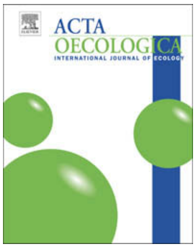
ACTA OECOLOGICA 35 (2009) 149–156

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/actoec



Original article

Would species richness estimators change the observed species area relationship?

Paulo A.V. Borges[a,*,1], Joaquín Hortal[a,b,1], Rosalina Gabriel[a], Nídia Homem[a]

[a]Departamento de Ciências Agrárias, CITAA (Azorean Biodiversity Group), Universidade dos Açores, Campus de Angra, 9701-851 Angra do Heroísmo, Portugal
[b]NERC Centre for Population Biology, Division of Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK

ARTICLE INFO

Article history:
Received 26 November 2007
Accepted 15 September 2008
Published online 4 November 2008

Keywords:
Species–area relationship (SAR)
Non-parametric species richness estimators
Standardized surveys
Azores
Epigean arthropods
Canopy arthropods
Bryophytes

ABSTRACT

We evaluate whether the description of the species area relationship (SAR) can be improved by using richness estimates instead of observed richness values. To do this, we use three independent datasets gathered with standardized survey methods from the native laurisilva forest of the Azorean archipelago, encompassing different distributional extent and biological groups: soil epigean arthropods at eight forest fragments in Terceira Island, canopy arthropods inhabiting *Juniperus brevifolia* at 16 forest fragments of six different islands, and bryophytes of seven forest fragments from Terceira and Pico islands. Species richness values were estimated for each forest fragment using seven non-parametric estimators (ACE, ICE, Chao1, Chao2, Jackknife1, Jackknife2 and Bootstrap; five in the case of bryophytes). These estimates were fitted to classical log–log species–area curves and the intercept, slope and goodness of fit of these curves were compared with those obtained from the observed species richness values to determine if significant differences appear in these parameters. We hypothesized that the intercepts would be higher in the estimated data sets compared with the observed data, as estimated richness values are typically higher than observed values. We found partial support for the hypothesis – intercepts of the SAR obtained from estimated richness values were significantly higher in the case of epigean arthropods and bryophyte datasets. In contrast, the slope and goodness of fit obtained with estimated values were not significantly different from those obtained from observed species richness in all groups, although a few small differences appeared. We conclude that, although little is gained using these estimators if data come from standardized surveys, their estimations could be used to analyze macroecological relationships with non-standardized observed data, provided that survey incompleteness and/or unevenness are also taken into account.

---

[*] *Corresponding author*. Departamento de Ciências Agrárias, CITAA (Azorean Biodiversity Group), Universidade dos Açores, Campus de Angra, 9701-851 Angra do Heroísmo, Portugal. Tel.: +351 295402234; fax: +351 295402205.

E-mail addresses: pborges@uac.pt (P.A.V. Borges), j.hortal@imperial.ac.uk (J. Hortal).

[1] These authors contributed equally to this work.

1146-609X/$ – see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.actao.2008.09.003

1. Introduction

One of the strongest patterns observed in ecology is the relationship between the number of species and area, the so-called species–area relationship (SAR) (e.g., Rosenzweig, 1995, 2003; Lawton, 1999; Ricklefs and Lovette, 1999; Gaston and Blackburn, 2000). However, there is no consensus about the formulation of SAR models and the intercept and slope of the fitted models in each particular scale and study system (He and Legendre, 2002). For example, several mechanisms have been proposed to generate SAR (see Rosenzweig, 1995; Borges and Brown, 1999; Turner and Tjørve, 2005; Whittaker and Fernández-Palacios, 2007), each of them making different predictions about SAR behavior. In addition, the scale used to measure SAR also has an effect on the parameters and predictive power of the models (Palmer and White, 1994; Crawley and Harral, 2001; Palmer, 2007; J. Hortal, K.I. Ugland, P.A.V. Borges and R.J. Whittaker, unpublished).

Uncertainty is an essential component of the measurement of any biodiversity descriptor, and the source and degree of uncertainty varies according to the data of origin and the purpose of the analyses (Hortal, 2008). Species richness is usually measured as the observed number of species in a given area. However, it is extremely difficult to obtain a complete inventory of the fauna or flora at a given place, and biodiversity data commonly suffer from heterogeneity in sampling strategies and/or sample size, as well as from survey unevenness (see discussion in Hortal and Lobo, 2005; Hortal et al., 2007, 2008; Lobo et al., 2007). Because of this, sampling success is not always the same in all the surveyed areas, leading to potentially important biases in the proportion of the total number of species that are inventoried in each place. The way species richness is measured might affect SAR, due to sampling effects and survey unevenness (see, e.g., Rosenzweig, 1995; Barnosky et al., 2005; Ouin et al., 2006). Therefore, SAR models might not be useful if richness figures are biased due to sampling unevenness (Rosenzweig, 1995), because the largely unknown errors associated with observed richness could have an impact on the slope and goodness of fit of SAR models based on empirical data (see Turner and Tjørve, 2005).

A possible way to minimize sampling biases is to estimate the number of species at each site (Colwell and Coddington, 1994; Gotelli and Colwell, 2001; Hortal et al., 2004; Walther and Moore, 2005). Richness estimators may correct sampling bias and adequately estimate species richness at each site, improving the ecological models obtained with these data (e.g., SAR models), even in studies where a proper standardized sampling program was applied. Several SAR studies have applied species richness estimators to correct for sampling bias (e.g., Smith, 2001; Oertli et al., 2002). Here, Cam et al. (2002) showed that the slopes of SARs based on observed species numbers are steeper than those of curves based on estimates of richness; they suggested that these changes in slope were produced by the correction made by estimators, which will minimize the bias in observed richness values caused by undetected species. Interestingly, the biases produced by survey unevenness in predictive maps of species richness are eliminated when estimated values instead of observed numbers are used to develop the models (see Hortal et al., 2004). Therefore, methods correcting for biases in observed richness values, such as estimators, might improve the reliability of SAR models. However, no formal test on the comparative performance of species richness estimators for SAR studies has been carried out yet.

Here, we investigate whether there is a significant gain in using non-parametric richness estimators to build SAR models with data coming from standardized surveys, as opposed to using the observed number of species (S_{obs}). In addition, we aim to ascertain if the SAR models obtained from estimated richness scores differ significantly from those built from observed data. More precisely, we test if the use of estimated instead of observed richness values either improves the goodness of fit or alters the intercept and slope of SAR models. For this comparison, we use several non-parametric estimators that can be easily calculated using available software (e.g., EstimateS – Colwell, 2006; Species Diversity and Richness – Henderson and Seaby, 2006). We test three hypotheses:

(i) the intercept of the SAR models built from species richness estimators will be higher than when built from observed data, when the estimated richness values are typically higher than the observed ones;
(ii) the slope of the SAR will be different when calculated from observed and estimated richness values; if, according to Cam et al. (2002), the richness of small areas is underestimated by the surveys (due to a sampling problem caused by the difficulty of detecting rare species), the SAR based on observed number of species would be the steeper of the two; if, on the contrary, the increments between observed and estimated richness provided by the estimators are comparatively higher in richer places, the SAR based on estimated values would be the steeper;
(iii) the goodness of fit will be higher in SAR models based on estimated species richness than in models based on observed values, due to their higher *precision* (i.e., fewer random errors, see Walther and Moore, 2005) in the estimates of species richness per site and in the relative magnitude of variations in species richness between sites.

We used three independent datasets from different islands in the Azores archipelago, in the eastern Atlantic Ocean, to test these hypotheses: epigean arthropods in reserves of different sizes in Terceira Island; canopy arthropods on an endemic Azorean tree (*Juniperus brevifolia*) in reserves in several islands with a difference in cover of the target plant; and soil and rock bryophytes in reserves of different sizes in Terceira and Pico.

2. Methods

2.1. Biological datasets

We used three datasets from two large inventory projects which used standardized sampling protocols to describe the spatial variation of species diversity of arthropods and bryophytes at the Azores. This archipelago is included in the

Macaronesian biogeographic region, which also comprises Madeira, the Canary Islands and Cape Verde Islands. The Azores comprise nine main islands and some small islets, extending for about 615 km in a WNW–ESE axis across the Mid-Atlantic Ridge, which separates the western group (Flores and Corvo), from the central (Faial, Pico, São Jorge, Terceira and Graciosa) and the eastern (São Miguel and Santa Maria) island groups. These islands were originally covered by laurisilva, a dense forest of laurel-leaved evergreen hardwood trees. Arthropods living on the remaining fragments of native vegetation were surveyed within Natural Forest Reserves and/or NATURA 2000 protected areas on seven of the Azorean Islands (excluding the smaller, Graciosa and Corvo) for the Biodiversidade de Artrópodes da Laurisilva dos Açores project (1999–2003; herein 'BALA'; see Borges et al., 2005, 2006; Ribeiro et al., 2005; Hortal et al., 2006). Bryophytes were also surveyed in several of these areas using standardized sampling protocols (see Gabriel and Bates, 2005).

2.1.1. *Dataset 1 (EAR): epigean arthropods in reserves of Terceira Island*

We used the eight native forest fragments surveyed by BALA in Terceira. Transects (150 × 5 m) were randomly placed within these fragments of protected native forest areas. The number of transects per forest fragment was established using a logarithmic scale, assuming a SAR with a slope (z) of 0.35 in a log–log scale (i.e., a 10 fold increase in area implies a doubling in the number of species): placing one transect in 1-ha fragments, two transects in 10-ha fragments, and so on. Consequently, larger reserves received higher sampling effort (i.e., "proportional sampling"); such a protocol is intended to capture not only "area *per se* effects" but also spatial beta diversity that prevails in larger areas (Borges et al., 2005; unpublished data; see also Schoereder et al., 2004). The forest fragments sampled include a large protected forest area ("Serra de Santa Bárbara e Mistérios Negros" that is included in NATURA 2000 European Community Conservation scheme) sampled with eight transects, three medium-sized forest fragments sampled with four transects, three small fragments sampled with two transects and one 3-ha fragment sampled with a single transect (Table 1). Transects were allocated by randomly selecting the starting points in a map, regardless of accessibility (though excluding cliffs), and then randomly selecting the direction of the transect in the field. In each transect the same number of samples was set up to sample epigean arthropods: 30 pitfall traps set in the ground for at least a 2-week period during the summer (see details at Borges et al., 2005). Additional survey techniques were used in these transects to sample canopy arthropods (see Section 2.1.2).

2.1.2. *Dataset 2 (CAJ): canopy arthropods in* Juniperus brevifolia *trees*

We used data for 16 native forest fragments from six islands (Table 2). These forest fragments were surveyed during BALA, following the area-dependent proportional sampling design described in Section 2.1.1 to allocate transects in each forest fragment. At each transect, 10 replicates of the three most abundant and most common woody plant species (trees and shrubs) were sampled in a standardized way: one branch of each of the most common species was sampled at a 15-m interval along the transect. Usually, three plant species were clear dominants, but sometimes fewer than three woody plant species were present, so the survey was restricted to these species (see Ribeiro et al., 2005 for details). Due to this, we used only the arthropod community in the canopy of the most common tree species (*Juniperus brevifolia*) to avoid habitat diversity effects. The area available for this community was estimated as the number of samples taken on *J. brevifolia* canopy available per forest fragment. Since the selection of a plant for surveying implied that it constitutes an important part of the canopy in the transect, if the plant was surveyed in many transects, it probably has a wide distribution and abundance in a particular forest fragment. Thus, the number of samples per fragment can be taken as a surrogate of the "habitat area" available for the canopy arthropod fauna.

2.1.3. *Dataset 3 (BD): bryophyte dataset*

Moss and liverwort communities were surveyed in several habitats of seven native forest fragments from two islands (Terceira and Pico). For this work we used only data from soil and rocks, equivalent to the habitat of epigean arthropod communities. A proportional sampling approach was also used, with larger forest fragments receiving more samples. However, the rate of species accumulation with area is smaller in bryophytes than in arthropods (see Gabriel and Bates, 2005, and Section 4), so the number of transects per fragment was set up according to the following model: $n = 5.78 \times \text{Area}^{0.25}$. The number of samples taken ranged from 10 to 40 (Table 3). As more samples than necessary were available for some of the forest fragments, we randomly

Table 1 – List of the native forest fragments included in the EAR dataset, with their respective codes, names, areas (in ha) and numbers of sampled transects (T). Estimator abbreviations as in the text

Name	Code	Area	T	S_{obs}	ACE	ICE	Chao1	Chao2	Jack1	Jack2	Bootstrap
Serra de St^{a} Bárbara e M. Negros	SB	1274	8	61	88.19	87.68	87.41	84.21	82.9	95.82	70.6
Guilherme Moniz	GM	408	4	43	60.16	59.66	62.91	58.86	55.89	64.77	48.62
Biscoito da Ferraria	BF	391	4	52	76.53	78.76	65.69	73.52	70.84	82.69	60.21
Terra Brava	TB	143	4	57	75.28	76.56	66.1	66.03	74.84	76.95	65.8
Pico do Galhardo	GH	66	2	40	72.32	79.3	98.06	104.4	59.67	77.1	47.84
Algar do Carvão	AC	28	2	30	41.96	44	39.2	41.22	40.82	47.65	34.75
Matela	M	25	2	45	55.69	55.88	50.56	50.11	56.79	57	51.09
Corrente de Lava	CL	3	1	22	42.73	34.41	72	72	31.67	41	25.73

Table 2 – List of the native forest fragments included in the CAJ dataset, with their names, islands of occurrence, codes and numbers of available arthropod samples of *Juniperus brevifolia*. Estimator abbreviations as in the text

Name	Island	Code	Samples	S_{obs}	ACE	ICE	Chao1	Chao2	Jack1	Jack2	Bootstrap
Serra de St^{a} Bárbara e M. Negros	Terceira	TER-SB	150	79	116.91	118.03	142.39	141.45	108.8	132.52	91.18
Morro Alto e Pico da Sé	Flores	FLO-MA	80	46	63.61	65.51	72.72	81.83	60.81	73.51	51.99
Biscoito da Ferraria	Terceira	TER-BF	60	49	57.11	61.03	55.53	60.28	61.78	68.65	54.76
Mistério da Prainha	Pico	PIC-MP	50	61	72.7	77.02	68.41	74.24	75.7	83.52	67.66
Pico do Galhardo	Terceira	TER-GH	40	51	60.96	63.17	57.96	57.05	62.7	64.84	56.78
Lagoa do Caiado	Pico	PIC-LC	40	44	52.37	50.11	49.23	47.31	52.78	52.07	48.69
Caveiro	Pico	PIC-C	40	48	66.13	68.84	63.15	63.13	63.6	72.32	54.88
Terra Brava	Terceira	TER-TB	30	41	55.41	53.57	63.67	63.67	52.6	62	45.74
Pico Pinheiro	S. Jorge	SJG-P	20	42	51.12	53.6	49.97	46.62	52.45	52.97	47.54
Pico Frades – Topo	S. Jorge	SJG-T	20	31	35.22	36.73	32.14	35.78	38.6	41.54	34.62
Caldeiras Funda e Rasa	Flores	FLO-FR	20	35	72.96	88.58	69.53	99.44	54	70.29	42.5
Pico da Vara	S. Miguel	SMG-PV	10	32	38.66	39.16	39.25	35.56	39.2	40.62	35.66
Manhenha	Pico	PIC-M	10	28	32.91	37.92	29.92	32.06	36.1	37.61	32
Graminhais	S. Miguel	SMG-G	10	21	25.72	32.35	26.33	33.5	29.1	34.88	24.4
Atalhada	S. Miguel	SMG-A	10	36	40.22	45.53	37.39	41.7	45	48.02	40.43
Cabeço do Fogo	Faial	FAI-CF	8	26	41.69	39.8	37.56	32.96	35.63	39.38	30.49

eliminated some of them, selecting equivalent numbers of soil and rock samples. Each sample consists of a small quadrat (30 × 30 cm). From each quadrat, the abundance of each species of bryophytes and other plants (macrolichens and vascular plants) was recorded, using a scale with 10 classes of percentage cover (see Gabriel and Bates, 2005). However, percentage cover is not a direct measure of abundance, so for this dataset we used only presence–absence data and tested only incidence-based estimators, to avoid violating the assumptions of abundance-based estimators.

2.2. *Data analysis*

We tested seven non-parametric estimators available in EstimateS 8.0 software (Colwell, 2006; available at http://purl.oclc.org/estimates) to estimate species richness (S_{est}) in alternative to the observed number of species (S_{obs}) in the SAR models: ACE, ICE, Chao1, Chao2, Jackknife1 (Jack1), Jackknife2 (Jack2) and Bootstrap (see Colwell and Coddington, 1994; Gotelli and Colwell, 2001; or Hortal et al., 2006 for more details on these estimators and their performance). We used small grain sizes (i.e., the way sampling data is grouped) for the analyses (individual traps for epigean arthropods, individual canopy samples for canopy arthropods, and individual quadrats for bryophytes) because it produced the most precise and unbiased estimations with the estimators used (see Hortal et al., 2006). S_{est} figures were calculated as the average of 100 randomizations. According to the Coleman test for heterogeneity (see Colwell and Coddington, 1994), all matrices were derived from a homogeneous (stable) community, so heterogeneity was expected from unsystematic (random) sampling errors.

We used the power law function (the most commonly used model) to describe SAR: $S = cA^{z}$, expressed as a log–log model: $\log_{10}Y = c + z\log_{10}X$, where Y is the response variable species richness (S), and X is the explanatory variable number of units of area (A). The parameters are: the intercept (c), a fitting coefficient or estimated number of species per "unit" area; and the slope (z), that indicates the rate of species number increasing with area. The exponential model, often used by botanists (Rosenzweig, 1995) showed no relevant improvement for bryophyte data. Thus, we applied the log–log function to all three datasets, and the parameters of SAR models were obtained using least squares linear regression.

Our three hypotheses were tested by assessing whether the intercept, slope and goodness of fit differed between the regression models obtained from either observed or estimated data. For example, to assess if SAR functions created from estimated richness values would show a different slope from that obtained with empirical data (our second prediction), the slope of a SAR model from S_{est} was considered different than the slope obtained from S_{obs} if their estimates were

Table 3 – List of the native forest fragments included in the BD dataset, with their codes, names, areas (in ha) and number of sampled transects (*T*) and quadrats (Q, point samples). Estimator abbreviations as in the text

Name	Island	Code	Area	T (Q)	S_{obs}	ICE	Chao2	Jack1	Jack2	Bootstrap
Serra de St^{a} Bárbara e M. Negros	Terceira	TER-SB	1274	4 (40)	66	88.27	79.17	86.48	92.54	75.8
Caveiro	Pico	PIC_C	199	2 (20)	54	74.52	73.1	73	83.33	62.59
Terra Brava	Terceira	TER-TB	143	2 (20)	51	62.28	58.04	64.3	66.67	57.73
Lagoa do Caiado	Pico	PIC-LC	131	2 (20)	46	60.24	59.24	60.25	67.78	52.47
Pico do Galhardo	Terceira	TER-GH	66	2 (20)	55	65.23	68.24	69.25	76.78	61.84
Matela	Terceira	TER-M	25	1 (10)	39	54.89	56.31	53.4	61.93	45.42
Corrente de Lava	Terceira	TER-CL	3	1 (10)	42	65.6	57.39	58.2	66.2	49.34

significantly different in a two-tailed *t*-test. The other two hypotheses were tested in the same way, by comparing the intercepts (*c*) and coefficients of determination (r^2) and the overlap of their confidence intervals.

3. Results

The estimated richness values were different (and higher) than the observed ones, although the magnitude of these differences varied widely among estimators and cases (Tables 1, 2 and 3). These differences caused significant increments in the intercept in the EAR and BD datasets for almost all estimators, although no significant differences were found in the CAJ dataset (Table 4). Importantly, the slope of SAR models for S_{obs} and S_{est} were quite similar in the three datasets (Table 4). In fact, significant differences in the slope of SAR models obtained with observed and estimated richness appeared only in five cases, four of them showing poor goodness of fit values. The SAR model obtained using ACE estimator in the epigean arthropods dataset was the only one showing a slope significantly smaller than that obtained with S_{obs} while retaining a significant goodness of fit, although such goodness of fit was similar (albeit smaller) to the one obtained from observed values. Four other SAR models based in S_{est} did show a significant decrease in slope (Chao1 and Chao2 in the Epigean Arthropods dataset and ICE and Chao2 in the Ground Bryophytes dataset), but these models showed no significant relationship between estimated values and area (Table 4).

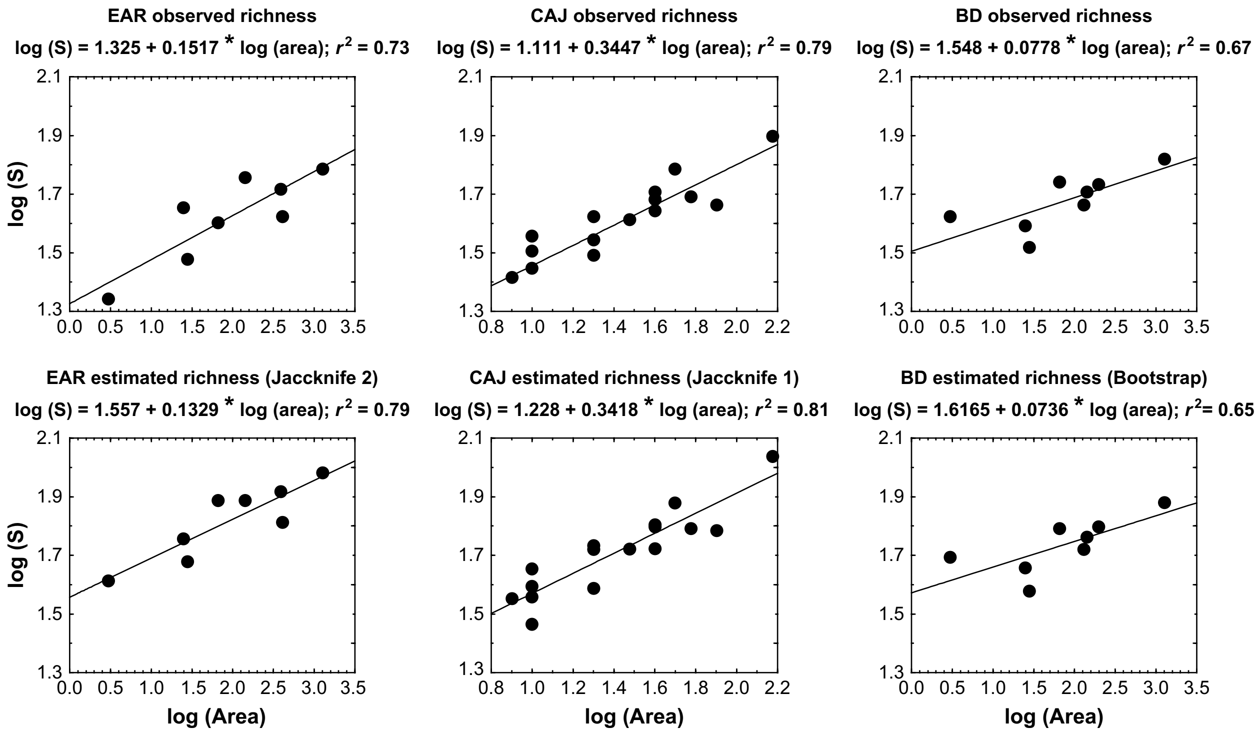
In general, z values were always smaller for S_{est} in EAR and BD datasets, although these differences were not significant. No slope differed significantly from the observed in CAJ, but in this case some slopes were steeper than the S_{obs} slope (0.35) (their average z value [0.37] was also higher) (Table 4). The dispersion of the data points in the SAR models obtained from S_{est} was smaller that from S_{obs} in some estimators for the two arthropod datasets (Fig. 1). Thus, some estimators (Jackknife 1 and 2 and Bootstrap) produced slightly better fits than that obtained with S_{obs}. However, improvements in r^2 were never significant, and all estimators performed worse than the observed values in the bryophyte dataset (sometimes with significantly smaller fits; Table 4, Fig. 1). Nevertheless, although all SAR models obtained from observed data were significant, some models calculated from S_{est} were not (Table 4).

4. Discussion

Our analyses show that using estimated instead of observed richness values can significantly alter the intercept of SAR models, but it does not produce significant changes in the

Table 4 – Parameters, coefficient of determination and significance of SARs based on the observed species richness (S_{obs}) and species richness estimators for three different datasets. As ACE and Chao1 need abundance data, they have been excluded from the presence/only Bryophyte dataset. *c* and *z* are the intercept and slope of the SAR models; r^2 and *p* measure the goodness of fit of these models. SE is the standard error in the estimation of *c*, *z* and r^2, and *t* is the result of a two-tailed *t*-test comparing the differences in the intercept, slope and goodness of fit between SAR models developed with estimated and observed richness; $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Richness measure	*c*	SE	*t*	*z*	SE	*t*	r^2	SE	*t*	*p*
Epigean arthropods (EAR)										
S_{obs}	*1.32*	*0.08*		*0.150*	*0.037*		0.73	0.08		0.0069
ACE	1.56	0.07	4.61^{**}	0.118	0.033	-2.69^{*}	0.68	0.07	−0.77	0.0114
ICE	1.51	0.08	3.51^{*}	0.143	0.037	−0.59	0.71	0.08	−0.32	0.0086
Chao1	1.73	0.12	6.49^{***}	0.045	0.058	-7.24^{***}	0.09	0.13	-8.85^{***}	0.4640
Chao2	1.74	0.12	6.64^{***}	0.042	0.059	-7.43^{***}	0.08	0.13	-8.99^{***}	0.5087
Jackknife1	1.47	0.07	2.86^{*}	0.144	0.033	−0.54	0.76	0.07	0.41	0.0051
Jackknife2	1.56	0.06	4.59^{**}	0.133	0.028	−1.47	0.79	0.06	0.94	0.0033
Bootstrap	1.39	0.08	1.28	0.150	0.036	−0.07	0.74	0.08	0.15	0.0061
Canopy arthropods (CAJ)										
S_{obs}	*1.11*	*0.07*		*0.351*	*0.048*		0.79	0.07		<0.0001
ACE	1.21	0.09	1.17	0.358	0.063	0.18	0.70	0.09	−1.18	<0.0001
ICE	1.28	0.10	2.00	0.327	0.065	−0.60	0.64	0.10	−1.85	0.0002
Chao1	1.12	0.10	0.07	0.418	0.067	1.64	0.74	0.10	−0.69	<0.0001
Chao2	1.14	0.11	0.38	0.415	0.075	1.51	0.69	0.11	−1.27	<0.0001
Jackknife1	1.23	0.06	1.52	0.342	0.044	−0.25	0.81	0.06	0.25	<0.0001
Jackknife2	1.24	0.07	1.64	0.367	0.049	0.40	0.80	0.07	0.10	<0.0001
Bootstrap	1.18	0.07	0.84	0.338	0.045	−0.34	0.80	0.07	0.10	<0.0001
Ground Bryophytes (BD)										
S_{obs}	*1.55*	*0.05*		*0.078*	*0.034*		0.67	0.05		0.0240
ICE	1.73	0.06	8.49^{***}	0.050	0.029	-3.09^{*}	0.37	0.05	-15.86^{***}	0.1443
Chao2	1.71	0.05	8.08^{***}	0.052	0.022	-3.01^{*}	0.52	0.04	-8.02^{***}	0.0663
Jackknife1	1.69	0.05	7.10^{***}	0.067	0.024	−1.22	0.62	0.05	-2.69^{*}	0.0352
Jackknife2	1.76	0.05	10.47^{***}	0.056	0.024	−2.54	0.53	0.05	-7.70^{***}	0.0648
Bootstrap	1.62	0.05	3.44^{*}	0.074	0.024	−0.49	0.65	0.05	−0.99	0.0278



EAR observed richness
log (S) = 1.325 + 0.1517 * log (area); r^2 = 0.73

CAJ observed richness
log (S) = 1.111 + 0.3447 * log (area); r^2 = 0.79

BD observed richness
log (S) = 1.548 + 0.0778 * log (area); r^2 = 0.67

EAR estimated richness (Jaccknife 2)
log (S) = 1.557 + 0.1329 * log (area); r^2 = 0.79

CAJ estimated richness (Jaccknife 1)
log (S) = 1.228 + 0.3418 * log (area); r^2 = 0.81

BD estimated richness (Bootstrap)
log (S) = 1.6165 + 0.0736 * log (area); r^2= 0.65

Fig. 1 – Species–area curves for the three datasets. The empirical ones (obtained with observed richness, S_{obs}) are shown in the upper row, and the ones obtained with the best-performing estimators (S_{est}) are shown in the lower row; epigean arthropods (EAR), canopy arthropods (CAJ) and epigean bryophytes (BD) are shown in the left, centre and right columns, respectively.

slope and goodness of fit of SAR models, at least when data come from standardized surveys. Therefore, while we accept our first hypothesis (intercepts will be higher in SAR models calculated from estimated data), we reject our second and third hypotheses (there will be consistent changes in slope and significant improvements in the goodness of fit of the models). When important changes in slope occur, they are associated with unreliable SAR models (except in one case). Some slight changes in slopes occur in good models developed from estimated values, but not in a consistent way: although most estimators result in smaller z values (as expected), some were steeper in the CAJ dataset (Table 4). Moreover, no estimator outperformed the fitting of SAR models based on observed data; although some estimators (the two Jackknives and Bootstrap) produce slight improvements for both arthropod datasets, they were not significant (Table 4). On the contrary, some estimators produced unrealistic estimated values.

Some authors have argued that the rate of increase in species richness with area can be overestimated if data on the observed number of species is used (Cam et al., 2002; Turner and Tjørve, 2005); thus, lower slopes would be expected when SAR is calculated from richness estimates. Such a prediction is made on the assumption that the relative effect of undetected species in the richness estimations will be higher in places with smaller observed richness values, because it is more difficult to describe the distribution of abundances in the whole community in places with fewer species (see Cam et al., 2002). Although some of our results might give some support to these ideas, others do not (CAJ dataset), and, more importantly, the possible improvements provided by richness estimates are not enough to produce significant changes in the measured slope of the SAR.

Are our results context-specific, or can they be extrapolated to other cases? The slopes (z) of SAR models differed between the three datasets studied: arthropod datasets presented steeper slopes than bryophytes, and within arthropods CAJ fauna presented a higher species replacement than EAR (Table 4, Fig. 1). In spite of their high local richness (see also Gabriel and Bates, 2005), Azorean mosses and liverworts show low β-diversity levels; important increments in area do not result in significant increments in species richness, as expected due to their high colonization ability. Arthropod communities also follow two well known empirical patterns, that is, low slopes for forest fragments within an island (EAR dataset), as expected for ‘islands’ of similar habitat within continental areas, and high slopes for fragments across islands (CAJ dataset), following a typical oceanic archipelago z value around 0.35 (see Rosenzweig, 1995; Whittaker and Fernández-Palacios, 2007). Therefore, the SARs we used as examples constitute a representative sample of the most commonly found relationships between richness and area.

Our results mean that there is no significant gain in using estimators for SAR studies when data come from standardized studies. However, they also indirectly imply that, *a priori*, SAR models obtained from estimated data might not be different from the true SAR that would be obtained if the actual richness in each land patch was known. This has important implications for the utility of species richness estimators as a mean of data standardization. The potential advantage of species richness estimators is that they might provide unbiased richness values, eliminating the effect of uneven sampling effort (see Hortal et al., 2004). However, they present some critical problems, such as the lack of stabilization with increasing sampling effort (Hortal et al. 2006) and our lack of knowledge on how close their estimates are from true richness values. Also, biodiversity data are often biased and uneven (Hortal et al., 2007, 2008). A number of sampling artefacts are likely to arise according to the kind of organisms surveyed and the design of the study. Such data unevenness often results in a picture of richness patterns different than the actual one.

If surveys follow a standardized protocol designed to avoid (i) biases in the effort invested in different places (and not the number of individuals surveyed) and (ii) misrepresentation of some geographic areas or ecological conditions, sampling artefacts are minimized (see Schoereder et al., 2004; Hortal and Lobo, 2005; Borges et al., 2005; Funk et al., 2005; Hortal et al., 2007). This is the case with the datasets used here and many other standardized ecological datasets, where sampling effort is not uneven. In our case, the spatial design of the surveys (i.e., the number of sampling plots per forest fragment) accounts for differences in area, and the effort devoted to each sampling plot is constant. Due to this, observed richness values are a realistic picture of the actual richness patterns; thus, neither significant changes in the slope of the SAR nor model fitting improvements (maximum gain in r^2 was 0.6) appear when estimators are used. This does not mean that richness estimators are not worth using in any case. If survey design or sampling effort has been uneven, the picture of richness patterns offered by observed data might be unrealistic, and estimated values are a necessary step to obtain reliable models of the species–area or any other ecological relationship. This will change the magnitude of the richness scores analyzed (i.e., increase richness scores and also the intercept of the SAR), but will result in more reliable descriptions of the slope and explanatory power of the actual SAR. In this case, a previous assessment of the quality of the data and its spatial coverage within each of the studied land patches (as developed in Hortal and Lobo, 2005; Hortal et al., 2007, 2008) is needed to ascertain if the observed picture of species richness is reliable enough for the estimators to provide reliable richness values (Hortal et al., 2004).

5. Concluding remarks

All measures of the diversity of biological assemblages, such as richness, have their own sources of uncertainty, which must be acknowledged and taken into account when analyzing these data (Hortal, 2008). When studying species–area relationships, we recommend careful thought about the origin of the data used, its potential drawbacks, and the treatment needed. If data comes from standardized surveys, it can be assumed that observed richness provides a good picture of the actual distribution of richness in the studied dataset, and therefore SAR based on the observed data would be preferred. The same can be true for local checklists of natural areas or islands gathered from exhaustive surveys (see Hortal, 2008). If, however, data come from non-standardized sources and/or non-exhaustive surveys, species richness estimators could provide a means of standardizing these data in order to obtain comparable richness values, provided that the possible biases and drawbacks in these data have been investigated, and their effect on richness estimates is likely to be negligible.

Acknowledgements

The authors are especially grateful to Clara Gaspar for her contribution to BALA, to Shai Meiri for his critical review of the manuscript, as well as to Regina Meneses for the English editing of a former version. We are also indebted to all the taxonomists who helped identify the arthropod morphospecies (H. Enghoff, F. Ilharco, V. Manhert, J. Ribes, A.R.M. Serrano and J. Wunderlich) and bryophyte species (J.W. Bates, R. Schumacker, C. Sérgio and E. Sjögren). We also wish to thank C. Aguiar, J. Amaral, I. Amorim, G. André, A. Arraiol, E. Barcelos, P. Cardoso, F. Dinis, P. Gonçalves, S. Jarroca, C. Melo, F. Pereira, H. Mas i Gisbert, S. Ribeiro, A. Rodrigues, A. Serrano, L. Vieira and A. Vitorino for their contribution to field and/or laboratory work. Funding for field data collecting was provided by the Direcção Regional dos Recursos Florestais Secretaria Regional da Agricultura e Pescas) through the project Reservas Florestais dos Açores: Cartografia e Inventariação dos Artrópodes Endémicos dos Açores (PROJ. 17.01 – 080203). This work was also supported by the project ATLANTICO, financed by the European Community programme INTERREG III B. JH was supported by Portuguese FCT grant BPD/20809/2004, a CITA A grant and the UK Natural Environment Research Council. NH is currently supported by a grant from CITA A.

REFERENCES

Barnosky, A.D., Carrasco, M.A., Davis, E.B., 2005. The impact of the species–area relationship on estimates of paleodiversity. PLoS Biology 3, e266.

Borges, P.A.V., Brown, V.K., 1999. Effect of island geological age on the arthropod species richness of Azorean pastures. Biological Journal of the Linnean Society 66, 373–410.

Borges, P.A.V., Aguiar, C., Amaral, J., Amorim, I.R., André, G., Arraiol, A., Baz, A., Dinis, F., Enghoff, H., Gaspar, C., Ilharco, F., Mahnert, V., Melo, C., Pereira, F., Quartau, J.A., Ribeiro, S., Ribes, J., Serrano, A.R.M., Sousa, A.B., Strassen, R.Z., Vieira, L., Vieira, V., Vitorino, A., Wunderlich, J., 2005. Ranking protected areas in the Azores using standardized sampling of soil epigean arthropods. Biodiversity and Conservation 14, 2029–2060.

Borges, P.A.V., Lobo, J.M., Azevedo, E.B., Gaspar, C., Melo, C., Nunes, V.L., 2006. Invasibility and species richness of island endemic arthropods: a general model of endemic vs. exotic species. Journal of Biogeography 33, 169–187.

Cam, E., Nichols, J.D., Hines, J.E., Sauer, J.R., Alpizar-Jara, R., Flather, C.H., 2002. Disentangling sampling and ecological explanations underlying species–area relationships. Ecology 83, 1118–1130.

Colwell, R.K., Coddington, J.A., 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society of London B 345, 101–118.

Colwell, R.K., 2006. EstimateS: statistical estimation of species richness and shared species from samples. Version 8. User's Guide and application published at: http://purl.oclc.org/estimates/, Department of Ecology & Evolutionary Biology, University of Connecticut, Storrs.

Crawley, M.J., Harral, J.E., 2001. Scale dependence in plant biodiversity. Science 291, 864–868.

Funk, V.A., Richardson, K.S., Ferrier, S., 2005. Survey-gap analysis in expeditionary research: where do we go from here? Biological Journal of the Linnean Society 85, 549–567.

Gabriel, R., Bates, J.W., 2005. Bryophyte community composition and habitat specificity in the natural forests of Terceira, Azores. Plant Ecology 177, 125–144.

Gaston, K.J., Blackburn, T.M., 2000. Pattern and Process in Macroecology. Blackwell Science, Oxford.

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.

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

Henderson, P.A., Seaby, R.M.H., 2006. Species Diversity and Richness v4.0. Pisces Conservation Ltd., Pennington, Lymington.

Hortal, J., 2008. Uncertainty and the measurement of terrestrial biodiversity gradients. Journal of Biogeography 35, 1355–1356.

Hortal, J., Lobo, J.M., 2005. An ED-based protocol for the optimal sampling of biodiversity. Biodiversity and Conservation 14, 2913–2947.

Hortal, J., Garcia-Pereira, P., García-Barros, E., 2004. Butterfly species richness in mainland Portugal: predictive models of geographic distribution patterns. Ecography 27, 68–82.

Hortal, J., Borges, P.A.V., Gaspar, C., 2006. Evaluating the performance of species richness estimators: sensitivity to sample grain size. Journal of Animal Ecology 75, 274–287.

Hortal, J., Lobo, J.M., Jiménez-Valverde, A., 2007. Limitations of biodiversity databases: case study on seed–plant diversity in Tenerife (Canary Islands). Conservation Biology 21, 853–863.

Hortal, J., Jiménez-Valverde, A., Gómez, J.F., Lobo, J.M., Baselga, A., 2008. Historical bias in biodiversity inventories affects the observed realized niche of the species. Oikos 117, 847–858.

Lawton, J.H., 1999. Are there general laws in ecology? Oikos 84, 177–192.

Lobo, J.M., Baselga, A., Hortal, J., Jiménez-Valverde, A., Gómez, J.F., 2007. How does the knowledge about the spatial distribution of Iberian dung beetle species accumulate over time? Diversity & Distributions 13, 772–780.

Oertli, B., Joye, D.A., Castella, E., Juge, R., Cambin, D., Lachavanne, J. B., 2002. Does size matter? The relationship between pond area and biodiversity. Biological Conservation 104, 59–70.

Ouin, A., Sarthou, J.-P., Bouyjou, B., Deconchat, M., Lacombe, J.-P., Monteil, C., 2006. The species–area relationship in the hoverfly (Diptera, Syrphidae) communities of forest fragments in southern France. Ecography 29, 183–190.

Palmer, M.W., White, P.S., 1994. Scale dependence and the species–area relationship. American Naturalist 144, 717–740.

Palmer, M., 2007. Species–area curves and the geometry of nature. In: Storch, D., Marquet, P.A., Brown, J.H. (Eds.), Scaling Biodiversity, Ecological Reviews. Cambridge University Press, Cambridge, pp. 15–31.

Ribeiro, S., Borges, P.A.V., Gaspar, C., Melo, C., Serrano, A.R.M., Amaral, J., Aguiar, C., André, G., Quartau, J.A., 2005. Canopy insect herbivores in the Azorean laurisilva forests: key host plant species in a highly generalist insect community. Ecography 28, 315–330.

Ricklefs, R.E., Lovette, I.J., 1999. The roles of island area per se and habitat diversity in the species–area relationships of four Lesser Antillean faunal groups. Journal of Animal Ecology 68, 1142–1160.

Rosenzweig, M.L., 1995. Species Diversity in Space and Time. Cambridge University Press, Cambridge.

Rosenzweig, M.L., 2003. Reconciliation ecology and the future of species diversity. Oryx 37, 194–205.

Schoereder, J.H., Galbiati, C., Ribas, C.R., Sobrinho, T.G., Sperber, C.F., DeSouza, O., Lopes-Andrade, C., 2004. Should we use proportional sampling for species–area studies? Journal of Biogeography 31, 1219–1226.

Smith, F., 2001. Historical regulation of local species richness across a geographic region. Ecology 82, 792–801.

Turner, W.R., Tjørve, E., 2005. Scale-dependence in species–area relationship. Ecography 28, 721–730.

Walther, B.A., Moore, J.L., 2005. The definitions 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.

Whittaker, R.H., Fernández-Palacios, J.M., 2007. Island Biogeography. Ecology, Evolution, and Conservation, second ed. Oxford University Press, Oxford.