PLOS ONE

RESEARCH ARTICLE

Disentangling the effects of sampling scale and size on the shape of species abundance distributions

Renato A. Ferreira de Lima[1¤]*, Paula Alves Condé[1], Cristina Banks-Leite[2], Renata C. Campos[3], Malva I. Medina Hernández[3], Ricardo R. Rodrigues[4], Paulo I. Prado[1]

1 Laboratório de Ecologia Teórica (LET), Departamento de Ecologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil, **2** Department of Life Sciences, Imperial College London, Ascot, United Kingdom, **3** Departamento de Ecologia e Zoologia, Centro de Ciências Biológicas, Universidade Federal de Santa Catarina, Florianópolis, Brazil, **4** Departamento de Ciências Biológicas, Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, Brazil

¤ Current address: Naturalis Biodiversity Center, CR Leiden, The Netherlands
* raflima@usp.br

Check for updates

OPEN ACCESS

Citation: Lima RAFd, Condé PA, Banks-Leite C, Campos RC, Hernández MIM, Rodrigues RR, et al. (2020) Disentangling the effects of sampling scale and size on the shape of species abundance distributions. PLoS ONE 15(9): e0238854. https://doi.org/10.1371/journal.pone.0238854

Editor: Paulo De Marco Júnior, Universidade Federal de Goias, BRAZIL

Received: January 30, 2020

Accepted: August 25, 2020

Published: September 11, 2020

Copyright: © 2020 Lima et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: Data collection was supported by the grant 1999/09635-0 and 2013/50718-5, São Paulo Research Foundation (FAPESP) and by CNPq/BMBF (process 690144/01-6). RAFL was supported by the 'Coordenação de Aperfeiçoamento de Pessoal de Nível Superior' (CAPES) and by grant 2013/08722-5, São Paulo Research Foundation (FAPESP). PIP was

Abstract

Many authors have tried to explain the shape of the species abundance distribution (SAD). Some of them have suggested that sampling spatial scale is an important factor shaping SADs. These suggestions, however, did not consider the indirect and well-known effect of sample size, which increases as samples are combined to generate SADs at larger spatial scales. Here, we separate the effects of sample size and sampling scale on the shape of the SAD for three groups of organisms (trees, beetles and birds) sampled in the Brazilian Atlantic Forest. We compared the observed SADs at different sampling scales with simulated SADs having the same richness, relative abundances but comparable sample sizes, to show that the main effect shaping SADs is sample size and not sampling spatial scale. The effect of scale was minor and deviations between observed and simulated SADs were present only for beetles. For trees, the match between observed and simulated SADs was improved at all spatial scales when we accounted for conspecific aggregation, which was even more important than the sampling scale effect. We build on these results to propose a conceptual framework where observed SADs are shaped by three main factors, in decreasing order of importance: sample size, conspecific aggregation and beta diversity. Therefore, studies comparing SADs across sites or scales should use sampling and/or statistical approaches capable of disentangling these three effects on the shape of SADs.

Introduction

The species abundance distribution (SAD) is a simple yet compelling way of describing ecological communities [1–3]. Because differences in the shape of SADs may represent changes in diversity across communities, many ecologists have tried to explain what would be the main factors shaping the SAD. There are explanations that rely on how species share niches or

supported by CNPq (process 303878/2008-8) and by grant 09/53413-5, São Paulo Research Foundation (FAPESP). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

resources [4, 5] or on stochastic population dynamics [6–8]. Nevertheless, it has been known for decades that the number of individuals sampled from the community (i.e. sample size) greatly affects the shape of the SADs, a phenomenon known among ecologists as the "veil line" effect [9]. In rank-abundance plots, the increase in sample size leads to less steep SADs, while in log-abundance class plots (i.e. Preston plots) this leads to more symmetrical ones (but see [10] for a discussion). Independently of the way the SAD is presented, the change in its shape is mainly due to the decline of the relative contribution of rare species in the sample (i.e. species in the first abundance classes), when compared to the contribution of more common species [9, 11, 12]. Since the publication of Preston's pioneer work [9, 13], a sampling theory was developed to deduce how the shape of SADs is affected by sample size [1, 11, 14]. It was also shown that species aggregation and detectability patterns may also play a role in shaping SADs [15–17].

More recently, the spatial scale of sampling (i.e. the geographical extent covered by the sampling units) was suggested as an explanation for changes in the shape of SADs [18, 19], a process that some years later was termed as scale-dependent SAD by [20]. Briefly, the SAD would become less steep or more symmetrical as samples from separate sites are pooled together. This trend was observed for herbs sampled over tens of meters [21], and for trees sampled over hundreds of meters [18]. The same result was obtained for corals and reef fishes [22], arthropods [17] and for birds and island fishes [18] sampled in scales of hundreds to thousands of kilometers. Indeed, computer simulations showed that the SAD do change in shape as non-overlapping samples are combined, a process that depends on the degree of species aggregation and species turnover between samples [19, 23]. Therefore, there is evidence of a sampling scale effect on the shape of the SAD, as shown for the species-area and similarity distance-decay curves [24, 25].

However, previous work did not disentangle the effects of sample size and sampling spatial scale in shaping SADs. In these studies, larger sampling scales are obtained by pooling entire samples taken from separate sites [18, 22], leading to an increase in sampling spatial scale *and* sample size as samples are combined [21]. Although some authors have even treated the effect of sample scale and sample size practically as equivalents [22, 26, 27], it is easy to conceive sampling designs where one can increase sampling scale without increasing sample size and vice-versa (*e.g.* [24]). One can also use randomization or rarefaction methods to tell apart the effects of sample size and sampling scale in the shape of SADs [20, 28]. Independently of the approach chosen, the control of sample size effects shaping the SAD is a crucial step in the attempt to understand the ecological processes behind community patterns [29, 30]. This is particularly relevant to explain differences in community diversity sampled at different places or times [28], becoming crucial if these communities were sampled differently [31, 32]. Surprisingly, none of the previous studies claiming an effect of sampling scale on the shape of SADs have explicitly addressed such an important issue.

Here, we disentangle the role of sampling spatial scale and sample size in shaping SADs. Based on similar frameworks that take into account sampling intensity [1, 11, 14, 15], we simulate samples of communities that preserve the properties of observed communities at a given reference scale, but that do not vary in their average sample sizes. This procedure controls for possible effects of sample size while increasing the spatial scale of observation of the community. Following the early work of Preston [9, 13], our basic prediction is that sample size has a greater influence on the shape of SADs than sampling scale. Our prediction would be supported if the shape of an observed SAD matches the shape of the simulated SAD generated from higher-scale observed SADs but controlling for sample size. To assess the generality of our results, we examine our prediction using datasets of three very distinct groups of

organisms (i.e. plants, invertebrates, and vertebrates) and SADs constructed using counts of individuals and biomass.

Materials and methods

Observed SADs

We used data from three very distinct groups of organisms sampled in the Brazilian Atlantic Forest: trees (including palms and tree-ferns), dung beetles (Coleoptera, Scarabaeidae) and understorey birds (see details in S1 Appendix in S1 File). Sampling was divided into four spatial scales: sample (tens of meters), habitat (one to few hundreds of meters), site (hundreds of meters to kilometers), and regional scales (hundreds of kilometers). For trees, the sample scale corresponded to 256 40×40 m permanent forest plots (total of 40.96 ha, 60 838 trees of 483 species sampled) placed in four different sites, which are 75–300 km apart from each other. For beetles, the sample scale corresponded to 134 pitfall traps baited with human feces (7505 beetles of 71 species captured over more than 24 000 trap-hours) placed at three sites (250–300 km apart from each other). For birds, the sample scale corresponded to 65 sampling sites, each of them being sampled using 10 mist nets (6127 individuals and 140 species captured over a total of more than 41 000 net-hours) placed at three sites (50–150 km apart from each other).

For all groups, habitat scale was defined by contiguous samples (plots, pitfalls and mist nets) that could be grouped based on the existence of marked environmental differences within sites, such as topographic positions, size of forest fragment, soil types or differences in disturbance history of the forest. The site scale for all groups was defined by the combination from samples taken from the same site, while the regional scale was obtained by merging all samples from all sites. Because the three groups were sampled using different methods and sampling designs, the mean number of individuals sampled at each scale, used here as a measure of sample intensity, varied among groups as well (S1 Table in S1 Appendix). The population biomass of trees was obtained using allometric equations for woody species, palms, and tree-ferns. For beetles, population biomass was obtained directly from the dry weight of individuals captured in the field. For birds, the average body mass obtained from the literature was used to obtain the population biomass.

Simulated SADs

To control the effect of sample size on SADs at different spatial scales, we used an approach similar to parametric bootstrapping *sensu* [33]. In general terms, we combined subsamples with size equal to n_i/x taken from x separate sites, where n_i the sample size at scale i and x is the number of sites available at the same scale. This procedure creates community samples at a larger sampling scales but with the same average sample size $\sum n_i/x$ of the lower sampling scale. In practical terms, the expected abundance of species i in simulated samples pooled at scale j, ordered from the smallest to the largest scale (j = 1, 2, 3, 4), was:

$$\boldsymbol{E}[\mathbf{x}_{ij}]=\mathbf{x}_{ij}\times\boldsymbol{N}_j/\boldsymbol{N}_k \quad (1)$$

where k is the index for any given upper scale (k = 2, 3, 4); x_{ik} is the observed abundance of species i in the pooled samples at scale k; N_j and N_k are the observed abundances of all species at scale j and k and, thus, N_j/N_k is the sampling intensity. When we combine the expected abundances of all i species, then the resulting SAD kept the same relative abundances of the observed SAD, but with a sample size that matches the mean size of the sample units observed at the scale below (S2 Appendix in S1 File). The mean size used to generate the simulated SADs varied according to the scale and group of organisms considered (S1 Table in S1 File).

For instance, the mean sample sizes at the sample scale were 238, 58 and 94 individuals for trees, beetles and birds, respectively.

We simulated community samples by generating random draws from the Poisson distribution and the Negative Binomial (NB) distribution. Both distributions have a long tradition in ecological literature to simulate data under the assumption of spatial randomness and spatial aggregation of conspecific individuals, respectively [1, 11, 15, 16, 34]. Generating random Poisson draws is similar to performing an individual-based rarefaction of the observed SAD [22, 28] and the comparison between the methods used here and rarefaction techniques indeed yielded similar results (not shown). On the other hand, the use of the NB adds complexity to the analysis which is related to the additional aggregation parameter (). The estimation of is generally scale and density-dependent: species with larger populations tend to have less aggregated distributions (higher) and vice-versa [35]. In this study, we use scale-specific and species-specific values of that were estimated from the empirical data using maximum likelihood techniques. Whenever the estimation of this parameter was not possible, we assigned an arbitrary value of that was equal to the community average of this parameter at the same sampling scale. Details on the estimation of are given in S2 Appendix in S1 File.

Since both Poisson and NB are discrete distributions, they cannot be used directly to assess changes in species biomass distributions. Thus, the generation of simulated samples for biomass distributions was conducted in two steps. The first step was identical to the Poisson and NB draws described above, used to obtain random numbers of individuals per species. Then, for each species with n individuals, we sampled with replacement n values of biomass from the empirical biomass distribution of the same species at the same scale. Although it may not be the ideal solution, this bootstrapping procedure was very easy to implement and provided results that were appropriate enough for the purposes of this study (S2 Appendix in S1 File). For each group of organisms and for both measures of abundance, we draw 10000 simulated community samples.

Observed versus simulated SADs

For each group of organisms (*i.e.*, trees, beetles and birds) and measure of abundance (*i.e.*, counts of individual and biomass), the observed and simulated SADs were presented using rank-abundance plot on logarithmic scales, as suggested by [36]. We used relative measures of abundance in order to make the rank-abundance plots comparable among groups of organisms. Each rank-abundance plot presents only the mean of the observed and simulated SADs at each sampling scale. For instance, the observed SAD at the sample scale represents the mean SAD obtained in 256, 134 and 65 sample units for trees, beetles and birds, respectively. For the simulated SADs, plots present the mean SAD obtained from the 10000 draws for each sampling scale. For the sake of comparison, we also represented SADs using histograms of the mean proportion of species per abundances classes on a $\log_2$ scale (*i.e.*, Preston plots), which are presented along with the 5% and 95% quantiles of simulated abundance values per abundance class (S3 Appendix in S1 File). Differently from [9], we used true octaves that are closed on the right (1, 2, 3–4, 5–8, 9–16 and so on). Although this is not the commonly adopted choice of bins and it is different from the binning proposal of [37], the results remain essentially the same using both binning approaches (results not shown).

We used the Kolmogorov-Smirnov test to compare the observed and simulated SADs. We compared each of the 10 000 simulated SADs obtained from each reference sampling scales (site, habitat and sampling scales) with the mean observed SAD at the same sampling scale. Prior to the Kolmogorov-Smirnov test, we discretized the mean observed SAD and excluded ranks with zero individuals from both observed and simulated SADs. For biomass, we also

excluded the ranks with values of relative biomass smaller than the minimum observed, which were 0.00001, 0.00005 and 0.00009 for trees, beetles and birds, respectively. Then, for each pair of observed and simulated SADs, we obtained the value of the test and the *p*-value associated with the hypothesis that the two SADs come from the same distribution. Here, we used the proportion of simulated SADs being regarded as equal to the mean observed SAD as a measure of goodness of fit. If 5% of the 10000 simulated SADs or more had a *p*-value $\geq$0.05, then the simulated and observed SADs were considered as being equal. The results of the Kolmogorov-Smirnov test applied using the mean observed SAD and the mean simulated SADs (i.e. only one test instead of using the mean of 10000 tests) resulted in very similar results for almost all groups of organisms and measures of abundance (S2 and S3 Tables in S1 File). All analyses were performed in R (www.R-project.org) using the package *sads* [38] and *dgof* [39].

Results

All group of organisms presented very similar trends in their results using counts of individuals: the observed SADs became more concave (less linear on a log-log scale rank-abundance plots – Fig 1) or more symmetrical (on $\log_2$ Preston plots – S1 Fig in S1 File) as we increase the spatial scale of observation. For SADs constructed using biomass instead of counts of individuals, the observed biomass distributions were already concave even at the smaller sampling scales (Fig 2 and S2 Fig in S1 File). However, we still observed the same tendency of more concave shapes at higher scales for all groups, particularly for trees. The change of the shape of the SADs with increasing scale was more conspicuous when results are presented using Preston plots (S1 and S2 Figs in S1 File) instead of rank-abundance plots (Figs 1 and 2), but the latter provides a more complete representation of the response of each individual species in the community.

The simulated SADs presented shapes very similar to the observed ones (Figs 1 and 2) and results were mostly the same for all groups of organisms and for both measures of abundance. The Kolmogorov-Smirnov test confirmed that simulated and observed SADs had very similar shapes in almost all possible combinations of reference sampling scales and community sample sizes (Table 1). Beetles had more deviations between simulated and observed SADs, but the Kolmogorov-Smirnov test did not suggest significant differences between their shapes. This lack of significance for this group, despite the visual differences between simulated and observed SADs (Figs 1 and 2), may be due to the much smaller number of species of beetles in comparison to the two other groups. Simulated samples taken from Negative Binomial distribution (NB) did improve the match between the observed and simulated SADs for trees at all spatial scales when compared to samples taken from the Poisson distribution (S3 Fig in S1 File). The same was not true for beetles and birds, although the latter group did present a non-significant tendency of improvement when samples were taken from the NB. Therefore, we present simulated SADs (Figs 1 and 2; S1 and S2 Figs in S1 File) generated by NB sampling for trees and by Poisson sampling for beetles and birds.

Discussion

As previously found for corals, fish, trees, and birds [18, 22], pooling samples together to increase the spatial scale of observation does indeed change the shape of the SAD. For all groups of organisms studied here, the pooled SADs were progressively more concave down (rank-abundance plots) or more symmetrical (Preston plots). However, this change in the shape of the SAD was close to what would be expected from the increase in sample size caused by pooling sampling units. Thus, the suggested effect of sampling spatial scale on the shape of the SAD is largely due to the well-known 'veil line' effect: increasing sample size decreases the

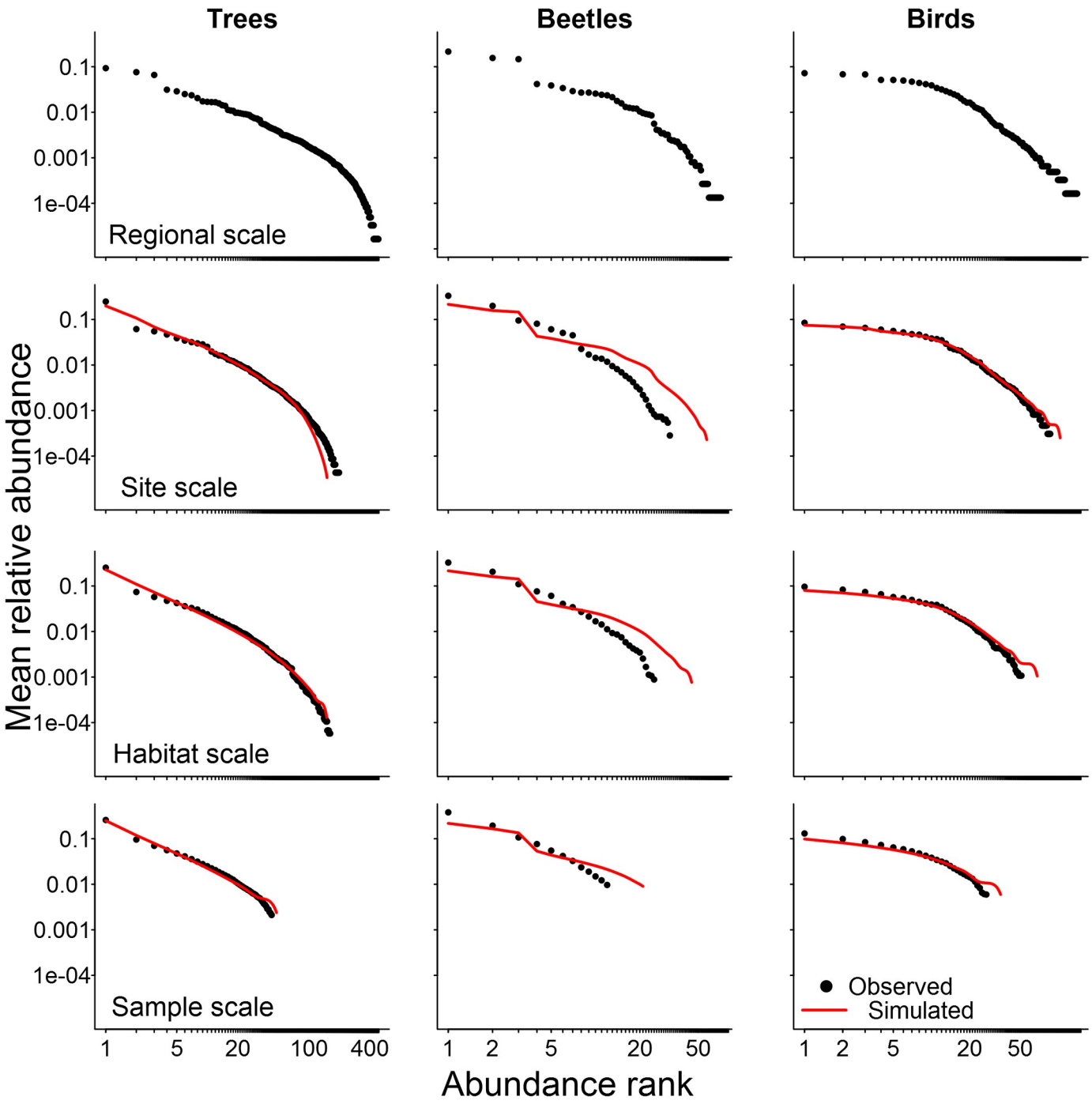

Fig 1. Disentangling sampling scale and sample size effects in Species Abundance Distributions (SAD) using counts of individuals. Observed (black circles) and simulated (red lines) are the mean relative abundance of species in each rank of abundance for trees, beetles and birds at four sampling spatial scales. For all scales except for the regional one, we present the expected values of simulated samples taken from the observed SAD at the upper scale but with the same mean sample size observed at the scale immediately below (see details on S2 Appendix in S1 File). Samples were simulated assuming conspecific aggregation (Negative binomial sampling) for trees and complete spatial randomness (Poisson sampling) for birds and beetles.

https://doi.org/10.1371/journal.pone.0238854.g001

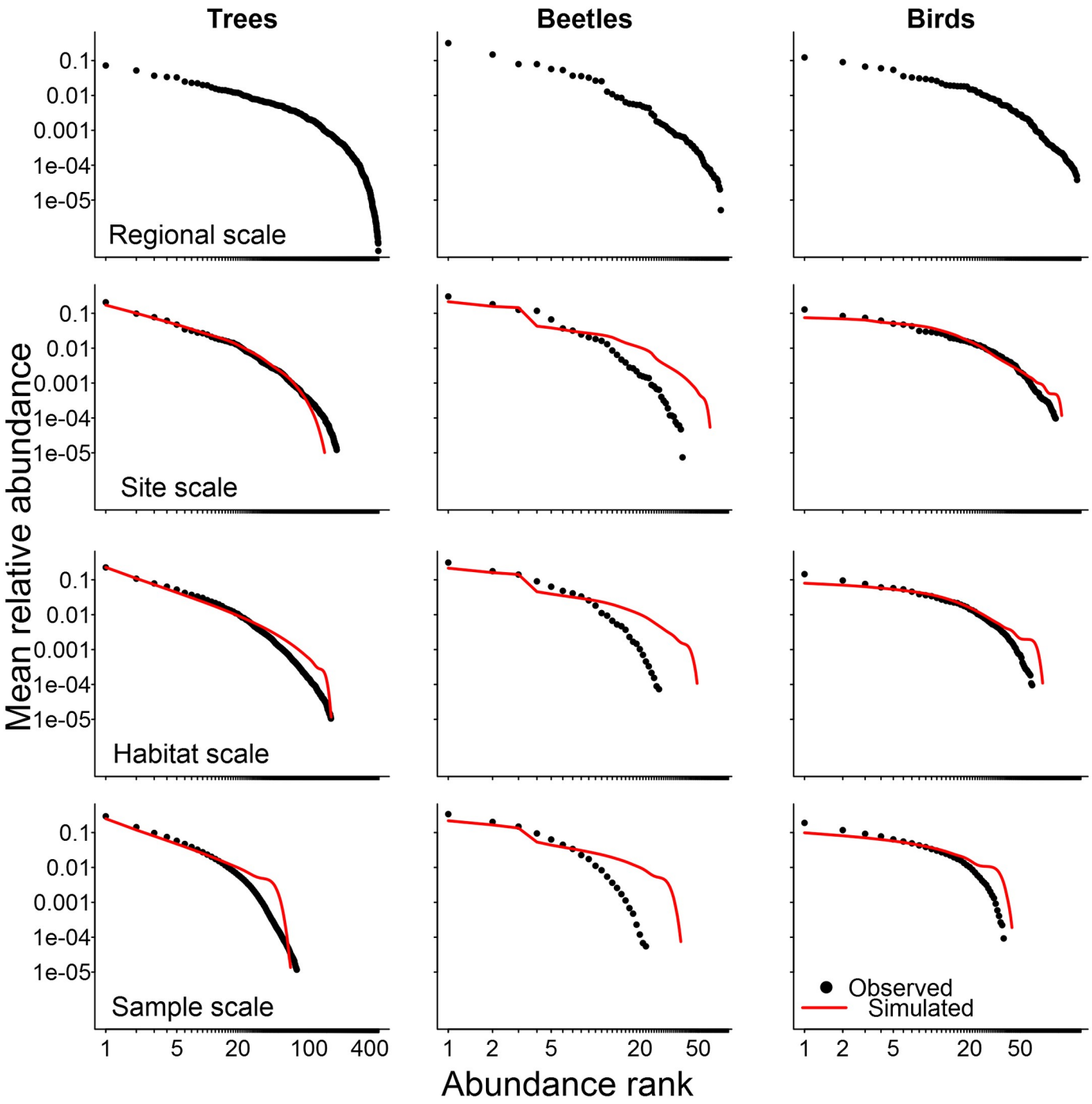

Fig 2. Disentangling sampling scale and sample size effects in Species Abundance Distributions (SAD) using biomass. As in Fig 1, observed (black circles) and simulated (red lines) are the mean relative abundance of species in each rank of abundance for trees, beetles and birds at the four sampling spatial scales. Other details are provided in the caption of Fig 1.

https://doi.org/10.1371/journal.pone.0238854.g002

Table 1. Summary of the Kolmogorov-Smirnov tests comparing the observed and simulated SADs.

Simulation scale	Organism	Reference scale					
		Counts of individuals			Biomass		
		Regional	Site	Habitat	Regional	Site	Habitat
Site	Trees	0.815			0.370		
	Beetles	1			1		
	Birds	0.999			1		
Habitat	Trees	0.993	0.977		0.581	0.519	
	Beetles	1	1		1	1	
	Birds	0.995	0.541		0.999	1	
Sample	Trees	0.982	0.999	**0.000**	**0.013**	0.754	0.945
	Beetles	1	1	0.973	1	0.999	1
	Birds	0.999	0.164	**0.000**	0.733	0.999	0.571

Values correspond to the proportion of the 10000 simulated SADs being regarded as equal (p-value $\geq$0.05) to the mean observed SAD for each reference sampling scale. Values in bold mean that observed and simulated SADs are different (α = 5%). For each reference sampling scale (*i.e.*, regional, site and habitat), we simulated communities using the same number of species and relative densities but with a mean sample size of the corresponding scale of simulation (*i.e.*, site, habitat and sample).

https://doi.org/10.1371/journal.pone.0238854.t001

probability of having rare species in the sample [9, 11, 12]. The similarity of trends found for the abundance and biomass of distinct organisms, such as trees, beetles, and birds, reinforces this conclusion. Besides the indirect effect of sample size, the shape of SADs is also influenced by conspecific aggregation, which tends to increase the number of both rare and very common species in the sampled SAD [15, 17]. In our study, aggregation was important for trees (S3 Fig in S1 File), confirming that tropical tree species are often spatially aggregated [40, 41]. This also suggests that the role of conspecific aggregation in shaping SADs may be more important for sessile organisms, at least for the spatial scales considered here. And if we consider that the patterns of species aggregation depend on the scale considered [35, 42], then conspecific aggregation may even be regarded as an indirect effect of the sampling spatial scale.

Although sample size is the preponderant effect on the SAD, we found some deviations between observed and simulated SADs that could not be attributed to sample size or conspecific aggregation. Thus, there is evidence of a sampling scale effect, which is related to the beta diversity of the metacommunity [19]. Beta diversity depends on the correlation of species abundances across communities, which in turn defines the joint SAD of two or more communities as a multivariate probability distribution [31, 32]. Taking two communities as a simple example, there is a small probability that a given species is very abundant or very rare in both samples when beta diversity is high (negative correlation, inset plot in Fig 3a). This probability is higher when beta diversity is low (positive correlation, inset plot in Fig 3b). Therefore, when we combine samples the resulting SAD will have larger variance as beta diversity increases, because the proportion of rare species in the pooled SAD increases. In our study, beta diversity was higher for beetles at all sampling scales (probably due to the larger extent of the sampling coverage for this group), which may explain the deviations between observed and simulated SADs. Moreover, conspecific aggregation also increases the variance in the correlation between samples and of the pooled SAD (Fig 3c and 3d – see S4 Fig in S1 File for the same results presented using Preston plots). Unlike the effect of sample size and species aggregation, which can be efficiently handled using sampling theory, beta diversity is more related to ecological and evolutionary processes, such as habitat partitioning, dispersal limitation, biogeography and disturbance history. Since these processes are hardly perceived at very small scales (*e.g.* [3]), beta diversity will probably affect the shape of SADs at larger spatial scales, by

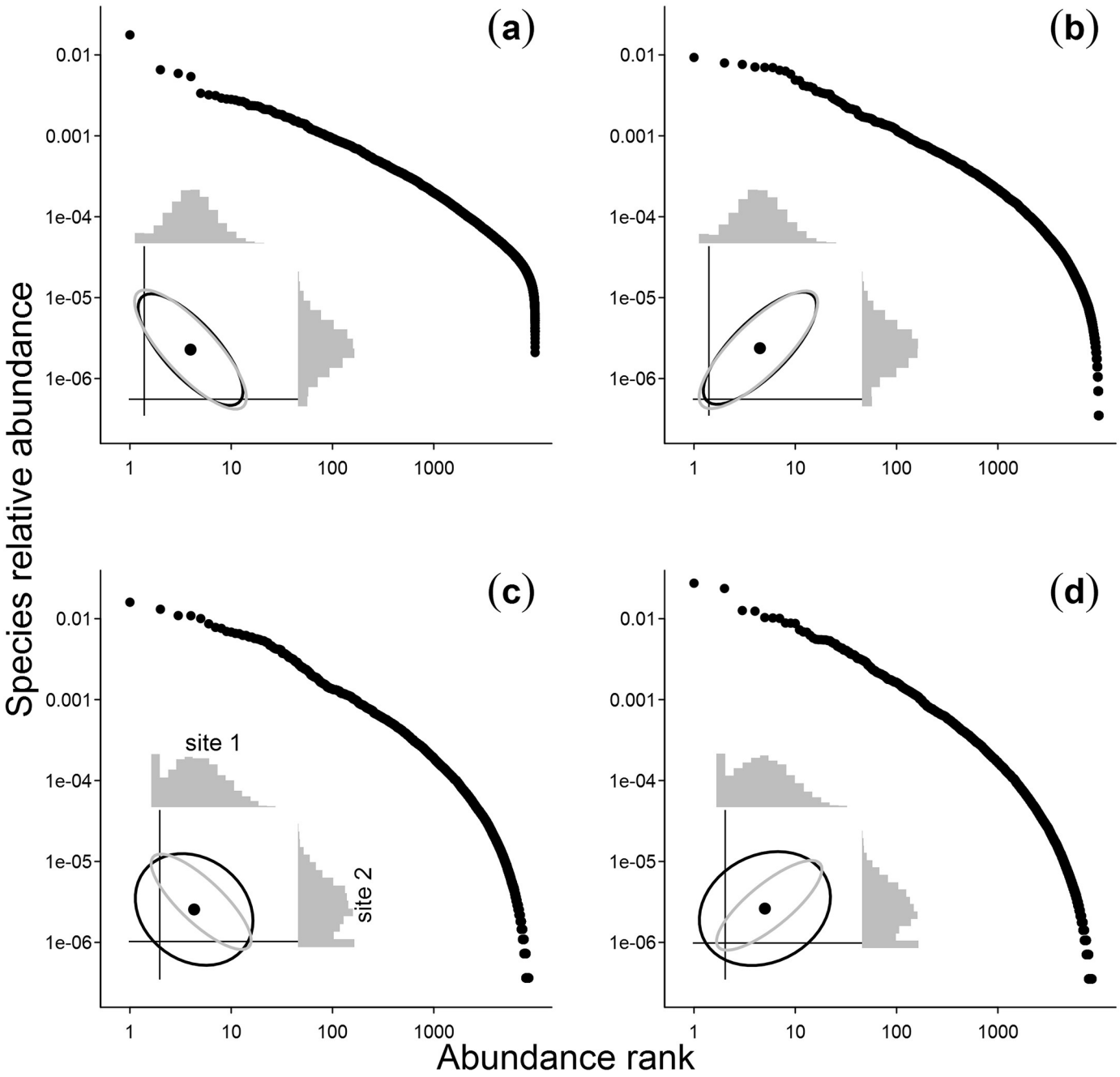

Fig 3. The influence of conspecific aggregation and of the correlation of species abundances among communities (beta diversity) in the shape of the SAD, while combining samples to increase the sampling scale. Here, we exemplify this influence by pooling together two samples taken from a bivariate Lognormal distribution with the exact same parameters and controlling for possible sample size effects. We present the resulting rank-abundance plots (on log-log scale) from the combination of sites where individuals of all species are randomly distributed and beta diversity is high (a) or low (b) and when there is conspecific aggregation with high (c) or low beta diversity (d). In each panel, the inset plot shows the ellipses that encompass 95% of the bivariate distributions and the marginal Preston plots of the sample SADs. Gray ellipses delimit the joint distribution of abundances in the communities and black ellipses delimit the distribution of abundances in the samples.

https://doi.org/10.1371/journal.pone.0238854.g003

increasing relative contribution of rare species (Fig 3 and S4 Fig in S1 File). Thus, beta diversity may explain the increase in the contribution of rare species at very large spatial scales found by [17, 20, 43].

The use of multivariate distributions to describe joint SADs was advanced by [7]. By assuming no conspecific aggregation and a Lognormal underlying SAD, they used a bivariate Poisson-Lognormal distribution to describe joint SADs where (i) communities can have different underlying SAD parameters, (ii) with different correlation between species abundances, (iii) which will present a sampling size effect on the SAD of the pooled samples. Sampling decreases the correlation between species abundances, an effect that is stronger if there is conspecific aggregation (gray and black ellipses in the inset plots of Fig 3 and S4 Fig in S1 File). Parametric multivariate distributions are still hard to fit, especially if we want to take into account spatial aggregation. So far, the existing investigations are based on simulations and they have shown that the similarity between sampled communities is important [19, 23]. However, these investigations were not fully parametric and/or they were based on presence-absence similarity indexes. Here we show that the simulation from parametric multivariate distributions with an explicit correlation between species abundances is feasible, as exemplified in Fig 3 (see S4 Appendix in S1 File for details), and thus one could use simulation-based methods of inference [44] to fit and compare multivariate SADs across space and time. We also present different ways of presenting pooled SADs from samples taken at different spatial scales (S5 Appendix in S1 File).

As found by [18, 22], SADs built using biomass instead of counts of individuals are more symmetrical in Preston plots, a phenomenon termed 'differential veiling' by [45]. Here, we attribute this difference to the fact that biomass SADs are essentially the distribution of sums of biomass values (i.e., rare species will not necessarily have the lowest biomass ranks). If we take biomass as random variables, then the higher symmetry of the biomass SAD can be the product of a weak variant of the Central Limit Theorem: the sum of independent variables will tend to a normal distribution [46]. It is reasonable to assume that biomass values are independent within species, but biomass distributions are probably not identically among species. Thus, biomass SAD will tend to a more symmetrical distribution, but more slowly and deviating from normality. This is a different explanation than those provided by [45, 47], both relying on differential consequences of sampling individuals (discrete) and biomasses (continuous variable). Ecologically speaking, the higher symmetry of the biomass SAD suggests that the dominance in ecological communities is lower than previously thought. If biomass expresses better the partitioning of resources among species [48], a few but large individuals could suffice to maintain populations, especially regarding size-structured organisms. And since growth is more frequent than births and deaths, the biomass SADs is more variable and may be more effective to capture changes on community structure [18, 47].

Conclusions

Sample size had a much greater influence than sampling spatial scale on the shape of SADs. Although sample size was the leading effect, we advocate that the shape of SADs is influenced by three factors, in decreasing order of importance: sample size [9, 13], conspecific spatial aggregation [15] and beta diversity [19, 23]. The changes related to the first two factors can be taken into account using an approach based on univariate distributions, while the third may require the use of multivariate distributions [31, 32]. Analytical solutions for such distributions may be hard to find or to deal with, and simulation-based methods of inference offer a more feasible alternative for most ecologists. The present study and some previous work [19, 23] showed how computer simulations can disentangle the 'process and observation layers' [49]

that influence the shape of SADs (see [28] for an rarefaction-based approach to disentangle similar effects in the number of species in samples). Future studies comparing results with different underlying SADs, species aggregation patterns and beta diversity scenarios are welcome to refine our understanding of the effects of sample size, conspecific aggregation and beta diversity on SADs and how such effects interact with existing explanations of the shape of SADs based on species dispersal distances [2, 17], species-range distributions [34, 50] or species strategies (e.g. core-occasional species hypothesis [51]). In addition, there is still uncertainty if these hypotheses apply similarly to numerical abundance and biomass [18, 52]. Thus, once the effects of sample size and species aggregation have been taken into account, it is possible to explore which are the mechanisms driving the shape of SADS and more broadly the diversity of ecological communities in space and time. An assessment of how the relative role of such mechanisms may change across taxa, environmental gradients and spatial scales is still needed, so there is still plenty of room to explore the ecological determinants of SADs [29, 30].

Supporting information

S1 File.
(DOCX)

S2 File.
(R)

S1 Dataset.
(CSV)

Acknowledgments

We thank Vinícius C. Souza and Fernando Z. Vaz-de-Mello for the identification of tree and beetle species, respectively. We also thank Hélène Morlon and Jerôme Chave for their valuable suggestions on an early version of this manuscript.

Author Contributions

Conceptualization: Renato A. Ferreira de Lima, Paula Alves Condé, Paulo I. Prado.

Data curation: Renato A. Ferreira de Lima, Paula Alves Condé, Cristina Banks-Leite, Renata C. Campos, Malva I. Medina Hernández, Ricardo R. Rodrigues.

Formal analysis: Renato A. Ferreira de Lima, Paulo I. Prado.

Funding acquisition: Renato A. Ferreira de Lima, Cristina Banks-Leite, Ricardo R. Rodrigues.

Methodology: Renato A. Ferreira de Lima, Paulo I. Prado.

Project administration: Ricardo R. Rodrigues, Paulo I. Prado.

Writing – original draft: Renato A. Ferreira de Lima, Paulo I. Prado.

Writing – review & editing: Renato A. Ferreira de Lima, Paula Alves Condé, Cristina Banks-Leite, Renata C. Campos, Malva I. Medina Hernández, Ricardo R. Rodrigues, Paulo I. Prado.

References

1. Pielou EC. Ecological Diversity. New York: John Wiley and Sons; 1975.

2. Chisholm RA, Lichstein JW. Linking dispersal, immigration and scale in the neutral theory of biodiversity. Ecol Lett. 2009; 12: 1385–1393.
3. Cheng J, Mi X, Nadrowski K, Ren H, Zhang J, Ma K. Separating the effect of mechanisms shaping species-abundance distributions at multiple scales in a subtropical forest. Oikos. 2012; 121: 236–244.
4. MacArthur RH. On the relative abundance of species. Am Nat. 1960; 94: 25–36.
5. Tokeshi M. Species coexistence: ecological and evolutionary perspectives. Oxford: Blackwell; 1999.
6. Routledge RD. The form of species-abundance distributions. J Theor Biol. 1980; 82: 547–558.
7. Engen S, Lande R. Population dynamic models generating the Lognormal species abundance distribution. Math Biosci. 1996; 132: 169–183.
8. Hubbell SP. The unified neutral theory of biodiversity and biogeography. Princeton: Princeton University Press; 2001.
9. Preston FW. The commonness, and rarity, of species. Ecology. 1948; 29: 254–283.
10. Nekola JC, Šizling AL, Boyer AG, Storch D. Artifactions in the log-transformation of species abundance distributions. Folia Geobot. 2008; 43: 259–268.
11. Dewdney AK. A general theory of the sampling process with applications to the ''veil line''. Theor Popul Biol. 1998; 54: 294–302.
12. Chisholm RA. Sampling species abundance distributions: Resolving the veil-line debate. J Theor Biol. 2007; 247: 600–607.
13. Preston FW. The canonical distribution of commonness and rarity: part I. Ecology. 1962; 43: 185–215.
14. Bulmer MG. On fitting the Poisson Lognormal distribution to species-abundance data. Biometrics. 1974; 30: 101–110.
15. Green JL, Plotkin JB. A statistical theory for sampling species abundances. Ecol Lett. 2007; 10: 1037–1045.
16. Alonso D, Ostling A, Etienne RS. The implicit assumption of symmetry and the species abundance distribution. Ecol Lett. 2008; 11: 93–105.
17. Borda-De-Água L, Whittaker RJ, Cardoso P, Rigal F, Santos AMC, Amorim IR, et al. Dispersal ability determines the scaling properties of species abundance distributions: A case study using arthropods from the Azores. Sci Rep. 2017; 7: 1–13.
18. Morlon H, et al. Taking species abundance distributions beyond individuals. Ecol Lett. 2009; 12: 488–501.
19. Šizling AL, Storch D, Šizlingová E, Reif J, Gaston KJ. Species abundance distribution results from a spatial analogy of central limit theorem. P Natl Acad Sci USA. 2009; 106: 6691–6695.
20. Rosindell J, Cornell SJ. Universal scaling of species-abundance distributions across multiple scales. Oikos 2013; 122: 1101–1111.
21. Wilson JB, Gitay H, Steel JB, King WM. Relative abundance distributions in plant communities: effects of species richness and of spatial scale. J Veg Sci. 1998; 9: 213–220.
22. Connolly SR, Hughes TP, Bellwood DR, Karlson RH. Community structure of corals and reef fishes at multiple scales. Science. 2005; 309: 1363–1365.
23. McGill BJ. Does Mother Nature really prefer rare species or are log-left-skewed SADs a sampling artefact? Ecol Lett. 2003; 6: 766–773.
24. Palmer MW, White PS. Scale dependence and the species-area relationship. Am Nat. 1994; 144: 717–740.
25. Nekola JC, White PS. The distance decay of similarity in biogeography and ecology. J Biogeogr. 1999; 26: 867–878.
26. Zillio T, He F. Inferring species abundance distribution across spatial scales. Oikos. 2010; 119: 71–80.
27. Borda-de-Água L, Borges PAV, Hubbell SP, Pereira HM. Spatial scaling of species abundance distributions. Ecography. 2012; 35: 549–556.
28. McGlinn DJ, Xiao X, May F, Gotelli NJ, Engel T, Blowes SA, et al. Measurement of Biodiversity (MoB): A method to separate the scale-dependent effects of species abundance distribution, density, and aggregation on diversity change. Methods Ecol Evol. 2019; 10: 258–269.
29. McGill BJ, Nekola JC. Mechanisms in macroecology: AWOL or purloined letter? Towards a pragmatic view of mechanism. Oikos 2010; 119: 591–603.
30. Yen JDL, Thomson JR, Mac Nally R. Is there an ecological basis for species abundance distributions? Oecologia. 2013; 171: 517–525.
31. Sæther BE, Engen S, Grøtan V. Species diversity and community similarity in fluctuating environments: parametric approaches using species abundance distributions. J Anim Ecol. 2013; 82: 721–738.

32. Engen S, Lande R, Walla TR, DeVries PJ. Analyzing spatial structure of communities using the two-dimensional Poisson Lognormal species abundance model. Am Nat. 2002; 160: 60–73.

33. Manly BFJ. Randomization, bootstrap and Monte Carlo methods in biology. 3rd ed. London: Chapman and Hall/CRC Press; 2006.

34. Brown JH, Mehlman DW, Stevens GC. Spatial variation in abundance. Ecology. 1995; 76: 2028–2043.

35. Taylor LR, Woiwod IP, Perry JN. The density-dependence of spatial behaviour and the rarity of randomness. J Anim Ecol. 1978; 47: 383–406.

36. Newman M. Power laws, Pareto distributions and Zipf's law. Contemp Phys. 2005; 46: 323–51.

37. Williamson M, Gaston KJ. The lognormal distribution is not an appropriate null hypothesis for the species-abundance distribution. J Anim Ecol. 2005; 74: 409–422.

38. Prado PI, Miranda MD. sads: Maximum likelihood models for species abundance distributions. R package version 0.1.10. 2014.

39. Arnold TB, Emerson JW. Nonparametric goodness-of-fit tests for discrete null distributions. R Journal; 2011; 3: 34–39. Available from: http://journal.r-project.org/archive/2011-2/RJournal_2011-2_Arnold+Emerson.pdf

40. Condit R, et al. Spatial patterns in the distribution of tropical tree species. Science. 2000; 288: 1414–1418.

41. Capretz RL, Batista JLF, Sotomayor JFM, Cunha CR, Nicoletti MF, Rodrigues RR. Spatial Pattern of Four Forest Communities by Ripley'S K Function in Sao Paulo State. Cienc Florest. 2012; 22: 552–565.

42. Fortin MJ, Dale MRT. Spatial analysis: a guide for ecologists. Cambridge: Cambridge University Press; 2005.

43. Hubbell SP, He F, Condit R, Borda-de-Agua L, Kellner J, ter Steege H. How many tree species are there in the Amazon and how many of them will go extinct? Proc Natl Acad Sci. 2008; 105: 11498–11504.

44. Hartig F, Calabrese JM, Reineking B, Wiegand T, Huth A. Statistical inference for stochastic simulation models–theory and application. Ecol Lett. 2011; 14: 816–827.

45. Williamson M. Why do species abundance distributions of individuals and of biomass behave differently under sampling? Oikos. 2010; 119: 1697–1699.

46. Johnson N, Kotz S, Balakrishnan N. Continuous univariate distributions. vol. 1. New York: John Wiley and Sons; 1994.

47. Anderson BJ, Chiarucci A, Williamson M. How differences in plant abundance measures produce different species-abundance distributions. Meth Ecol Evol. 2012; 3: 783–786.

48. Saint-Germain M. et al. Should biomass be considered more frequently as a currency in terrestrial arthropod community analysis? J App Ecol. 2007; 44: 330–339.

49. Royle JA, Dorazio RM. Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities. San Diego: Academic Press; 2008.

50. McGill BJ, Collins C. A unified theory for macroecology based on spatial patterns of abundance. Evol Ecol Res. 2003; 5: 469–492.

51. Magurran AE, Henderson PA. Explaining the excess of rare species in natural species abundance distributions. Nature. 2003; 422: 714–716.

52. Henderson PA, Magurran AE. Linking species abundance distributions in numerical abundance and biomass through simple assumptions about community structure. P Roy Soc B: Biol Sci. 2010; 277: 1561–1570.