



Spatial scaling of species abundance distributions

Luís Borda-de-Água, Paulo A. V. Borges, Stephen P. Hubbell and Henrique M. Pereira

Luís Borda-de-Água (lbagua@gmail.com) and H. M. Pereira, Centro de Biologia Ambiental, Faculdade de Ciências, Univ. de Lisboa, PT-1749-016 Lisboa, Portugal, HMP also at: Depto de Engenharia Civil e Arquitectura, Inst. Superior Técnico, PT-1049 001 Lisboa, Portugal. – P. A. V. Borges, Azorean Biodiversity Group (CITA-A), Univ. dos Açores, Angra do Heroísmo, PT-9700 Angra do Heroísmo, Azores, Portugal. – S. P. Hubbell, Dept of Ecology and Evolutionary Biology, Univ. of California, Los Angeles, CA 90095, USA, and Center for Tropical Forest Science, Smithsonian Tropical Research Inst., Unit 0948, APO AA 34002-0948, Republic of Panama.

Species abundance distributions are an essential tool in describing the biodiversity of ecological communities. We now know that their shape changes as a function of the size of area sampled. Here we analyze the scaling properties of species abundance distributions by using the moments of the logarithmically transformed number of individuals. We find that the moments as a function of area size are well fitted by power laws and we use this pattern to estimate the species abundance distribution for areas larger than those sampled. To reconstruct the species abundance distribution from its moments, we use discrete Tchebichef polynomials. We exemplify the method with data on tree and shrub species from a 50 ha plot of tropical rain forest on Barro Colorado Island, Panama. We test the method within the 50 ha plot, and then we extrapolate the species abundance distribution for areas up to 5 km². Our results project that for areas above 50 ha the species abundance distributions have a bimodal shape with a local maximum occurring for the singleton classes and that this maximum increases with sampled area size.

Understanding the processes that determine the relative abundance of species in a community is a central task in ecology (Brown 1995, Lawton 2000, Hubbell 2001). Usually, the information on species relative abundances is conveyed through the histogram of the number of species with a given number of individuals, the species abundance distribution (hereafter SAD). Not surprisingly, SADs have played a major role in the development of theories of biodiversity and biogeography (McGill et al. 2007).

There have been several approaches to the study of SADs. Some authors have put the emphasis on which probability density function best fits the SAD at one given spatial scale (McGill 2003a, Volkov et al. 2003, Williamson and Gaston 2005). Others, stemming from the original work by Fisher et al. (1943), have looked at SADs at different scales by considering that an observed SAD is a sample from the SAD of a larger regional pool of species. For example, Fisher arrived at his well known logseries distribution by assuming that the observed abundance distribution of Lepidoptera species in light traps resulted from Poisson sampling from a gamma distribution. Later, Preston (1948, 1962) questioned the generality of Fisher's logseries arguing that it was an artifact of small sample size, and that when more data were gathered the lognormal distribution gave a better fit. Preston described the evolution of the logseries to the lognormal as the shift to the left of the distribution of a veil line that progressively reveals more (rarer) species as sample size increases. In an important paper, Dewdney (1998) reassessed Preston's suggestions and showed that the veil line is not an appropri-

ate explanation for the shape of sampled distributions from lognormals. In the same vein, McGill (2003b) studied the shape of sampled distributions and concluded that their shapes and, in particular, the degree of (left-)skewness, was a function of sample size, in accordance with previous work by Gregory (1994, 2000). Recently, Alonso and McKane (2004) and Etienne and Alonso (2005) have studied the characteristics of sampled SADs under the framework of Hubbell's neutral theory, the former assuming random sampling and the later dispersal limitation. Dispersal limitation is a possible cause, but not the only one, for species aggregation. Recognizing the importance of aggregation, Green and Plotkin (2007) introduced a general framework to describe SADs assuming spatially aggregated species or, equivalently, heterogeneity in the sampling scheme. The framework of Green and Plotkin generalized the work by Dewdney (1998) who had considered only random sampling.

An important result that has emerged from these collected works (Dewdney 1998, Alonso and McKane 2004, Etienne and Alonso 2005, Green and Plotkin 2007) is that under random sampling the shape of the sampled (local) SADs is like that of the regional SAD, but that for nonrandom sampling or spatially aggregated species one expects the local SADs to exhibit often highly variable shapes. Studying the effects of spatial aggregation is warranted because there is good evidence that almost all species are clumped in distribution in nature (Condit et al. 2000). Empirical data, simulations, and theoretical developments reveal that SADs change their shapes when sample size changes. For example, consider

changes in the SAD for trees and shrubs in the 50 ha plot on Barro Colorado Island. The SAD for areas between 1 and 50 ha, plotted on a logarithmic scale of the number of individuals, changes from a monotonically decreasing function to a distribution with a maximum for intermediate abundance classes (Fig. 1). Using simulations based on the assumptions of the neutral theory and modeling different levels of aggregation, Borda-de-Água et al. (2007) obtained similar findings for small spatial scales, and, in addition, they showed that for very large scales the SAD is again a monotonically decreasing function. This result was predicted analytically by Hubbell's neutral theory (Hubbell 2001), which shows that Fisher's logseries describes the SADs of metacommunities (all trophically similar individuals in a biogeographic region, such as the Amazon basin (Hubbell et al. 2008)) and that the zero sum multinomial, a distribution with a shape similar to that of the lognormal, describes SADs of local communities (a subset of a metacommunity at scales where individuals interact). However, an important difference between the zero-sum multinomial and the lognormal distribution is that the former allows for a larger number of species in the low abundance classes. In fact, a large number of data sets show an excess of singletons when compared to those predicted by the lognormal (e.g. Fig. 1, and Volkov et al. 2003). As we will discuss, singletons play an important role in determining the shape of the SADs: monotonically decreasing or bimodal.

Beyond understanding the shape of sampled distributions, ecologists are also interested in the inverse problem, that is, how to predict the SAD at large scales based on samples from smaller scales. In an interesting set of papers, Šizling et al.

(2009a, b), and Kúrka et al. (2010) have shown that under general conditions the SADs converge to a specific distribution, and the convergence and final shape of the distribution is given by the Jaccard index–area relationship and the spatial autocorrelation of the abundances–area relationship. Their results stem from a generalization of the Central Limit theorem (Šizling et al. 2009a) and they have shown how the SAD of a given area can be obtained from the SADs of subplots (Šizling et al. 2009b). Zillio and He (2010) also addressed this problem using a Bayesian approach, and Harte et al. (2009) using maximum entropy methods, but the latter assuming only logseries SADs. Here we also introduce a method to extrapolate SADs to larger spatial scales but using a very different approach based on the scaling properties of the moments of SADs.

The premise of this paper is that there is not an inherent scale at which the SAD should be measured or described. We argue, instead, that attention should shift to the scaling properties of the distribution as a function of area. To understand and describe the scaling properties is important for at least two reasons. First, it may reveal intrinsic biological characteristics of the system and, second, it may allow extrapolation of the SAD to larger areas. This paper deals with the latter aspect with a view to providing tools for estimating the relative abundance of species at large spatial scales.

In order to study the scaling characteristics of SADs, we use the moments of the logarithmically transformed number of individuals. The choice of the moments to characterize the distribution and its scaling properties was motivated by the paper by Borda-de-Água et al. (2002). However, as we

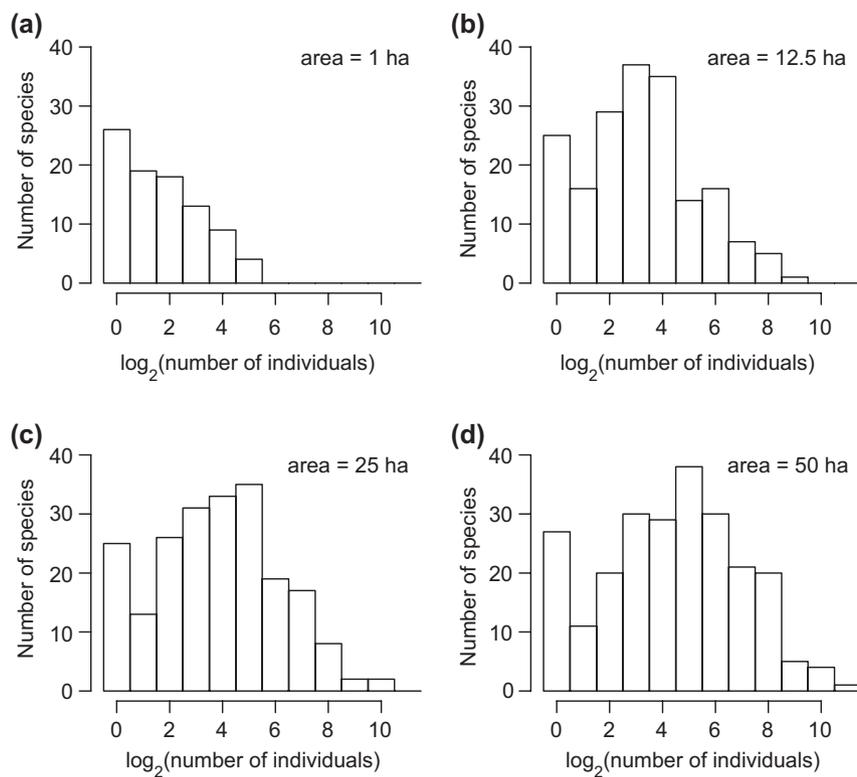


Figure 1. The species abundance distribution for 4 different sizes of area within the BCI 50 ha plot. Bins are centered on the logarithms of the powers of 2 and delimited by $\log(2^{n \pm 0.5})$.

explain in the Methods section, the direct application of the moments to reconstruct the distribution function is not efficient, therefore, we use a method based on Tchebichef moments and polynomials (Mukundan et al. 2001).

An important prediction arising from this work is that the number of the rarest species (singletons, doubletons, etc.) increases with sample sizes in such a way that the distribution becomes clearly bimodal, with one of the maxima occurring for the singleton class and the other for intermediate abundance classes. This shape is compatible with SADs reported in other works (Magurran and Henderson 2003) and it is important because it quantifies the role of rarity, in particular that of the singletons, in the diversity of tree rainforest communities (Rabinowitz 1981, Pitman et al. 1999).

Methods

Using moments to reconstruct the probability density function

We use the moments of the distribution of the logarithmically transformed number of individuals to characterize the scaling behavior of the SAD and to extrapolate and reconstruct it at larger spatial scales. This is possible because the characteristic function connects the moments of a distribution to the probability density function; the moments are the terms of the Maclaurin expansion of the characteristic function which, in turn, is the Fourier transform of the probability density function (Feller 1971). However, such a procedure is rarely practical, as we explain below and, instead, we use a method based on discrete scaled orthogonal Tchebichef polynomials and moments introduced by Mukundan et al. (2001). We now formalize this discussion.

Given a community composed of S species, of which the j th species has X_j individuals, we estimate the moment of order n , M_n , of the species abundance distribution of the \log_2 transformed number of individuals, $x_j = \log_2(X_j)$, by using the relationship

$$M_n = \frac{1}{S} \sum_{j=1}^S x_j^n, \quad (1)$$

from where $M_0 = 1$. Knowing the moments, we can, in principle, reconstruct the probability density function because the moments are the coefficients of the Maclaurin expansion of the characteristic function, $\varphi_X(t)$ (Hu 1962, Feller 1971), that is,

$$\varphi_X(t) = M_0 + itM_1 + \frac{(it)^2}{2!}M_2 + \frac{(it)^3}{3!}M_3 + \dots + \frac{(it)^p}{p!}M_p + \dots$$

where i is the imaginary number. Since the characteristic function is the Fourier transform of the probability density function, $f(x)$, the latter can be recovered by applying the inverse Fourier transform to $\varphi_X(t)$. However, the above procedure is not useful because it requires a very large number of moments (Teague 1980). For example, in our numerical experiments to reproduce a normal distribution with mean 10 and standard deviation 2, we estimated about 40 moments. This, compounded with the need to extrapolate moments, as we explain later, led us to explore other methods.

Here we use a method based on the scaled discrete orthogonal Tchebichef polynomials and moments introduced by Mukundan et al. (2001). The basic idea of the method is to fit a given discrete function $f(x)$ defined in N points with the sum of scaled discrete orthonormal Tchebichef polynomials, $\tilde{T}_n(x)$, weighted by the Tchebichef moments, T_n , that is,

$$f(x) = \sum_{n=0}^{N-1} T_n \tilde{T}_n(x). \quad (2)$$

and the following inequalities apply: $0 \leq n \leq N-1$ and $0 \leq x \leq N-1$. We provide the formulas to calculate $\tilde{T}_n(x)$ and T_n in Supplementary material Appendix 1.

Although we could estimate the Tchebichef moments directly from the data (Mukundan et al. 2001), such procedure would not be useful for our purposes because the scaling properties that are useful to extrapolate the SAD are observed for the moments, M_n , and not the Tchebichef moments, T_n . Fortunately, one can relate the Tchebichef moments, T_n , to the moments, M_n (Supplementary material Appendix 1, Eq. A1). In summary, the method to approximate a distribution using the Tchebichef moments and polynomials consists of the following steps: 1) estimate the moments using Eq. (1), 2) calculate the Tchebichef moments (Supplementary material Appendix 1, Eq. A1) and polynomials (Supplementary material Appendix 1, Eq. A2), and 3) approximate the probability density function using Eq. (2).

Data

We illustrate the application of the previous method using data on mapped populations of woody plant species (except lianas) from a 50 ha (500×1000 m) plot of old-growth tropical moist forest on Barro Colorado Island (BCI), Panama (Condit 1998, Hubbell et al. 1999, 2005). The data set contains information on the species and spatial location of all individuals with stem diameters ≥ 1 cm dbh (diameter at breast height). Here, however, we limit our analysis to stems with dbh ≥ 10 cm because if the BCI 50 ha plot will be expanded in the near future it is likely to be only for stems with dbh ≥ 10 cm, hence, it is the predictions using the dbh ≥ 10 cm class stems that are more likely to be tested. (For completeness, we show in Supplementary material Appendix 3 the results for all stems in the BCI data set, dbh ≥ 1 cm.)

Moments estimation and scaling

We estimate the moments as follows. First, for a given area smaller than 50 ha, we choose randomly the location of m (typically 20) subplots. Because the subplots are located randomly this means that some may overlap, and this is more likely to occur when the size of the subplots increase. We then log transform (using base 2) the number of individuals of each species, calculate for each subplot the moments up to a given order using Eq. (1), and calculate the average of the moments obtained from all subplots. In order to obtain the scaling properties of the moments as a function of area, we repeat this procedure for different subplot sizes (see Supplementary material Appendix 2 for details on the statistics of the moments as a function of the size of the subplot sampled).

To assess the scaling characteristics of the moments we use $\log_{10}\text{-}\log_{10}$ plots of the moments versus the area, A . As we will see, in these plots, and above a certain value of A , the moments are almost straight lines: therefore, it is reasonable to model their behavior as a set of relations of the form

$$\log(M_n(A)) = a_n + b_n \log(A), \quad (3)$$

i.e. $M_n(A) = 10^{a_n} A^{b_n}$. We can then use Eq. (3) to extrapolate the value of the moments for larger areas. The above procedure allows the determination of a single set of values (a_n , b_n) from where we can extrapolate the moments, and then estimate the Tchebichef moments and obtain the SAD. In order to obtain a final predicted SAD that corresponds to an average value and its associated confidence intervals, we repeat the above steps 100 times.

Binning method

For the low abundance classes, how species are 'binned' into \log_2 abundance classes, the binning method, affects considerably the shape of the histogram at the low abundance end of the distribution. We use the method suggested by Williamson and Gaston (2005) in which the bins are centered on the logarithms of the powers of two, that is, 0, 1, 2, et seq. (or in a linear scale, 1, 2, 4, et seq.) and with boundaries at $\log(2^{n \pm 0.5})$, that is, at 0.5, 1.5, 2.5, et seq. (or in a linear scale 1.414, 2.828, 5.656, et seq.). Hence, if S_n represents the number of species with n individuals, the first bin reports S_1 , the second S_2 , the third S_3 to S_5 , and so on. By using this scheme we make sure that the boundaries of the bins precisely double, and ensure that no number of individuals ever falls exactly on the boundaries of adjacent bins (Williamson and Gaston 2005). We also considered the binning schemes used by Hubbell (2001), where bins are for counts S_1 , S_2 to S_3 , S_4 to S_7 , et seq., and by Preston, where bins are for counts $S_1/2$, $S_1/2 + S_2/2$, $S_2/2 + S_3 + S_4/2$, $S_4/2 + S_5 + S_6 + S_7 + S_8/2$, et seq., but the results did not change qualitatively.

Results

We first consider the extrapolation of the SAD from areas smaller than 50 ha up to 50 ha. Figure 2a shows the logarithm of the moments from order 1 to 11 as a function of the logarithm of the area; there are 12 bins in the histogram, hence the maximum number of Tchebichef moments is 12, from order 0 to 11 (Supplementary material Appendix 1), and recall that $M_0(A) = 1$ (see Supplementary material Appendix 2 for the numerical values and statistics of the linear regressions). Figure 2b and c show the residuals for the moments of order 1 and 5, respectively; we show the residuals of the 5th moment because, as we will see, this is the highest order moment used in the extrapolations. Visual inspection reveals that the logarithm of the moments has a linear behavior as a function of the logarithm of the area. Nevertheless, closer inspection of the residuals obtained by least squares shows that the residuals have a clear downward trend for small areas. Therefore we restrict, conservatively, the region to estimate the slope and intercept, the scaling region,

to areas above 8 ha. Above this threshold the residuals have an oscillatory behavior, but note that the amplitude of the oscillations is very small compared to the values of the moments and that they tend to decrease for higher order moments. The breakdown of the power law relationship for small areas is similar to the one observed for the species area relationship for small areas (Hubbell 2001). For instance, Condit et al. (1996) reported for these data that below 1 ha the species area relationship is not well fitted by a power law.

Because the distribution we obtain after the application of the Tchebichef method is a probability density function, hence with bin densities adding to 1, in order to extrapolate the SAD, we first need to extrapolate the number of species, which we do by assuming a power law species area relationship. Figure 2d shows the log-log plot of the species area relationship, and the dashed line corresponds to the fit with $\log(S) = \log(c) + z \log(A)$. In addition, when extrapolating the SAD with the Tchebichef moments, we also need to have an estimation of the number of individuals of the most abundant species so that we know the number of bins required. Interestingly, we found that the logarithm of the number of individuals of the most abundant species, $\log_2 N_{max}$, also increases as a function of area in a way that is well approximated by a power law, $\log(\log_2 N_{max}) = \log(c') + z' \log(A)$, as shown in Fig. 2e.

We now extrapolate the SAD to 50 ha using the estimates of all moments obtained from linear regressions using scaling regions ranging from 8 ha to 10, 15, 20 and 25 ha. We obtained the best results when we limited the number of moments used in the projection to six because the projections were very sensitive to slight variations in the values of higher moments. This result is to be expected because the higher moments are raised to higher powers (Mukundan 2004). We show the predicted SADs in Fig. 3 and the distributions of the parameters of the regression in Supplementary material Appendix 2. In all 4 cases the predicted species abundance distribution for 50 ha has a bell shape with the maximum for intermediate classes. Observe that the shape of the SADs for the areas used to make the predictions can be considerably different from that of 50 ha, especially for the areas equal to 10 and 15 ha. As expected, when the size of the scaling region increases the quality of the fit increases as well, as can be seen visually or by calculating the difference between the predicted, f_j , and observed values, f_i , as $\varepsilon = \sum |f_j - \hat{f}_i|$, which decreases as $\varepsilon = 109.8, 93.1, 61.1$ and 40.38 when we increase the upper limit of the scaling region from 10 to 25 ha. Notice, as well, that the predicted SADs are able to track the large number of singletons.

We now extrapolate the species abundance distribution for areas hitherto not sampled using the data from the 50 ha plot up to 500 ha (Fig. 4), assuming that the moments retain their power law behavior. We use the scaling region from 8 to 50 ha, as shown in Fig. 2 and, based on our previous results, we use only moments up to order 5 (see the distributions of the parameters of the regression in Supplementary material Appendix 2). As before, we need to extrapolate the number of species and the number of individuals of the most abundant species, which we assume both cases to follow power laws, as shown in Fig. 2d and e. As expected, we can see from Fig. 4 that the maximum for intermediate classes moves towards more abundant classes when area increases. Notice,

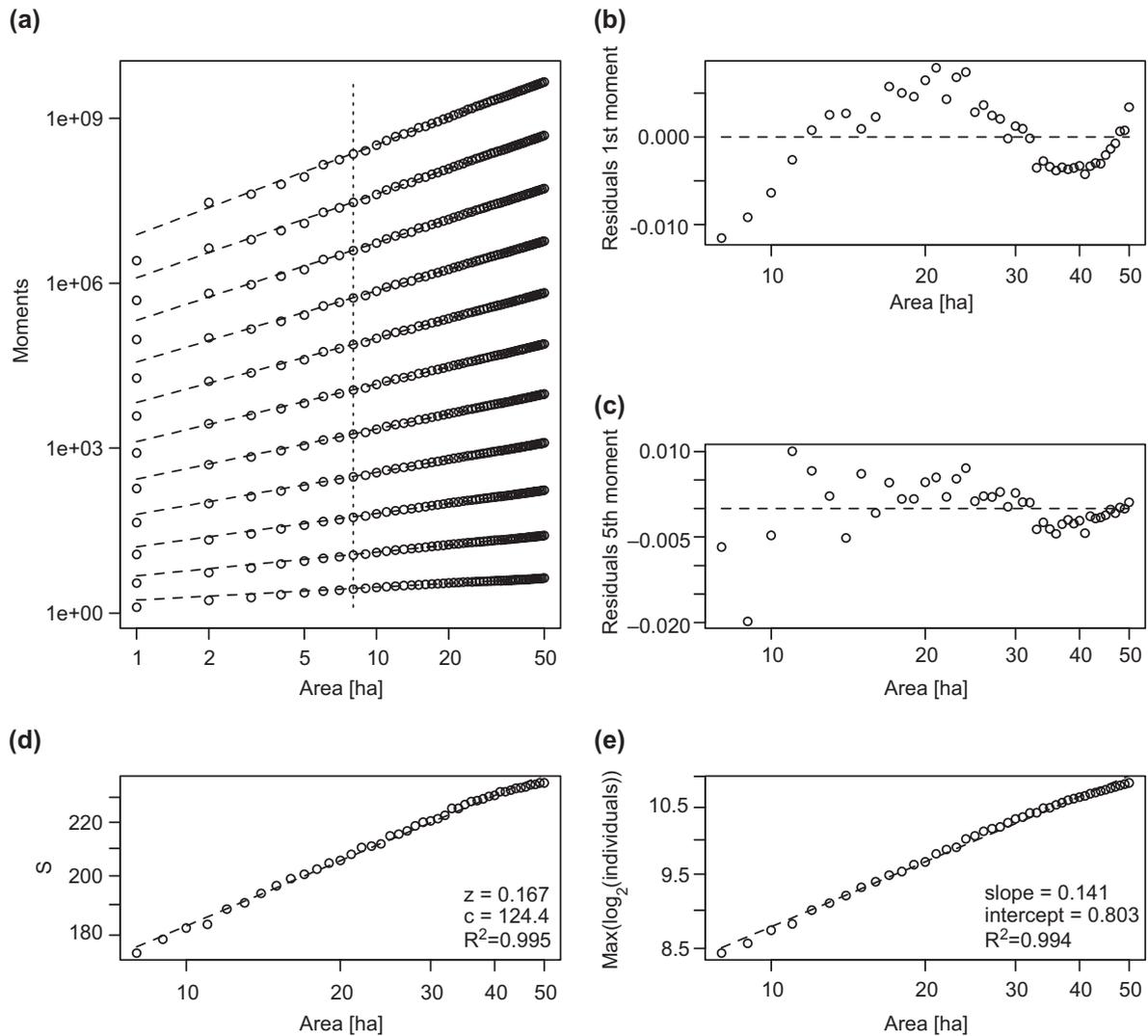


Figure 2. Plot (a) shows the logarithm of moments from 1 up to order 11 as a function of the logarithm of areas between 1 and 50 ha using data for stems with $\text{dbh} \geq 10$ cm. The order of the moments increases when we go from the bottom to the top lines. The dashed lines added to the each moment are the curves of best fit obtained from linear regressions of the logarithm of the moments versus the logarithm of the area between 8 and 50 ha. Plots (b) and (c) show the residuals from the linear regression of the logarithm of the moments of order 1 and 5 (the largest used in the reconstruction of SADs). Plot (d) is the species area relationship and the dashed line corresponds to fit obtained with least squares assuming a power law relationship, $S = cA^z$. Plot (e) shows the logarithm of the number of individuals of the most abundant species as a function of the logarithm of the area, and the dashed line was obtained by least squares assuming a power law relationship.

however, that the extrapolated distributions reveal an interesting shape characterized by a pronounced dip in the lower abundance classes, the bottom of which shifts to higher abundance classes for larger areas. The increase in the number of singletons also occurred when we imposed a curvature in the log-log evolution of the moments (Supplementary material Appendix 1 for details). Noticed that such a trend is already apparent in the SADs for subplots of the 50 ha plot, as can be observed in Fig. 1, and it reveals that the number of very rare species, in particular singletons, tend to increase when the size of the area sampled increases.

Discussion

In this paper we use the pattern exhibited by the scaling of moments of the SAD to predict how the SAD will change across a range of spatial scales. The straightforward method

of relating the moments of a distribution with the probability density function is not practical (Teague 1980), therefore we use discrete orthogonal Tchebichef polynomials and moments (Mukundan et al. 2001). We illustrated the application of the method with data on tree and shrub species from a 50 ha plot in a tropical forest, and predicted the species abundance distribution for areas ranging from 100 to 500 ha. As far as we know, a similar attempt has only been carried before by Zillio and He (2010), but using a completely different approach.

Zillio and He based their method on Bayes' theorem in order to upscale the SAD up from smaller samples. In contrast, our method uses the observed properties of the moments in order to extrapolate them and predict SADs for larger area sizes. Both methods are non-parametric because they do not assume any specific function for the SAD. Although some implicit assumptions are the same for the two methods of extrapolating the SAD, such as that the surrounding forest

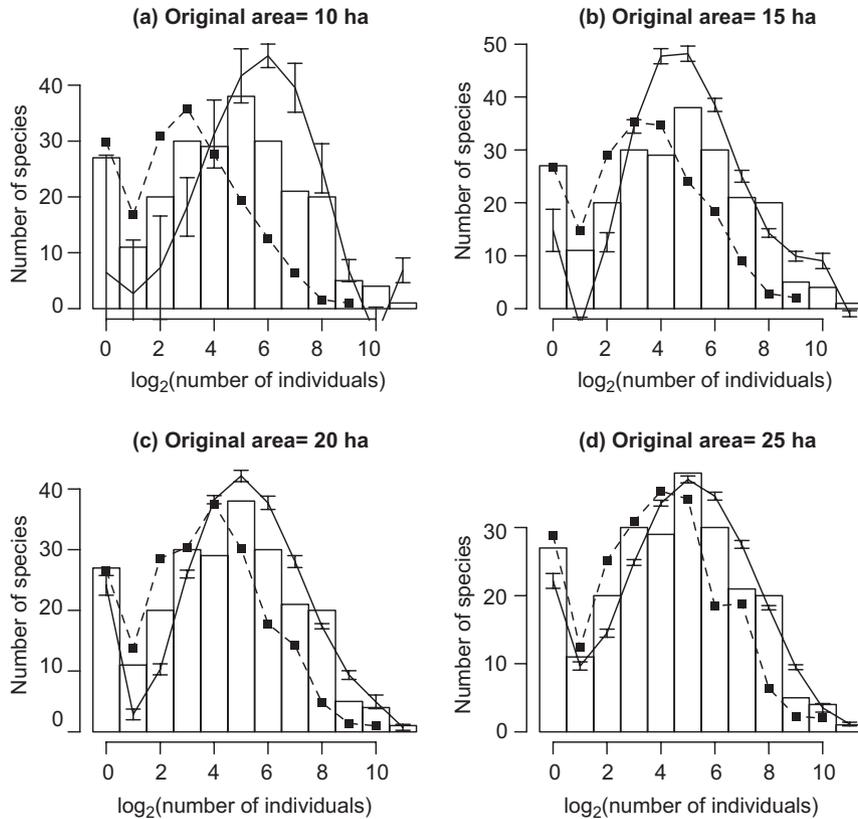


Figure 3. In all plots the histograms correspond to the species abundance distribution for all stems with $\text{dbh} \geq 10$ cm in the BCI 50 ha plot. The squared dots and dashed lines show the SAD at the original scale: plots (a, b, c, d) = 10, 15, 20, 25 ha, respectively. The continuous line is the distribution obtained with the method of Tchebichef polynomials using only the first 6 moments (orders 0 to 5) and the error bars represent ± 2 standard deviations obtained with 100 repetitions (see main text for details).

habitat has the same scaling of the species spatial distribution above 50 ha as it has within the 50 ha plot, others differ. For instance, Zillio and He describe the aggregation of the species using a negative binomial distribution and assume

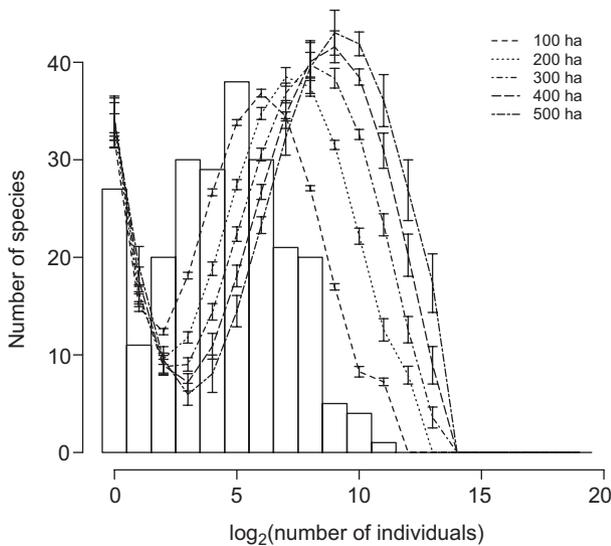


Figure 4. The histogram corresponds to the species abundance distribution for all stems in the BCI 50 ha plot with $\text{dbh} \geq 10$ cm. The 5 dashed lines are the curves obtained by the method of Tchebichef moments after extrapolation from the 50 ha plot data, and the error bars represent ± 2 standard deviations obtained with 100 repetitions (see main text for details).

that the clustering parameter is the same for all the species. This assumes that all species obey similar rules that govern their spatial dispersion, which they demonstrate is a reasonable first approximation. In contrast, our method does not assume any characteristics for the species; the present methods resides solely on the pattern exhibited by the moments, that we have modeled here as power laws. We did it because power laws gave a good fit to the spatial scales analyzed, however, if in the future, after sampling larger areas, we find that other expressions give a much better fit, then we can extrapolate the moments using these other expressions. Further improvements of our method will likely come from a better understanding of the scaling behavior of the moments and from exploring other methods for reconstructing the species abundance distribution from the moments. Along these lines, in addition to Tchebichef moments, we also explored Legendre moments (Teague 1980), but we found that the method of Tchebichef moments had two advantages. First, it does not require normalization of the data (in the Legendre method one must normalize the data to the interval from -1 to 1) and, second, and more importantly, it gave a better fit to the data, especially for the number of species in the classes of the least abundant species (see Supplementary material Appendix 1, and Mukundan et al. (2001) for other advantages of the Tchebichef moments).

An interesting result of our extrapolations for areas above 50 ha is the dip observed in the SAD among the least abundant classes. Specifically, this result predicts that the number of the rarest species increases when sampling effort increases,

which implies, in particular, that larger samples keep accumulating singleton species. Although the general evolution is the same, this specific prediction is in contrast with those of Zillio and He (2010) or the theoretical ones by Chisholm and Lichstein (2009) who predicted a reduction of singletons when area increases. However, it should be noticed from Fig. 1 that when we move from 12.5 to 50 ha the number of singletons increases and the abundance classes of species with 2 individuals and 3 to 5 individuals decrease, a trend that is not present in Chisholm and Lichstein (2009) or Zillio and He (2010). The method of Tchebichef moments simply propagated and amplified this trend for larger areas. Therefore, these shapes may indeed reveal the true evolution of the shape of the SAD on larger spatial scales.

Notice, as well, that the shapes of the extrapolated SADs are compatible with theoretical predictions (Kúrka et al. 2010) and empirical SADs reported by other researchers (Ugland and Gray 1982, Magurran and Henderson 2003, Gray et al. 2005, Dornelas and Connolly 2008). Concerning the theoretical predictions, Kúrka et al. (2010) concluded that multi modal distributions are possible when the Jaccard index is sufficiently high, in accordance with our results, because for the smallest considered area size (8 ha) the Jaccard index is approximately equal to 0.8, a value similar to that reported by Kúrka et al. Overall, the relationship between the approach by these authors based on the Jaccard-index-area and the spatial autocorrelation of abundances-area relationship and ours, is a promising direction for future research.

Concerning empirical SADs, the multi modal characteristics of some empirical SADs have led some authors to suggest more than one distribution to fit the data. For example, Magurran and Henderson (2003) modeled the species abundance distribution of the fish community of Hinkley Point, Bristol Channel, with a logseries for species that occurred infrequently in the time record, and a lognormal distribution for those that were persistent and abundant. The authors interpreted the results by the differences in the biological requirements of the persistent and abundant species (those that form the bulk of the lognormal) and of those of the rare and occasional species (those that give rise to the logseries); the latter have different habitat requirements from those where they were collected and, hence, they are rare and their presence is short lived. Gray et al. (2005) also argued that some SADs are better described by two distributions, but they suggested two lognormals, one for the rare species, the other for the common ones, and Ugland and Gray (1982) proposed that marine benthic communities are described by three lognormal distributions. More recently, Dornelas and Connolly (2008) identified a multimodal distribution with three local maxima for the SAD of coral colonies, which they fitted with Poisson lognormal distributions. Clearly, the predicted SADs by the Tchebichef moments resemble closer the two-distribution model suggested by Magurran and Henderson (2003), and their explanation may apply here as well. It is indeed likely that there is a large number of species occurring at very low abundances, such as singletons, that are only revealed when one increases the sample size. Following Magurran and Henderson (2003) interpretation, it is possible that these rare species have other habitat requirements

than those observed in BCI, therefore only establishing occasionally and being present for short periods, but possibly being abundant elsewhere (Pulliam 1988). Alternatively, these may be species sparsely distributed and, hence, rare everywhere (Pitman et al. 1999). In any case, it remains to be seen if indeed the SAD of Barro Colorado evolves to a clearly bimodal distribution with the number of singletons increasing as sampled area sizes increase.

We still know very little about the general behavior of the moments of SADs. According to our predictions for the SAD up to 500 ha (and also according to Zillio and He (2010)) it has a hump for intermediate abundance classes. On the other hand, neutral theory, predicts that for very large scales (the metacommunity) and under point mutation the SAD is a monotonically decreasing function given by Fisher's logseries. The difference between these results is likely to be due to the spatial scales involved. Indeed, we conjecture that the moments of the species abundance distribution exhibit multiple scaling regions, such as it is found for the triphasic behavior of the species area-relationship (Rosenzweig 1995, Hubbell 2001). The problem of how the moments of the SADs scale is complex and likely to be related to that of the species area-relationship. However, as explored in this paper, analyses of the moments of the species abundance distribution can be useful in practical applications, and we predict that further studies on the moments of SAD will bring insights on the assembly of ecological communities.

Acknowledgments – We thank B. McGill, and A. L. Šizling for suggestions that substantially improved the manuscript. LBA thanks the Azorean Biodiversity Group for the kind hospitality during his stay at the Univ. of Azores. This work was supported by the ABAFOBIO (PTDC/AMB/73901/2006) grant from Fundação para a Ciência e Tecnologia, Portugal. SPH thanks the National Science Foundation, the Smithsonian Tropical Research Inst., and numerous private foundations for support of both his theoretical and empirical work on and off Barro Colorado Island.

References

- Alonso, D. and McKane, A. J. 2004. Sampling Hubbell's neutral theory of biodiversity. – *Ecol. Lett.* 7: 901–910.
- Borda-de-Água, L. et al. 2002. Species-area curves, diversity indices, and species abundance distributions: a multifractal analysis. – *Am. Nat.* 159: 138–155.
- Borda-de-Água, L. et al. 2007. Scaling biodiversity under neutrality. – In: Storch, D. et al. (eds), *Scaling biodiversity*. Cambridge Univ. Press, pp. 348–375.
- Brown, J. H. 1995. *Macroecology*. – Chicago Univ. Press.
- Chisholm, R. A. and Lichstein, J. W. 2009. Linking dispersal, immigration and scale in the neutral theory of biodiversity. – *Ecol. Lett.* 12: 1385–1393.
- Condit, R. 1998. *Tropical forest census plots*. – Springer and R. G. Landes Company.
- Condit, R. et al. 1996. Species-area and species-individual relationships for tropical trees: a comparison of three 50-ha plots. – *J. Ecol.* 84: 549–562.
- Condit, R. et al. 2000. Spatial patterns in the distribution of tropical trees species. – *Science* 288: 1414–1418.
- Dewdney, A. K. 1998. A general theory of the sampling process with applications to the 'veil line'. – *Theor. Popul. Biol.* 54: 294–302.

- Dornelas, M. and Connolly, S. R. 2008. Multiple modes in a coral species abundance distribution. – *Ecol. Lett.* 11: 1008–1016.
- Etienne, R. S. and Alonso, D. 2005. A dispersal-limited sampling theory for species and alleles. – *Ecol. Lett.* 8: 1147–1156.
- Feller, W. 1971. An introduction to probability theory and its applications, Vol. 2. – Wiley.
- Fisher, R. A. et al. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. – *J. Anim. Popul.* 12: 42–58.
- Gray, J. S. et al. 2005. The impact of rare species on natural assemblages. – *J. Anim. Ecol.* 74: 1131–1139.
- Green, J. L. and Plotkin, J. B. 2007. A statistical theory for sampling species abundances. – *Ecol. Lett.* 10: 1037–1045.
- Gregory, R. D. 1994. Species abundance patterns of British birds. – *Proc. R. Soc. B* 257: 299–301.
- Gregory, R. D. 2000. Abundance patterns of European breeding birds. – *Ecography* 23: 2001–2008.
- Harte, J. et al. 2009. Biodiversity scales from plots to biomes with a universal species–area curve. – *Ecol. Lett.* 12: 789–797.
- Hu, M.-K. 1962. Visual pattern recognition by moment invariants. – *IRE Trans. Inf. Theory* IT-8: 179–187.
- Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography. – Princeton Univ. Press.
- Hubbell, S. P. et al. 1999. Light gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. – *Science* 283: 554–557.
- Hubbell, S. P. et al. 2005. Barro Colorado Forest census plot data. – <<https://ctfs.arnarb.harvard.edu/webatlas/datasets/bci>>.
- Hubbell, S. P. et al. 2008. How many tree species are there in the Amazon and how many of them will go extinct? – *Proc. Natl Acad. Sci. USA* 105: 11498–11504.
- Kürka, P. et al. 2010. Analytical evidence for scale-invariance in the shape of species abundance distributions. – *Math. Biosci.* 223: 151–159.
- Lawton, J. H. 2000. Community ecology in a changing world. – International Ecology Inst.
- Magurran, A. E. and Henderson, P. A. 2003. Explaining the excess of rare species in natural species abundance distributions. – *Nature* 422: 714–716.
- McGill, B. J. 2003a. A test of the unified neutral theory of biodiversity. – *Nature* 422: 881–885.
- McGill, B. J. 2003b. Does Mother Nature really prefer rare species or are log-left-skewed SADs a sampling artefact? – *Ecol. Lett.* 6: 766–773.
- McGill, B. J. et al. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. – *Ecol. Lett.* 10: 995–1015.
- Mukundan, R. 2004. Some computational aspects of discrete orthonormal moments. – *IEEE Trans. Image Process.* 13: 1055–1059.
- Mukundan, R. et al. 2001. Image analysis by Tchebichef moments. – *IEEE Trans. Image Process.* 10: 1357–1364.
- Pitman, N. C. A. et al. 1999. Tree species distributions in an upper Amazonian forest. – *Ecology* 80: 2651–2661.
- Preston, F. W. 1948. The commonness, and rarity, of species. – *Ecology* 29: 254–283.
- Preston, F. W. 1962. The canonical distribution of commonness and rarity. – *Ecology* 43: 182–215.
- Pulliam, H. R. 1988. Sources, sinks, and population regulation. – *Am. Nat.* 132: 652–661.
- Rabinowitz, D. 1981. Seven forms of rarity. – In: Syngé, H. (ed.), *The biological aspects of rare plant conservation*. Wiley, pp. 205–217.
- Rosenzweig, M. L. 1995. *Species diversity in space and time*. – Cambridge Univ. Press.
- Šizling, A. L. et al. 2009a. Species abundance distribution results from a spatial analogy of central limit theorem. – *Proc. Natl Acad. Sci. USA* 106: 6691–6695.
- Šizling, A. L. et al. 2009b. Invariance in species-abundance distributions. – *Theor. Ecol.* 2: 89–103.
- Teague, M. R. 1980. Image analysis via the general theory of moments. – *J. Opt. Soc. Am.* 70: 920–930.
- Ugland, K. I. and Gray, J. S. 1982. Lognormal distributions and the concept of community equilibrium. – *Oikos* 39: 171–179.
- Volkov, I. et al. 2003. Neutral theory and relative species abundance in ecology. – *Nature* 424: 1037–1053.
- Williamson, M. and Gaston, K. J. 2005. The lognormal distribution is not an appropriate null hypothesis for the species-abundance distribution. – *J. Anim. Ecol.* 74: 409–422.
- Zillio, T. and He, F. 2010. Inferring species abundance distribution across spatial scales. – *Oikos* 119: 71–80.

Supplementary material (Appendix E7128 at <www.oikosoffice.lu.se/appendix>). Appendix 1–3.