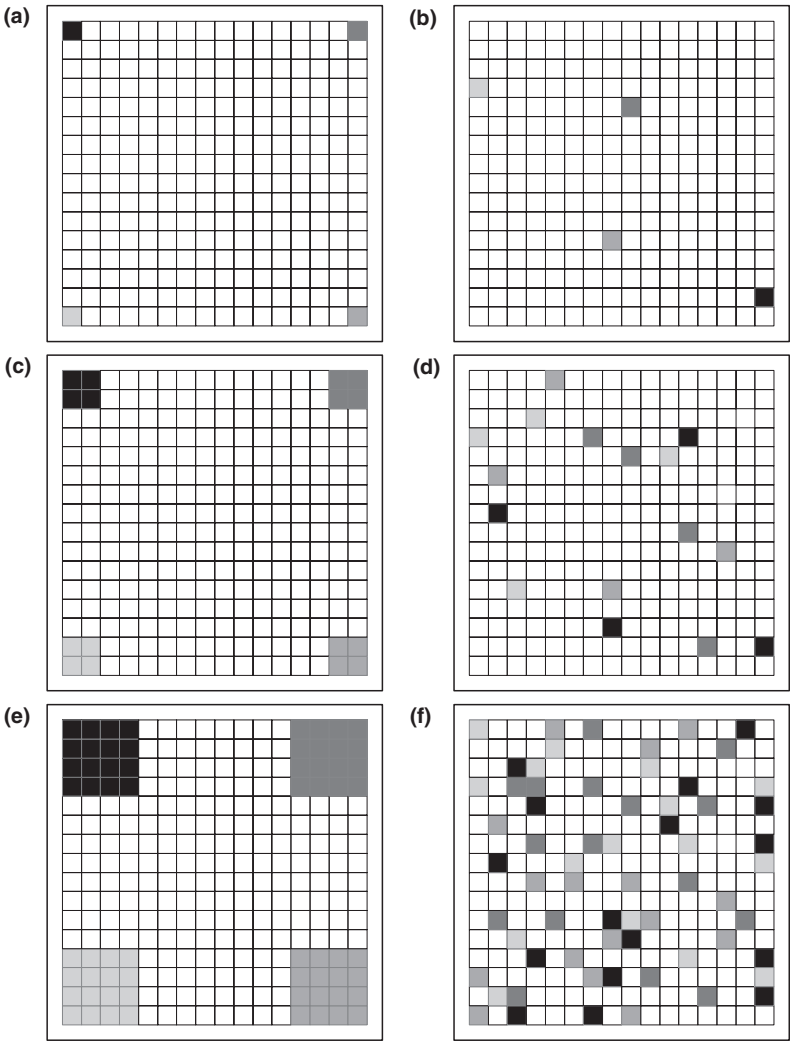


Figure 1 Spatial arrangement of sample units in the case of species–area relationships (SARs: a, c, e) and species–sampling relationships (SSRs: b, d, f), illustrated for a 16×16 grid. Four examples corresponding to areas of 1, 4 and 16 grid cells are shown in each case. In each of the six partial figures, the grids that belong to one 'plot' are marked by the same tone. For each of the 'plot' sizes, first the total species richness for the combined area of all cells of a certain 'tone' was determined, and then these richness counts were averaged over all 'tones'.

We applied the same set of five functional models as previously used by Dengler & Boch (2008). These models (Table 1) were chosen from the multitude of possible functions (see Tjørve, 2003, 2009; Dengler, 2009) because they represent a wide range of different curve properties (saturation versus unlimited; sigmoid versus non-sigmoid; different numbers of parameters), while at the same time having been shown to be particularly suitable within a certain category of functions (e.g. Stiles & Scheiner, 2007; Dengler, 2009). Following Dengler (2008, 2009), the curve fitting for each of the functions was done both with species richness (S) and with the logarithm of species richness ($\log S$) as dependent variables.

Finally, the performance of the different models was compared separately for each of the five random communities

Table 1 The five function types used to model the species–area relationships in this study and their characteristics (for more details see Dengler, 2009).

Curve name	Model	General shape	No. parameters	Upper asymptote
Power	$S = b_0 A^{b_1}$	Unbound, convex	2	No
Power, quadratic	$S = 10^{[b_0 + b_1 \log A + b_2 (\log A)^2]}$	U-shaped ($b_2 > 0$) or inverse U-shaped ($b_2 < 0$)	3	No
Logarithm*	$S = b_0 + b_1 \log A$	Unbound, convex	2	No
Michaelis–Menten	$S = b_0 A/(b_1 + A)$	Saturation, convex	2	Yes (b_0)
Lomolino	$S = b_0/(1 + b_1^{\log(b_2/A)})$	Saturation, sigmoid	3	Yes (b_0)

S = species richness; A = area; b_0, b_1, b_2 = fitted parameters; log = logarithm of a certain base (here $\log_{10}$ is used). In the case of the power function, b_0 and b_1 are often termed c and z. Note that for the U-shaped or inverse U-shaped quadratic power functions, the point beyond which richness would decrease with increasing area normally lies far outside the fitted range of areas.

*The logarithmic function is often erroneously termed the 'exponential function'.

with the Akaike information criterion for small n (AIC_c; Burnham & Anderson, 2002). Here, we compared the five models fitted in S-space with AIC_c calculated for S-space and then separately, the five models fitted in log S-space with AIC_c calculated for log S-space. To allow averaging of the results for different communities and comparison between different sampling strategies, we transformed the AIC_c values to Akaike weights (w_i), which correspond to the probability of model i being the best among the models compared (Burnham & Anderson, 2002; for a previous application to SARs, see Dengler, 2010). In Appendix S2, we additionally provide mean R^2 values, which may be more widely understood than Akaike weights.

We calculated the extrapolation capability of the different function types in the different situations by fitting the functions to the data without the values for 4096 grid cells and comparing the predicted value for 4096 grid cells with the known value at that scale (100 species). Following Walther & Moore (2005), extrapolation capability can be separated into the three concepts of bias, precision and accuracy. Thus, in Table 2, we present the bias of the extrapolation as mean predicted value (to be related to the true value of 100), the precision as standard error of the predicted mean and the accuracy as the mean absolute value of the log error of extrapolation (LEE; see Dengler, 2009).

RESULTS

The richness curves obtained with SAR and SSR sampling, respectively, differed greatly and consistently (Fig. 2). Apart from the endpoints, where both sampling schemes necessarily had the same values, the species–sampling curves showed consistently higher values than the species–area curves. While the two SARs did not give any indication of having an upper

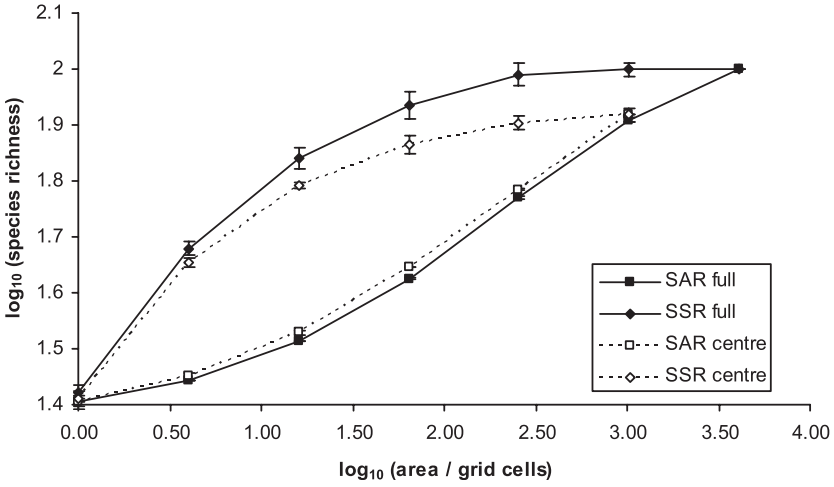


Figure 2 Dependence of species richness on area for four sampling schemes in simulated communities of 100 species occurring in a 4096 grid-cell field. Mean values for five random communities and their standard errors are indicated, both calculated on a log-scale. Note that in the case of species–area relationships (SARs), the standard errors are so minimal that the error bars are hardly visible.

Table 2 Goodness-of-fit expressed as mean Akaike weights (w_i) ± standard deviations for five functions fitted to species–area relationships (SARs) and species–sampling relationships (SSRs) constructed from data of five simulated ecological communities. The models with a support of 20% or more on average are highlighted in bold face. All functions were constructed both for the 64 × 64 grid field (Full) and for its central 32 × 32 square (Centre), and fitted both for species richness (S) and for log S as dependent variables. For details of the functions, see Table 1.

			Power	Power, quadr.	Logarithm	Michaelis–Menten	Lomolino
p			2	3	2	2	3
Saturation	Extent	S-space	No	No	No	Yes	Yes
SAR	Full	S	**92.8 ± 1.8**	5.0 ± 2.0	0.1 ± 0.1	0.0 ± 0.0	2.1 ± 0.2
SAR	Full	log S	**49.1 ± 14.5**	**49.7 ± 14.9**	0.3 ± 0.1	0.0 ± 0.0	0.9 ± 0.3
SAR	Centre	S	**26.1 ± 41.3**	**73.4 ± 42.1**	0.1 ± 0.1	0.0 ± 0.0	0.3 ± 0.6
SAR	Centre	log S	**20.4 ± 43.2**	**79.4 ± 43.7**	0.1 ± 0.1	0.0 ± 0.0	0.2 ± 0.3
SSR	Full	S	0.0 ± 0.0	3.9 ± 5.5	0.0 ± 0.0	0.2 ± 0.1	**95.9 ± 5.6**
SSR	Full	log S	0.0 ± 0.0	9.7 ± 21.5	0.0 ± 0.0	0.0 ± 0.0	**90.2 ± 21.5**
SSR	Centre	S	0.5 ± 1.1	11.8 ± 25.0	3.3 ± 6.5	13.8 ± 27.6	**70.6 ± 40.6**
SSR	Centre	log S	0.1 ± 0.2	3.8 ± 7.9	2.0 ± 2.7	14.3 ± 28.5	**79.8 ± 31.6**

Values are given as percentage.

The row p denotes the number of fitted parameters per function and the row Saturation indicates whether for large areas the function approaches a maximum richness value asymptotically. Note that the number of estimated parameters (k) used for the calculation of the information criteria, is $p + 1$ (i.e. the model parameters + the variance of the residuals; see Burnham & Anderson, 2002).

Table 3 Extrapolation capability of the species richness models for different sampling schemes, constructed from data of 1 to 1024 grid cells. Both species–area relationships (SARs) and rarefaction curves of species–sampling relationships (SSRs) were analysed, with the 1024 grid cells sampled either distributed over the whole extent of the 64 × 64 grid (Full) or restricted to the central 32 × 32 square (Centre). Five model functions were fitted in species richness (*S*)-space and in log *S*-space. For each sampling scheme, the predicted value (mean ± standard error) for 4096 grid cells is given, based on five simulated communities with a true richness of 100 species at that scale. Further, the accuracy of the estimate is presented as the mean of the absolute values of the log error of extrapolation (|LEE|), with the most accurate extrapolations (|LEE| ≤ 0.05) highlighted in bold face.

Sampling design	SAR				SSR			
Extent	Full		Centre		Full		Centre	
Parameter	Predicted	|LEE|	Predicted	|LEE|	Predicted	|LEE|	Predicted	|LEE|
Power (*S*)	103.3 ± 1.1	**0.01**	106.5 ± 4.9	**0.04**	134.1 ± 0.3	0.13	107.8 ± 3.7	**0.04**
Power (log *S*)	93.8 ± 0.8	**0.03**	98.6 ± 5.4	**0.04**	159.1 ± 1.3	0.20	124.4 ± 5.0	0.09
Power, quadratic (*S*)	122.8 ± 1.8	0.09	122.3 ± 3.5	0.09	88.0 ± 8.5	0.06	72.7 ± 2.2	0.14
Power, quadratic (log *S*)	127.7 ± 1.8	0.11	127.3 ± 3.4	0.10	84.3 ± 1.6	0.07	68.8 ± 2.4	0.16
Logarithmic (*S*)	82.8 ± 1.6	0.08	85.8 ± 3.5	0.07	124.4 ± 6.2	0.10	101.9 ± 3.2	**0.02**
Logarithmic (log *S*)	73.4 ± 1.4	0.13	76.7 ± 3.7	0.12	133.0 ± 1.9	0.12	109.5 ± 3.5	**0.04**
Michaelis–Menten (*S*)	64.9 ± 4.8	0.19	67.5 ± 3.1	0.17	95.7 ± 0.1	**0.02**	79.2 ± 2.3	0.10
Michaelis–Menten (log *S*)	50.5 ± 9.2	0.30	52.8 ± 2.5	0.28	91.3 ± 7.4	**0.04**	76.8 ± 2.3	0.12
Lomolino (*S*)	102.2 ± 1.1	**0.01**	104.9 ± 3.4	**0.03**	102.7 ± 9.4	**0.01**	84.0 ± 2.7	0.08
Lomolino (log *S*)	92.3 ± 6.8	**0.03**	95.5 ± 5.8	**0.05**	102.6 ± 5.8	**0.01**	83.7 ± 2.8	0.08

limit, the two SSRs turned nearly horizontal (even on double-log representation) to the right-hand side of the curves (Fig. 2). The SARs for 'Full' and 'Centre' nearly perfectly overlapped (Fig. 2). By contrast, the SSRs for full extent and 'Centre' shared a similar curvature, but the values for 'Centre' increasingly fell below those of 'Full' towards larger areas (Fig. 2).

For the full extent (4096 grid cells) assessed in *S*-space (i.e. with *S* as a response variable), the unbound normal power function was by far the best model for SAR data in all five communities, while the Lomolino function (a saturation function) was superior throughout for the SSR data (Table 2). The prevalence of unbound functions for SAR data and saturation functions for SSR data essentially remained the same when assessing the full-extent data in log *S*-space as well as for the 'Centre' data (Table 2). However, apart from the normal power function (for SARs) and the Lomolino function (for SSRs), in some cases the quadratic power function and the Michaelis–Menten function, respectively, were rated best according to AIC_c (detailed results not shown; see Table 2 for summary). The logarithmic function did not work well for either SARs or for SSRs (Table 2).

The most accurate extrapolations for SARs were achieved with the power function and the Lomolino function, while the quadratic power function, logarithm and Michaelis–Menten functions proved to be unsuitable for this sampling scheme (Table 3), the former predicting values that were too high and the latter two predicting values that were far too low. When the extrapolation was made from data from the central square only, the overall results were very similar to those of the full-extent data, and the extrapolation accuracy of the power and Lomolino functions were only slightly lower (Table 3).

In the case of SSRs, the two saturation functions, Lomolino and Michaelis–Menten, were best in extrapolation for the full-extent data (Table 3). However, the same functions performed very poorly when applied to the 'Centre' data, resulting in predicted values that lie approximately 20% below the true value for 4096 grid cells.

DISCUSSION

Differences between SARs and SSRs

Our simulation showed that SARs and area-based SSRs for the same geographical area studied have fundamentally different shapes, the first well approximated by power functions and the latter by saturation functions. The shape of the area-based SSRs was very similar to curves typically obtained by individual- or sample-based rarefaction models (e.g. Gotelli & Colwell, 2001). According to our results, both SARs and SSRs are suitable for extrapolating species richness within the area in which the sample units were spread, but only SARs allow extrapolation beyond that target area. Here again, area-based SSRs and individual- and sample-based rarefaction curves show similarities. This is apparent in that rarefaction curves predict richness within the study area (i.e. the 'universe' in which a uniform sampling process was carried out; Colwell & Coddington, 1994), but not beyond it. These findings support the view of Dengler (2009) that the fundamental division within species richness curves, both regarding curve shapes and potential applications, does not depend on whether the *x*-axis is related to 'area' (as argued by Scheiner, 2009) but on the spatial coherency of the individual grains.

Although being 'only' a typology, the terminological proposal of Dengler (2009) can help avoid misunderstandings and misapplications as they widely occur in biodiversity research (see review by Dengler, 2008). For example, the findings of

Stiles & Scheiner (2007) that their curves of type IIIB *sensu* Scheiner (2003) were generally best fitted by saturation functions (such as the Lomolino function), while their only type IV curve was much better fitted by unbound functions, do not need any ecological explanation (as brought forward by Stiles & Scheiner, 2007, and Scheiner, 2009). Instead, they could easily be (and as we understand it: most probably are) solely the consequence of the different sampling schemes (i.e. the type IIIB curves are SSRs *sensu* Dengler, 2009).

It should be stressed that the fundamental differences between true SARs and area-based SSRs regarding curve shape and extrapolation capability as demonstrated with our simulation are independent of the specific properties of the model applied. SARs and SSRs would only then be identical when all species were distributed completely randomly (Ney-Nifle & Mangel, 1999), a situation not found in nature. As soon as any kind of spatial autocorrelation (distance decay) is added to the species distribution model, SARs and SSRs will always differ systematically (see Chiarucci *et al.*, 2009). In other words, each of Scheiner's curve types IIB, IIIA and IIIB comprises a wide range of non-comparable sampling schemes (i.e. with different distances between the sample units), a fact that is extensively discussed by Chiarucci *et al.* (2009). Such curves only could provide biologically meaningful information if the spacing between the individual sampling units was standardized for all studies compared (see Dengler, 2008; Chiarucci *et al.*, 2009). Accordingly, Chiarucci *et al.* (2009) agree with us that 'SARs' built from non-contiguous plots are equivalent to sample-based rarefaction curves.

Methods for extrapolating species richness

Scheiner (2009) asked how the best possible estimate of species richness could be attained for an area of 1000 km^2 when resources are sufficient only to sample 1000 1-ha plots. Naturally, it is not useful to lump all these 1000 plots together somewhere within a single portion of the study area.

The first reasonable approach is random (or systematic) placement of the plots within the 1000-km^2 area (as proposed by Scheiner, 2009), followed by the construction of rarefaction curves to which a suitable type of saturation function is fitted. However, in contrast to Scheiner's claims, this SSR approach is not the only suitable approach.

As shown with the simulation, SARs offer another viable means of estimating the total richness of the 1000-km^2 area. A reasonable SAR approach would be to sample nested-plot SARs from 1 m^2 up to 100 ha in 10 different locations randomly (or systematically) placed within the whole study area. These 10 SARs would be averaged and used for an extrapolation to 1000 km^2, using the most suitable function type according to the LEE (see Dengler, 2009). Alternatively, each of the 10 curves may be used individually for extrapolation, giving a confidence interval for the probable richness value at 1000 km^2. As in the case of SSRs, a total of 1% of the study would be sampled, though with a different spatial arrangement.

A third method is to use the randomly distributed 1-ha plots of the SSR approach and to apply richness estimators with the species composition data obtained (for overviews of this approach see Magurran, 2004, and Walther & Moore, 2005). In this study, we compared only the first two methods because we wanted to address the differences between SSRs and SARs and because most richness estimators require abundance data while our simulation only generated incidence data. Theoretically, all three approaches (SSRs, SARs, richness estimators) should lead to reliable results so long as only the richness value for a single target area is required. However, to our knowledge these alternative approaches have never been compared in a single study. Beyond accepting, on theoretical grounds, the basic suitability of SSRs and SARs when appropriately applied, more extensive simulation and field studies would also be valuable to determine which approach is preferable in certain situations regarding accuracy, precision, bias and cost-effectiveness.

Validity of the simulation approach

Despite being rather simple, our simulation approach yielded distribution patterns that approximate patterns often observed in nature (see Appendix S1). The predominantly coherent patches inhabited by individual species simulated with Gaussian spherical fields reflect both biotic processes (prevalence of short-distance dispersal, clonal growth) and the patchiness of the abiotic environment. The predominance of the power model (over the logarithmic model and the saturation models) corresponds to the situation in many real communities, and the determined z-values of 0.17–0.20 (see b_1 in Appendix S3) lie well within the range of typical plant communities (e.g. Dolnik, 2003; Dengler & Boch, 2008; Dengler, 2009).

However, the graphic representation (Fig. 2) still shows small but consistent deviations from a 'pure' power function (i.e. one with a constant z-value) in as far as the z-values were 'triphasic', low–high–low. The reason for this deviation could lie in an over-simplistic representation of species distribution patterns in our simulations. First, we did not include environmental gradients over the full study extent but only small-scale, isotropic heterogeneity. The inclusion of large-scale environmental gradients would be likely to increase the z-values at the right-hand side of the graph because of the addition of new species at large scales. Second, we used identical scale parameters (see Methods) for all species, while varying scale parameters would probably better reflect the differences in mean sizes of species in communities. In consequence, a model reflecting the situation in nature more closely than our approach would most probably yield curves that are even closer to 'pure' power functions.

CONCLUSIONS AND OUTLOOK

The outcome of our simulation study complies with the theoretical predictions that prompted Dengler (2009) to

propose a new typology of species richness curves. The two major aspects of these findings can be summarized as follows.

1. SARs and area-based SSRs *sensu* Dengler (2009) have fundamentally different curve shapes. SSRs, irrespective of whether they are constructed from areas (as here) or from plotless sampling or individuals (e.g. Gotelli & Colwell, 2001), can be fitted very well by different saturation functions, while SARs cannot.

2. SARs can be used to extrapolate to any contiguous area bigger than the largest one sampled (of course, the accuracy will decrease with increasing size difference). SSRs can equally well be used to extrapolate to a single specific area for which they are sampled (i.e. the area in which the sample-units were spread: the 'target area'). However, SSRs fail to provide the foundation to extrapolate beyond the target area because they will predict practically the same richness value for all larger areas.

Thus, we are able to refute Scheiner's (2009) objections both theoretically and empirically. By contrast, our results are fully in line with his statement that 'different sampling schemes serve different purposes, and a variety of functional relationships are likely to hold'. Based on the findings presented above we can confirm the work of Dengler (2009) in as far as the sampling scheme is the basic determinant of the resulting curve shape. We are therefore convinced that the typology of Dengler (2009), which distinguishes curve types based on their purposes and their fundamental shapes, is appropriate and useful.

While there is a legacy of simulation-based analyses of diversity patterns (e.g. Ney-Nifle & Mangel, 1999; McGuiness, 2000; He & Legendre, 2002; Rosindell & Cornell, 2007; Gardner & Engelhardt, 2008), we are not aware of any other simulation study that tested the effect of SAR versus SSR sampling on extrapolation accuracy. Actually, the vast majority of simulation studies considered only true SARs based on nested squares (or rarely rectangles) of grid cells. Here, a more sophisticated simulation model than ours could be an excellent tool with which to study the effects of different parameter settings (corresponding to different processes) on the curve shapes both of SARs and SSRs. Such a model could produce predictions against which observed patterns could be tested (see Gotelli *et al.*, 2009).

ACKNOWLEDGEMENTS

We thank Samuel M. Scheiner for providing – with his correspondence – the opportunity for this article, Robert J. Whittaker, Nicholas Gotelli and two anonymous referees for their helpful comments on former versions of this article and Curtis Björk for polishing our English. We conducted this research within the framework of BIOTA Southern Africa, sponsored by the German Federal Ministry of Education and Research (BMBF) under promotion number 01 LC 0024A.

REFERENCES

Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference – a practical information-theoretic approach*, 3rd edn. Springer, New York.

Chiarucci, A., Bacaro, G., Rocchini, D., Ricotta, C., Palmer, M.W. & Scheiner, S.M. (2009) Spatially constrained rarefaction: incorporating the autocorrelated structure of biological communities into sample-based rarefaction. *Community Ecology*, **10,** 209–214.

Colwell, R.K. & Coddington, J.A. (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **345,** 101–118.

Dengler, J. (2008) Pitfalls in small-scale species–area sampling and analysis. *Folia Geobotanica*, **43,** 269–287.

Dengler, J. (2009) Which function describes the species–area relationship best? A review and empirical evaluation. *Journal of Biogeography*, **36,** 728–744.

Dengler, J. (2010) Robust methods for detecting a small island effect. *Diversity and Distributions*, **16,** 256–266.

Dengler, J. & Boch, S. (2008) Sampling-design effects on properties of species–area curves – a case study from Estonian dry grassland communities. *Folia Geobotanica*, **43,** 289–304.

Dolnik, C. (2003) Artenzahl-Areal-Beziehungen von Wald- und Offenlandgesellschaften – Ein Beitrag zur Erfassung der botanischen Artenvielfalt unter besonderer Berücksichtigung der Flechten und Moose am Beispiel des Nationalparks Kurischen Nehrung (Russland). *Mitteilungen der Arbeitsgemeinschaft Geobotanik in Schleswig-Holstein und Hamburg*, **62,** 1–183.

ESRI (2008) *ArcGIS release 9.3*. Environmental Systems Research Institute, Redlands, CA (http://www.esri.com).

Gardner, R.H. & Engelhardt, K.A.M. (2008) Spatial processes that maintain biodiversity in plant communities. *Perspectives in Plant Ecology, Evolution and Systematics*, **9,** 211–228.

Gotelli, N.J. & Colwell, R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4,** 379–391.

Gotelli, N.J., Anderson, M.J., Arita, H.T. *et al.* (2009) Patterns and causes of species richness: a general simulation model for macroecology. *Ecology Letters*, **12,** 873–886.

He, F. & Legendre, P. (2002) Species diversity patterns derived from species–area models. *Ecology*, **83,** 1185–1198.

Magurran, A.E. (2004) *Measuring biological diversity*. Blackwell, Malden, MA.

McGuiness, K.A. (2000) Distinguishing area and habitat heterogeneity effects: a simulation test of the MacNally and Watson (1997) protocol. *Austral Ecology*, **25,** 8–15.

Ney-Nifle, M. & Mangel, M. (1999) Species–area curves based on geographic range and occupancy. *Journal of Theoretical Biology*, **196,** 327–342.

R Development Core Team (2008) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna (http://www.R-project.org).

Rosindell, J. & Cornell, S.J. (2007) Species–area relationships from a spatially explicit neutral model in an infinitive landscape. *Ecology Letters*, **10,** 586–595.

Scheiner, S.M. (2003) Six types of species-area curves. *Global Ecology and Biogeography*, **12,** 441–447.

Scheiner, S.M. (2009) The terminology and use of species–area relationships: a response to Dengler (2009). *Journal of Biogeography*, **36,** 2005–2008.

Schlather, M. (2001) Simulation of stationary and isotropic random fields. *R-News*, **1,** 18–20.

StatSoft, Inc. (2008) *STATISTICA for Windows, version 8.0*. StatSoft, Inc., Tulsa, OK (http://www.statsoft.com).

Stiles, A. & Scheiner, S.M. (2007) Evaluation of species–area functions using Sonoran Desert plant data: not all species–area curves are power functions. *Oikos*, **116,** 1930–1940.

Tjørve, E. (2003) Shapes and functions of species–area curves: a review of possible models. *Journal of Biogeography*, **30,** 827–835.

Tjørve, E. (2009) Shapes and functions of species–area curves (II): a review of new models and parameterizations. *Journal of Biogeography*, **36,** 1435–1445.

Walther, B.A. & Moore, J.L. (2005) The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography*, **28,** 815–829.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Illustrations of examples from the simulated communities.

Appendix S2 Goodness-of-fit expressed as mean R^2 values for five functions fitted to species–area relationships (SARs) and species–sampling relationships (SSRs).

Appendix S3 Mean model parameters for five functions fitted to species–area relationships (SARs) and species–sampling relationships (SSRs).

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

BIOSKETCHES

Jürgen Dengler is a lecturer in botany and ecology at the Biocentre Klein Flottbek and Botanical Garden, University of Hamburg. He finds particular interest in diversity patterns and species–area relationships. He also conducts research in vegetation science, dealing with theoretical and methodological foundations, the establishment of large vegetation databanks, the development of consistent classifications, and is working toward applying science in furthering conservation. He focuses on European dry grasslands and ruderal communities as well as arid ecosystems in Africa.

Jens Oldeland is a PhD student in the BIOTA Southern Africa project, within the same working group as J.D. His major interests are remote-sensing and GIS applications in ecology, spatial and multivariate statistics, as well as habitat modelling. He focuses on arid ecosystems in Africa, mainly in Morocco and Namibia.

Author contributions: J.D. conceived the idea, conducted the statistical analyses, and composed the text. J.O. implemented the simulations and reviewed the text.

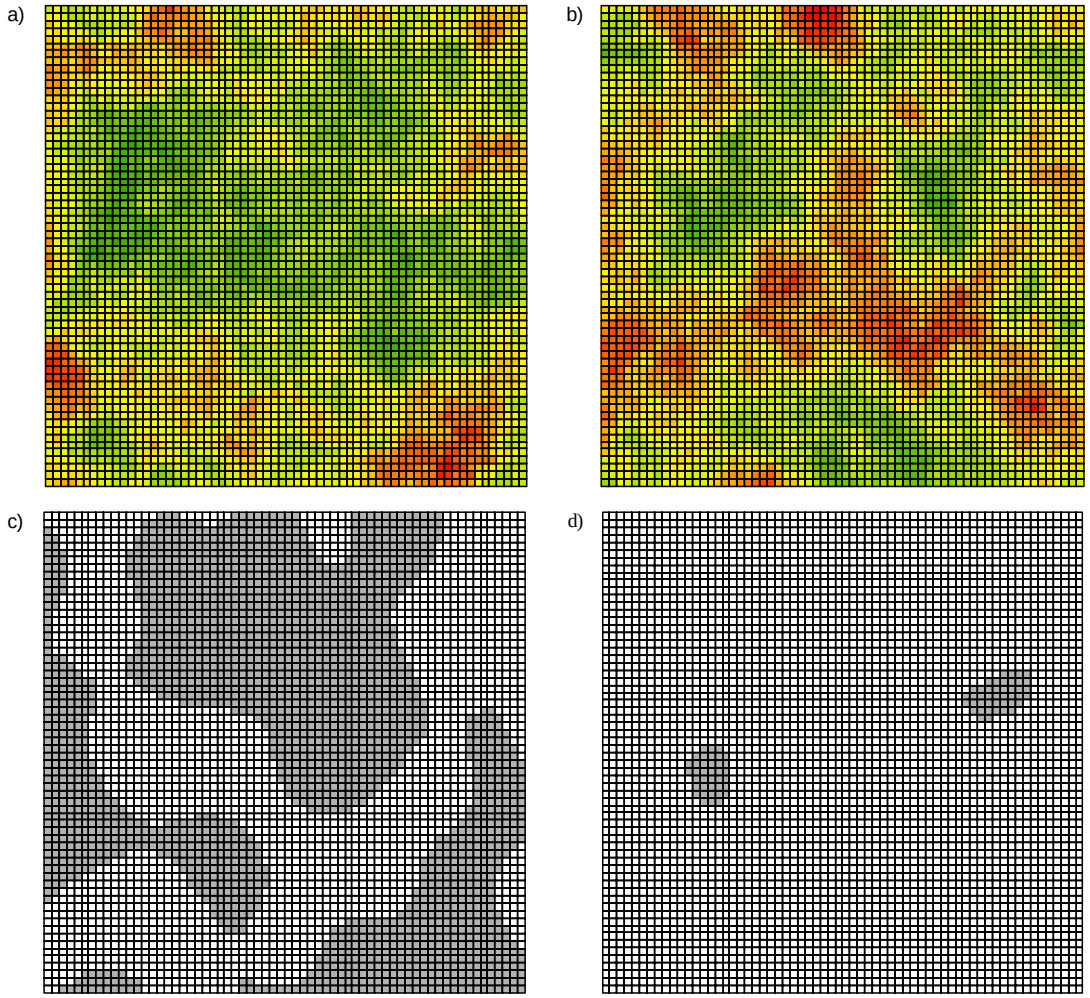
Editor: Nicholas Gotelli

SUPPORTING INFORMATION

Effects of sampling protocol on the shapes of species richness curves

by J. Dengler and J. Oldeland

APPENDIX S1 Examples of the simulated communities superimposed by the 64 × 64 grid. In the upper row (a, b), patterns of species richness at the one-grid cell level are presented for two of the simulated random communities (green = 14 species to red = 35 species). In the lower row, the simulated distribution patterns of a frequent species (species number *i* = 90: c) and a rare species (*i* = 20: d) are shown for one of these communities, with occupied cells being displayed in grey.



APPENDIX S2 Goodness-of-fit expressed as mean R^2 values for five functions fitted to species–area relationships (SARs) and species–sampling relationships (SSRs) constructed from data of five simulated ecological communities. All functions were constructed both for the 64 × 64 grid field (Full) and for its central 32 × 32 square (Centre), and fitted both for species richness (*S)* and for log *S* as dependent variables.

Sampling scheme	Extent	*S*-space	Power	Power, quadr.	Logarithm	Michaelis–Menten	Lomolino
SAR	Full	*S*	0.9885	0.9902	0.9156	0.6814	0.9879
SAR	Full	log *S*	0.9737	0.9905	0.8836	0.4833	0.9709
SAR	Centre	*S*	0.9796	0.9977	0.9008	0.6032	0.9781
SAR	Centre	log *S*	0.9649	0.9974	0.8864	0.5211	0.9584
SSR	Full	*S*	0.7943	0.9915	0.8890	0.9643	0.9977
SSR	Full	log *S*	0.8037	0.9917	0.9328	0.9593	0.9985
SSR	Centre	*S*	0.8434	0.9874	0.9216	0.9638	0.9954
SSR	Centre	log *S*	0.8421	0.9896	0.9442	0.9675	0.9972

APPENDIX S3 Mean model parameters for five functions fitted to species–area relationships (SARs) and species–sampling relationships (SSRs) constructed from data of five simulated ecological communities. For the meaning of b_0, b_1, and b_2, see Table 1. All functions were constructed both for the 64 × 64 grid field (Full) and for its central 32 × 32 square (Centre), and fitted both for species richness (*S)* and for log *S* as dependent variables.

Parameter	Sampling scheme	Extent	*S*-space	Power	Power, quadr.	Logarithm	Michaelis–Menten	Lomolino
b_0	SAR	Full	*S*	20.81	1.35	14.36	85.91	1082.45
	SAR	Full	log *S*	22.30	1.39	21.31	58.97	955.60
	SAR	Centre	*S*	21.03	1.39	17.67	67.64	1070.06
	SAR	Centre	log *S*	22.88	1.40	22.26	52.82	729.66
	SSR	Full	*S*	46.37	1.48	37.48	97.02	102.62
	SSR	Full	log *S*	36.99	1.44	29.39	93.32	102.82
	SSR	Centre	*S*	38.74	1.45	32.82	79.24	84.86
	SSR	Centre	log *S*	32.93	1.43	28.12	76.87	84.48
b_1	SAR	Full	*S*	0.1896	0.1524	21.16	42.57	1.59
	SAR	Full	log *S*	0.1763	0.0963	16.23	2.02	1.54
	SAR	Centre	*S*	0.1954	0.0944	18.87	10.56	1.62
	SAR	Centre	log *S*	0.1752	0.0749	15.06	1.52	1.56
	SSR	Full	*S*	0.1075	0.3588	20.94	4.28	4.59
	SSR	Full	log *S*	0.1495	0.4002	25.86	2.89	4.55
	SSR	Centre	*S*	0.1229	0.3507	19.12	2.88	4.61
	SSR	Centre	log *S*	0.1595	0.3882	22.54	2.19	4.74
b_2	SAR	Full	*S*		0.0081			3.9×10^8
	SAR	Full	log *S*		0.0221			6.3×10^8
	SAR	Centre	*S*		0.0273			4.2×10^8
	SAR	Centre	log *S*		0.0333			3.3×10^8
	SSR	Full	*S*		–0.0601			5.03
	SSR	Full	log *S*		–0.0694			5.06
	SSR	Centre	*S*		–0.0657			3.46
	SSR	Centre	log *S*		–0.0760			3.42