

Invariance in species-abundance distributions

Arnošt L. Šizling · David Storch · Jiří Reif ·
Kevin J. Gaston

Received: 11 April 2008 / Accepted: 24 October 2008 / Published online: 4 December 2008
© Springer Science + Business Media B.V. 2008

Abstract Many attempts to explain the species-abundance distribution (SAD) assume that it has a universal functional form which applies to most assemblages. However, if such a form does exist, then it has to be invariant under changes in the area of the study plot (the addition of neighboring areas or subdivision of the original area) and changes in taxonomic composition (the addition of sister taxa or subdivision to subtaxa). We developed a theory for such an area-and-taxon invariant SAD and derived a formula for

such a distribution. Both the log-normal and our area-and-taxon invariant distribution fitted data well. However, the log-normal distributions of two adjoined sub-assemblages cannot be composed into a log-normal distribution for the resulting assemblage, and the SAD composed from two log-normal distributions fits the SAD for the assemblage poorly in comparison to the area-and-taxon invariant distribution. Observed abundance patterns therefore reveal area-and-taxon invariant properties absent in log-normal distributions, suggesting that multiplicative models generating log-normal-like SADs (including the power-fraction model) cannot be universally valid, as they necessarily apply only to particular scales and taxa.

Electronic supplementary material The online version of this article (doi:10.1007/s12080-008-0031-3) contains supplementary material, which is available to authorized users.

A. L. Šizling (✉) · K. J. Gaston
Biodiversity and Macroecology Group,
Department of Animal and Plant Sciences,
University of Sheffield,
Sheffield S10 2TN, UK
e-mail: a.sizling@sheffield.ac.uk
e-mail: arnost.l.sizling@seznam.cz

K. J. Gaston
e-mail: k.j.gaston@sheffield.ac.uk

A. L. Šizling · D. Storch
Center for Theoretical Study,
Charles University in Prague and the Academy of Sciences of the
Czech Republic,
Jilská 1,
110 00 Praha 1, Czech Republic

D. Storch
e-mail: storch@cts.cuni.cz

D. Storch · J. Reif
Department of Ecology, Faculty of Science, Charles University,
Viničná 7,
128 44 Praha 2, Czech Republic

J. Reif
e-mail: jirireif@yahoo.com

Keywords Scale invariance · Area invariance ·
Taxon invariance · Population sizes ·
Frequency distributions · Relative abundances

Introduction

The relative frequency of the abundances, a , of species in samples, assemblages and communities has long intrigued ecologists. Perhaps foremost this has been manifested in the search for a common descriptor, $f(a)$, of species-abundance distributions (SADs). Indeed, a wide range of models have been fitted to different data sets, including distributions derived variously from considerations of simple statistics, population dynamics, niche partitioning, or fractal geometry (e.g. Motomura 1932; Fisher et al. 1943; Preston 1948, 1981; Magurran 1988; Magnussen and Boyle 1995; Engen and Lande 1996; Sichel 1997; Tokeshi 1999; Dewdney 2000; Hubbell 2001; Green et al. 2003; Magurran and Henderson 2003; Harte et al. 2005; Pueyo et al. 2007; for reviews see May 1975; Marquet et al. 2003; McGill et al. 2007).

Few would disagree that some basic characteristics are common to virtually all real SADs. These include that they are markedly right-skewed (i.e. that the largest number of species occurs in one of the smallest abundance classes). Beyond this, however, there is surprisingly little consensus. Various authors champion the importance of different broad classes of models to describe SADs (see Marquet et al. 2003; McGill et al. 2007), of particular different distributions within those classes, and of different methods of assessing the fit of models. They also place different emphases on issues of sampling and thus on the relationship between the underlying probability distribution and observed SADs. Nonetheless, identifying an appropriate function characterizing observed SADs has generally been considered as crucial for understanding mechanisms generating relative species abundances.

One of the reasons for the lack of clarity as to which functions provide the best models of SADs is that tests have focused almost exclusively simply on fitting one or more such functions or model results to one or more observed SADs, whatever the approach used (but see Kempton and Taylor 1974). This is problematic because the outcomes of such tests are especially sensitive to the fit in the tails of the distributions and these tails are usually biased in observational data (Fisher et al. 1943; Preston 1948; Williamson and Gaston 2005; Yin et al. 2005). The tail formed by the rarest species is biased by the difficulties associated with finding such species. In consequence it has been shown that, for example, small samples from log-normal and log-normal-like distributions tend to obey log-series distributions (Golicher et al. 2006). The tail of an SAD formed by the commonest species can also be biased by sampling, although the effects are probably less systemic, and dependent on the particulars of the methodology used (e.g. whether effectively there are upper bounds above which abundances cannot be differentiated).

The influence of biases in the tails of SADs on the statistical fit of distributions is particularly important given that discussion as to the relative merits of different distributions has come often to focus on rather narrow differences in such fit (e.g. McGill 2003). Indeed, it has increasingly been recognized that other testable properties need to be explored to help identify those distributions that most properly describe SADs (Wilson 1993; McGill et al. 2007). One such property is the consistency of the description of the SAD in its spatial and taxonomic complexity: if the observed data are correctly described by some distribution function, then the SAD of an assemblage (e.g. defined spatially, taxonomically, ecologically, or in other terms) comprising two sub-assemblages, each of them described by the distribution function, should be well described by the statistical composition of these two distribution functions. Moreover, if the distribution function describing the SAD is universal, the composed distribution

should be characterized by the same distribution function as describes the two sub-assemblages, albeit with different parameters (Pueyo 2006). We call this property of SADs area-and-taxon invariance (Šizling and Storch 2007, for broader concept see Storch and Šizling 2008), since such a universal distribution is invariant under changes in area of the study plot and taxonomic composition (i.e. changes in the set of species included; note that taxon invariance refers here to species in different taxonomic groups, not to the abundances of those taxonomic groups, e.g. not a genera-abundance distribution). Area-and-taxon invariances therefore occur when the same functional form that fits species data over a small area or a lower level taxonomic group can fit data over a larger area (comprising multiple small areas) and at a higher taxonomic level (comprising species from several lower taxa).

So far, area-and-taxon invariances have been implicit in the assumption of a general form of the SAD. If the functional form of SADs is universal, then it should not change with enlarging or reducing the size of the study area. Similarly, it should not change if we broaden the set of species studied by adding a sister taxon, or if we use only a subtaxon (i.e. one particular order instead of the whole of a class). Consequently, all models trying to explain or derive SADs (e.g. broken stick, power fraction, or neutral community dynamics) should assume either area-and-taxon invariance or some basic level of description to which the model applies and from which the other levels (and other functional forms of the SAD) can be mathematically derived. For instance, if we assume that a particular model generating the SAD concerns species population abundances across whole biogeographic units (continents), and the model is not area-and-taxon invariant, then the functional forms for smaller areas are different from that produced by the model, and have to be secondarily derived. In other words, if we assume that there is a universal mechanism generating the SAD, then it must either lead to an area-and-taxon invariant distribution, or it must apply only at some fundamental level of description defined by a particular area and taxon level. Both cases have interesting implications. Searching for area-and-taxon invariance is thus not simply a mathematical nicety, as it concerns the question of whether there is a privileged spatial scale that determines ecological patterns at other scales or whether all scales are more or less equivalent.

Here we develop a theory concerning the invariance of SADs under variation in area and taxonomic composition. Furthermore, we introduce a new test of abundance distributions based on their behavior when changing area and/or taxonomic scope, and apply it to the log-normal SAD which has been widely championed as a null hypothesis for the form of SADs (Preston 1948; for possible mechanisms see Whittaker 1970; May 1975; Engen and Lande 1996; Engen 2001; May et al. 2007)

and to a new functional form which fulfils the condition of an area-and-taxon invariant distribution. We will show that although the log-normal distribution provides a good fit to the observed SADs at different spatial scales, the composition of two log-normal SADs neither fits merged data nor leads to another log-normal SAD, i.e. it does not reveal consistency nor exhibit area-and-taxon invariance. In contrast, our new distribution retains its good fit to the data regardless of changing area. Although we obviously cannot prove that our functional form of an area-and-taxon invariant SAD is the most appropriate description of observed SADs, unlike the log-normal SAD it is universal and consistent when changing area and/or taxonomic composition. We argue that no mechanism proposed as the explanation of the log-normal SAD can be universal across scales and taxa.

We also offer a tool for all of the computations presented here and thus for explorations of any available dataset (see <http://www.cts.cuni.cz/wiki/ecology:start>).

Theory

Requirement of spatial and taxonomic consistency of SADs

Observed SADs can be fitted equally well using many functional forms. However, a good fit is not the only indication of appropriateness of a given functional form. A stronger requirement for an appropriate functional form of the SAD is consistency of its fit with data when changing the observed area and/or taxonomic scope. Assume that we have two SADs, f_1 and f_2 , characterizing assemblages on some adjoining areas, described by the same functional form (e.g. if $\lambda \neq \sigma$ then $f(a) = \lambda e^{-\lambda a}$ and $f(a) = \sigma e^{-\sigma a}$ are two functions representing one functional form). We call these functional forms consistent if the empirical SAD (i.e. the observed pattern) for the whole assemblage of the two areas taken together is fitted well by the SAD whose functional form, including parameters, can be generated by the mathematical composition (hereafter composite distribution) of f_1 and f_2 (hereafter component distributions). If the composite distribution for the whole assemblage, on the other hand, fits the empirical SAD considerably worse than the functional forms for individual sub-assemblages, then f_1 and f_2 cannot represent appropriate functional forms reflecting an underlying mechanism even if their fit to the sub-assemblages is perfect.

Importantly, the SAD may not be spatially and taxonomically consistent even if the composite distribution can be fitted well by the same functional form, but independently of the composition of SADs of the sub-assemblages, i.e. with parameters tuned according to the empirical distribution. The reason is that many functional forms of

SADs are sufficiently flexible to fit well if properly—and independently—parameterized. However, the parameters of the composite SAD depend on the component SADs, so that the comparison between fits of the composite and component distributions helps identify the appropriate analytical form of the SAD much better than a simple independent fitting.

The consistency does not necessarily imply area-and-taxon invariance; even a non-invariant SAD could in principle be spatially and taxonomically consistent if the composite SAD (in such a case differing in its functional form from the component SADs) fits the data well. However, in such a case this functional form obviously cannot be universal, as it changes with scales and taxa, and we never know a priori which is the fundamental scale (or fundamental taxon level) at which the mechanism that produces a given SAD should apply. A truly universal functional form of the SAD must be area-and-taxon invariant, i.e. its functional form does not change after the composition from the distributions characterizing sub-assemblages, although it naturally changes its parameters.

Composition of two SADs of adjoining sub-assemblages

It is obviously not a problem to determine the empirical SAD (i.e. the cumulative distribution function constructed over data) of a given assemblage from the sets of abundances of its sub-assemblages—it is a simple matter of summing the abundances of each species across these sub-assemblages. A greater challenge is if we assume that the SAD for an assemblage has a particular functional form (log-normal, geometric series etc.) and we seek to determine what is the resulting functional form and whether the original functional form is retained when changing the area by addition of other assemblages, or, conversely, by dividing the area into smaller plots. Equally, one might pose the question in terms of the addition of other taxa or the splitting of a given taxon into subtaxa.

Let us start with merging two subtaxa (regardless of whether these represent two monophyletic sister clades or otherwise delimited, mutually exclusive sets of species). This situation is simpler than merging two neighboring areas, as subtaxa by definition do not share any species. The functional form of the SAD that results by composing two SADs for respective subtaxa is

$$f_{1 \cup 2}(a) = \pi_1 f_1(a) + \pi_2 f_2(a), \quad (1)$$

where a s are abundances, f s are probability density functions, and π_i is a proportion of species in the focal taxon i ($\pi_1 + \pi_2 = 1$).

The case of merging two adjoining areas is more complex, as this requires two operations. One is the same as in the previous case, i.e. addition of the SADs of species

which occur in only one sub-assembly. The second operation comprises convoluting (splicing together) the SADs of species which are common to both sub-assemblies. The result (i.e. the form of the resulting SAD) of the first operation depends on relative species turnover between the sub-assemblies, as the effect of this operation depends on the proportion of species which are unique to just one sub-assembly. The result of the second operation (below marked by $*_c$ for correlated-convolution, hereafter c-convolution) depends on the correlation between species abundances in the two sub-assemblies (Table 1). If the abundances are perfectly correlated (i.e. they are proportional to each other), the composition of the respective distributions comprises simple proportional enlargement of species abundances, whereas if they are only poorly correlated (or uncorrelated), the c-convolution is much more complex (see Fig. 1 and Appendix 1). More formally, the entire composition of the two distributions represents the proportional sum (i.e. the linear combination) of particular SADs

$$f_{\text{plot 1} \cup \text{plot 2}} = \pi_{10} f_{\text{sp in plot 1 exclusively}} + \pi_{01} f_{\text{sp in plot 2 exclusively}} + \pi_{11} f_{\text{common sp in plot 1}} *_c f_{\text{common sp in plot 2}} \quad (2)$$

where the π s are the proportional sizes of the respective groups and π_{11} is equivalent to Jaccard index (J) of assemblage similarity (number of species in common divided by total species number; Table 1).

Area-and-taxon invariant SADs

If there is no fundamental scale (*sensu* area) which determines abundance patterns at other scales, the SAD is necessarily area and/or taxon invariant. Now we explain why the area invariance and taxon invariance are not separable from each other and derive an area-and-taxon invariant SAD.

Taxon invariance refers to the situation in which the functional form of the SAD does not change when merging two subtaxa. A taxon invariant distribution function can be found when taking Eq. 1 as the functional equation (i.e. equation whose solution is a function) and finding its solution. Apparently, there are two solutions. (1) If both f_1 and f_2 are identical (including parameters), then the composed function is obviously again the function f_1 ($\equiv f_2$). The solution is therefore any function if its parameters do not vary between the assemblages. This case seems to be biologically irrelevant, since taxa commonly vary in their average abundances and thus they cannot share identical SADs. (2) If the functions differ in parameters, the solution is any function which can be expressed as a sum (linear combination) of several functional forms (i.e. $f(a) = \sum_{i=1}^N c_i f_i(a)$; where N is a free parameter which may vary between areas). The only condition imposed on the sum is finite integrability (i.e. the integral between zero and infinity must be finite and greater than zero) and positivity

Table 1 Definitions of measures used

<i>K</i> – <i>S</i> distance	Kolmogorov–Smirnov statistics, i.e. the maximum distance between two cumulative distribution functions; it ranges between 0 and 1; units: proportions
<i>Consistency</i>	Difference between mean <i>K</i> – <i>S</i> distance for particular distributions and the distance for those composed plus one $\left(\text{cons} \stackrel{\text{Def}}{=} (KS_{\text{plot1}} + KS_{\text{plot2}})/2 - KS_{\text{comp}} + 1 \right)$; <i>K</i> – <i>S</i> distance is the distance between data and assumed distribution; if the composite distribution fits worse than component distributions the <i>consistency</i> is below one; an absolutely consistent model has <i>consistency</i> equal to 1; if the composite distribution fits data better than the components, the <i>consistency</i> falls above 1; it ranges between 0 and 2; units: proportional values
<i>J</i>	Jaccard index of similarity, i.e. proportion of the species found in two plots that are common to both; units: [%]
Correlation	The correlation quantifies a situation in which the abundances of species in a given plot are related to the abundances of the same species in the other plot; in the model, the dependence between abundances is imposed by two constraints—i.e. two increasing lines (Fig. 1b) within which the abundances are independent of each other. Correlation is calculated as the complement of an angle between the constraints $\left(\text{correlation} \stackrel{\text{Def}}{=} \pi/2 - \text{angle} \right)$; the bigger it is, the more the abundances constrain each other; it ranges between 0 and $\pi/2$; units: [rad]
Area invariant functional form (i.e. model) of SAD	Iff an analytical formula describing a SAD does not change when composing two formulas characterizing two SADs of two plots, we say that this formula (i.e. the model of SAD) is area invariant. The parameters of the formula may vary after the composition
Taxon invariant functional form (i.e. model) of SAD	Iff an analytical formula describing SAD does not change when composing two formulas characterizing two SADs of two taxa, we say that this formula (i.e. model of SAD) is taxon invariant. The parameters of the formula may vary after the composition
Fundamental area or taxon	Iff SAD is not area and/or not taxon invariant, it implies an existence of an area or/and a taxon, whose SAD determine SADs of other areas and/or other taxa. Such area and taxon are called fundamental in the text

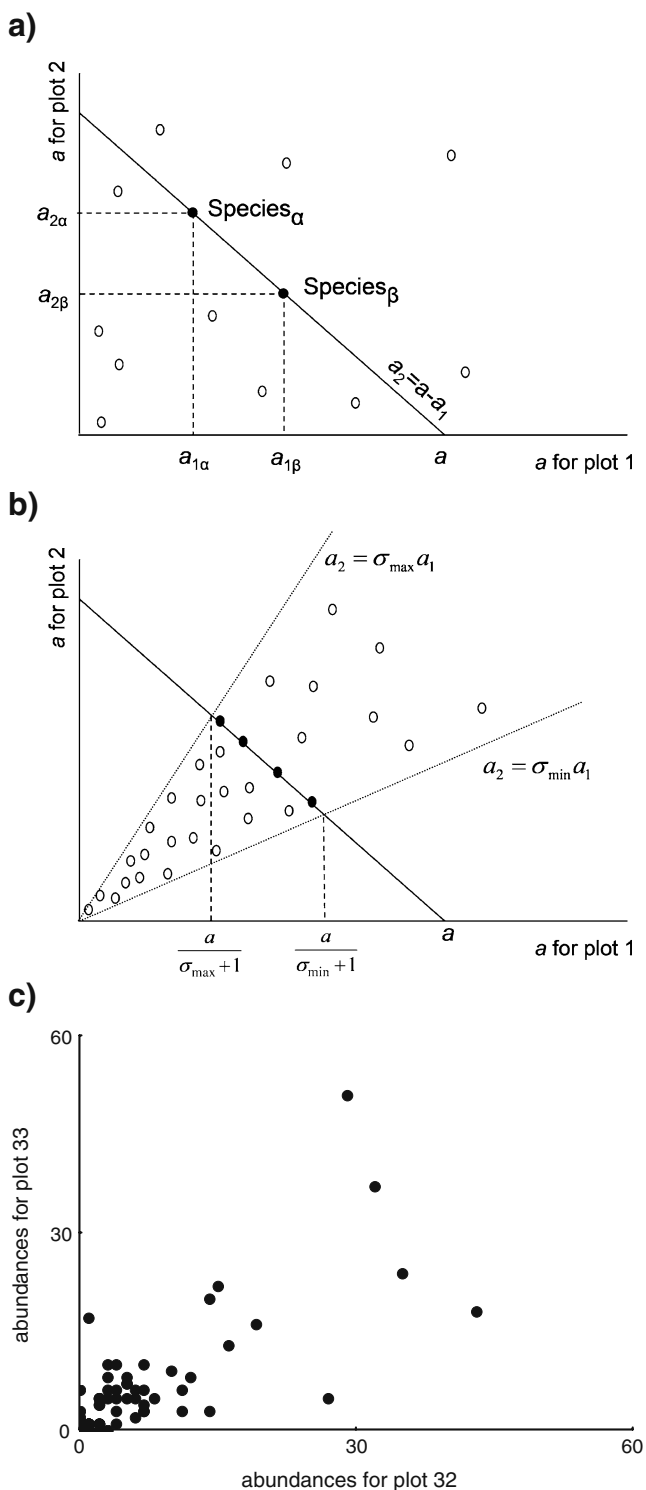


Fig. 1 Bivariate plots of abundances, a_1 and a_2 , for two disjunct plots in cases of **a** uncorrelated, **b** correlated and **c** observed abundances. Both *full* and *open circles* represent individual species. The species that occupy only one plot were excluded. **a** All the combinations of abundances giving the abundance a , if plots are joined, follow a line (*full line, full circles*); **b** if abundances are correlated, the points are constrained by two increasing lines (*dotted*); abundances for plot 1, a_1 , are thus bounded by $\frac{a}{\sigma_{\max}+1}$ and $\frac{a}{\sigma_{\min}+1}$ where σ s are slopes of the lines; **c** abundances observed on the transect were mostly weakly correlated with each other and followed the assumed pattern. The abundances of the species that occupy only one of the plots are shown in this graph

subtaxa comprises two sets of distinct species, whereas composing SADs from two adjoining areas comprises both merging two sets of distinct species and c-convolution of SADs of common species, area invariance implies also taxon invariance, provided that some species occur only in one sub-assembly. (Note that *taxon* here necessarily refers to any set of species, not only to a monophyletic species clade, i.e. this invariance comprises all possible taxonomic systems.) In such a case we can speak about area-and-taxon invariant SADs, and hereafter we will deal with this universal type of invariance. This invariance would be obvious if all species were shared by both areas and if abundances were fully correlated, so that the convolution would comprise simple rescaling of SADs for the respective assemblages (see above). However, the truly (i.e. non-trivially) area-and-taxon invariant SAD is characterized by the robustness against changes in species turnover (i.e. against the variation in π s) and changes in the level of correlation of abundances between the sub-assemblages. The natural condition imposed on the SAD by the area invariance is that every SAD must attain the origin (i.e. zero abundance has zero probability density; McGill et al. 2007). The reason is that nonzero probability density at zero abundance would imply that we know how many species are missing from a given area — but this is apparently arbitrary and relative to the definition of the whole species pool.

An area-and-taxon invariant function as defined above is a proportional sum of difponential distributions introduced by Preston (1981) (the name is based on ‘the difference between two exponentials’). We will therefore call it the multi-difponential distribution (see also the multi-exponential distribution in Šizling and Storch 2007), and it is expressed as

$$f(a) = \sum_{i=1}^N c_i (e^{-A_i a} - e^{-\alpha_i a}), \tag{3}$$

where a is abundance, the parameters A_i and α_i are positive ($0 < A_i < \alpha_i$), and c_i is a real number (see Fig. 2 for meaning of the parameters). For the rationale and the proof that it obeys the requirement of area-and-taxon invariance

($f(a) \geq 0$ for all a); for a similar idea in economics see Mandelbrot (1963).

Area invariance characterizes SADs whose functional form does not change when joining two areas, i.e. when it is transferred from small to large areas by the composition of the distributions of smaller areas. Since merging two

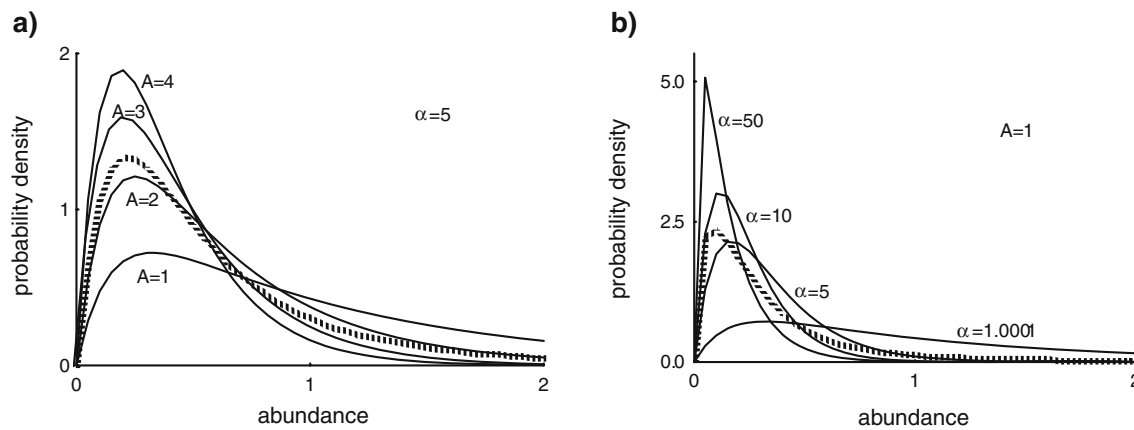


Fig. 2 Shapes of additive terms of the multi-diffonential distribution for different combinations of A and α parameters. Each term has a shape similar to a log-normal or log-series (high α in b) distribution, depending on proportion of A and α parameters (Eq. 3). The terms are shown after normalization (i.e. multiplying by a normalization constant $(= \alpha A / (\alpha - A))$ to have the area below the curve equal to one, so that the parameter c (Eq. 3) is replaced by $\delta(\alpha A / (\alpha - A))$ where δ is a real number which can be called *dominance* of the focal

term, because its value determines how close is the entire distribution to the focal term, which can then be called the dominant term. Sum of all dominances across all terms is one. *Full lines* in subplot **a** show variation in A ($=1, 2, 3,$ and 4), holding $\alpha=5$, while they show variation in α ($=1.0001, 5, 10,$ and 50), holding $A=1$ in subplot **b**. Note that $A < \alpha$ by definition. *Dashed lines* show the entire distribution (proportional sum of the terms) in the case when all the terms are equally dominant and thus equally contribute to the entire distribution

(Eq. 2; where model of c-convolution, $*_c$, is defined by Eq. 5 in Appendix I) see Appendix II.

The multi-diffonential distribution has a potentially large number of terms and consequently parameters, which could be interpreted as a disadvantage in comparison to simpler functional forms. However, this is actually a necessary property of any area invariant distribution, since species spatial turnover between sub-assemblages implies that the functional form is taxon invariant and thus actually *is* composed from a certain number of other functions, i.e. component SADs (see Eq. 1). Only a few additive terms might, however, be dominant (Fig. 2).

Methods

Data

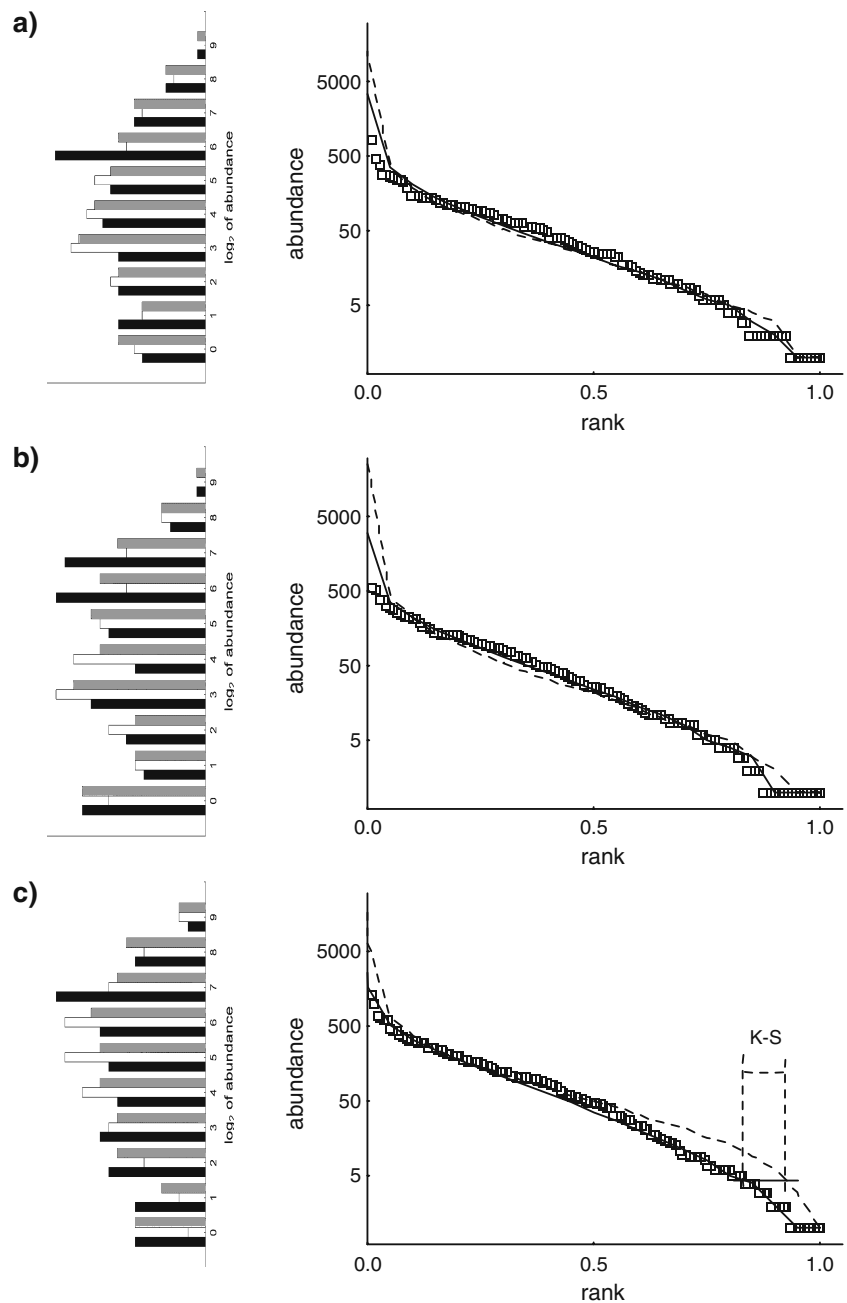
To test the area-and-taxon invariance of SADs we needed detailed abundance data of adjoined species assemblages at several spatial scales. Such data are not easily available, so we performed a multiscale bird survey across the Czech Republic. The data comprise the abundances of 144 bird species (excluding *Larus ridibundus* and *Riparia riparia*, i.e. the species breeding in large colonies) at each of 768 points along a linear East-West transect in south Bohemia and Moravia. Birds were mapped by the point count method (Bibby et al. 1992) within the distance limited by 150 m around each point during five early morning visits in

the breeding seasons (April–June) of 2004 and 2005. Points were separated by between 300 and 500 m, such that study plots of diameters 150 m at each point were approximately adjacent. The maximum recorded numbers of individuals of each species at each point from the five visits were used for analyses to obtain reliable abundance estimates (see Storch et al. 2002).

Tests

We performed two kinds of tests. Both focused on area-related variation of the SAD, because taxon-invariance inevitably follows from area-invariance, as explained above. First, we simply evaluated the fit of the multi-diffonential and the log-normal distributions to the observed SADs for different areas (in fact lengths of transects, see below). We will call this the test of the fit between the modeled and observed shape of the SAD. It consisted of calculating the Kolmogorov–Smirnov (K–S) distance (maximum difference between cumulative distribution functions; Table 1, Fig. 3c, and S1) between the observed SAD and the fitted distributions for abundance data for each sample of 50 consecutive points, of 100 consecutive points, and so on up to 750 points. All possible sets of a given number of consecutive points were used. The log-normal and multi-diffonential distributions vary in number of parameters, thus we also compare them using an information criterion (AIC; Akaike 1974), in the case of multi-diffonential distributions varying numbers of terms between 1–20 (log-normal distribution has two parameters).

Fig. 3 SAD as observed (*squares* in rank plots; *black bars* in Preston-style histogram), and modelled by log-normal (*dashed line* in rank plots; *white bars* in Preston-style histogram) and multi-diffonential (*full line* in rank plots; *dashed bars* in Preston's plot) approaches for **a** samples of the first 300 points (1–300) and **b** next 300 points (301–600) of the transect. Sub-plot **c** shows the merged data and composed distributions. K–S is the Kolmogorov–Smirnov distance, which is maximum distance in the direction of the axis of proportional rank in this case; it can be calculated in this way because the proportional rank plot is an inverted cumulative distribution function—the proportional rank estimates the cumulative probability. The rank plots for the log-normal and multi-diffonential approaches were constructed using 15,000 simulations



The weakness of the test of fit is that fit depends on the number of free parameters of a given distribution, and on the fitting procedure—and both these properties can naturally differ among different functional forms of SADs. Moreover, the fit of a distribution can be affected by sampling issues, as the functional form of the sampled distribution can differ from the underlying distribution (Green and Plotkin 2007). For this reason we focused on the second type of test, which evaluates consistency of the functional form of SADs across scales (areas) as explained above; hereafter the test of consistency. Since the correla-

tion of abundances and species turnover obviously affects consistency of the SAD, as we have shown, we evaluated the extent to which the consistency across scales (areas) is dependent on the correlation between species abundances in two given sub-assemblages, and on species turnover between them.

The test of consistency thus evaluated the deterioration of the fit of SADs which were mathematically composed from component SADs of two sub-assemblages (see Fig. 3), and the sensitivity of this deterioration on the variation in species turnover between sub-assemblages

and the correlation between abundances in each sub-assembly. This evaluation was performed as follows (see also SI):

1. The SADs for each species assemblage were fitted using both log-normal and multi-diffonential distributions. The species assemblages comprised non-overlapping sections of the point transect, starting with 20 points, then 40, 60,... 180 neighboring points.
2. Component SADs were constructed using the fitted SADs of each pair of composite assemblages, i.e. SADs of two neighboring sections on the transect. The composite assemblages thus consisted of 40, 80, 120,... 360 sample points.
3. Kolmogorov–Smirnov distances were calculated for the comparison between empirical and fitted (component) SADs for both sub-assemblages. We call these distances KS_{plot1} and KS_{plot2} .
4. Kolmogorov–Smirnov distance was calculated for the comparison between the composite SAD (composed from SADs of the given pair of sub-assemblages) and the empirical SAD for the respective assemblage. We call this distance KS_{comp} .
5. Consistency was calculated as a contrast between distances KS_{plot} and KS_{comp} , namely as $(KS_{plot1} + KS_{plot2}) / 2 - KS_{comp} + 1$. Consistency is thus a measure of the “success” of the composition of the two distributions in terms of the fit of the composite distribution to the merged data: a value of 1 indicates the same fit of the composite distribution as the fit of the component distributions, lower values indicate poorer fit of the composite distribution, whereas values higher than 1 indicate a better fit of the composite distribution in comparison to the component distributions.

Both the K–S distance and composed distributions were calculated numerically in all cases. The K–S distance was computed as a distance between the empirical cumulative distribution function for data and 15,000 abundances randomly drawn from the fitted or composed distribution. The log-normal distribution was fitted by calculating mean and variance from logarithmically transformed data, whereas for the fitting of the multi-diffonential distribution a specific algorithm was developed (see SII). Drawing from the log-normal distribution was based on the standard algorithm, as $\text{Exp}(\text{sum of } 1,000 \text{ random values drawn from the uniform distribution } (0,1) \text{ rescaled to } N(\mu, \sigma))$, and for drawing from the multi-diffonential distribution the cumulative distribution function was used (Eq. 7 in Appendix II, and SI). The number of 15,000 abundances was chosen according to the Dvoretzky–Kiefer–Wolfowitz (DKW) inequality (Wasserman 2004; wikipedia .../wiki/Dvoretzky_Kiefer-Wolfowitz_inequality) not to deviate more than about 1% from the

underlying distribution with confidence of 95%. Although the multi-diffonential distribution reveals good fit even in the case of two additive terms (SIII), we used 10 terms for all simulations.

Numerical simulations

Distributions of pairs of adjacent plots were composed numerically. This allowed us to use the same method for both the distributions and thus to make the results comparable with each other. We preferred a method based on random samples, each from one component distribution. The composition of fitted distributions comprised three steps: (1) generating abundances for species common to both plots, (2) generating abundances for species occupying only the first plot, and (3) generating abundances for species occupying only the second plot (see SI). All numbers of generated abundances were proportional to the numbers in the respective groups (π_{11} , π_{10} , and π_{01}), and together they comprise 15,000 values. The abundance for species common to both plots was generated as the sum of the two abundances drawn from the respective fitted distributions; however, the two abundances were selected so that the empirical observed correlation of species abundances between the respective plots was retained (the pairs of abundances of respective species from adjacent subplots must lie between two increasing constraints; see Fig. 1b, Eq. 5 in Appendix I, and SI.f, and o–q). This approach to the c-convolution is only one of the possible definitions and should properly be called a constrained-convolution. We preferred it for its simplicity. The constraints were extracted from the data for each composition separately.

Results

Both the distributions can be accepted at the $p=0.05$ level and rejected at $p=0.01$ level for the test of the *fit* using a test for K–S statistics (DKW inequality; $p \cong 0.02$), although the multi-diffonential distribution apparently fit the data better (Fig. 4). There were also no dramatic changes in the fit between different sized areas (i.e. plots differing in number of census points). However, the mean K–S for a given area ($N=14$) was correlated with the size of the area (although with marginal significance for a log-normal distribution, $p \cong 0.052$, $r=-0.52$, whereas for a multi-diffonential distribution $p < 10^{-5}$, $r=-0.98$), both the distributions fitting better in larger areas (in agreement with Preston 1948; Green and Plotkin 2007). The slightly but systematically better fit of the multi-diffonential distribution (Fig. 4) could in principle be caused by the fact that this distribution has more free parameters (Fig. 2). However, the

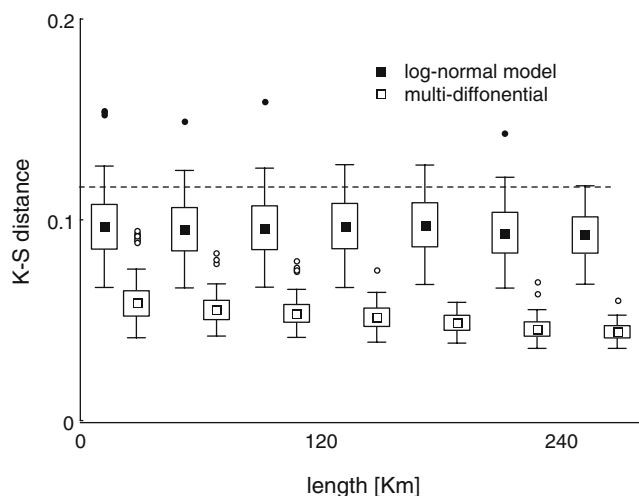


Fig. 4 Box (50%) and whiskers (95%) of the K–S distances (data-fitted distributions) for plots of various areas measured as the length of transect in km. Medians of distances for log-normal (black squares) and multi-diffonential (white squares) distributions are shown as well as the outliers (full and empty circles, respectively). The dotted line shows the limit ($p < 0.05$) for rejecting the fit (DKW test for K–S distance)

AIC identified all multi-diffonential distributions with 1–20 terms as providing a better fit than the log-normal (for details see Supplement: SIV).

Our main aim, however, was not to test the shape of the distribution, but its properties, as only the properties might indicate possible underlying mechanisms. Since many underlying mechanisms can produce very similar shapes, we tested both the models for consistency. The test of consistency revealed remarkable inconsistency for the log-normal distribution (Fig. 5) and consequently for all of its underlying processes. The consistency of the multi-diffonential distribution was much higher and could not be rejected. This was apparent when comparing the numbers of cases which fell below and above the level of absolute consistency (consistency=1). While all of the 104 cases fell

below one (i.e. reveal inconsistency) for the log-normal distribution, 74 of 104 cases fell above one (i.e. reveal consistency) for the multi-diffonential distribution (for details including significances see Table 2).

The inconsistency of the log-normal distribution and overall consistency of the multi-diffonential distribution were also revealed by the dependence of consistency on the

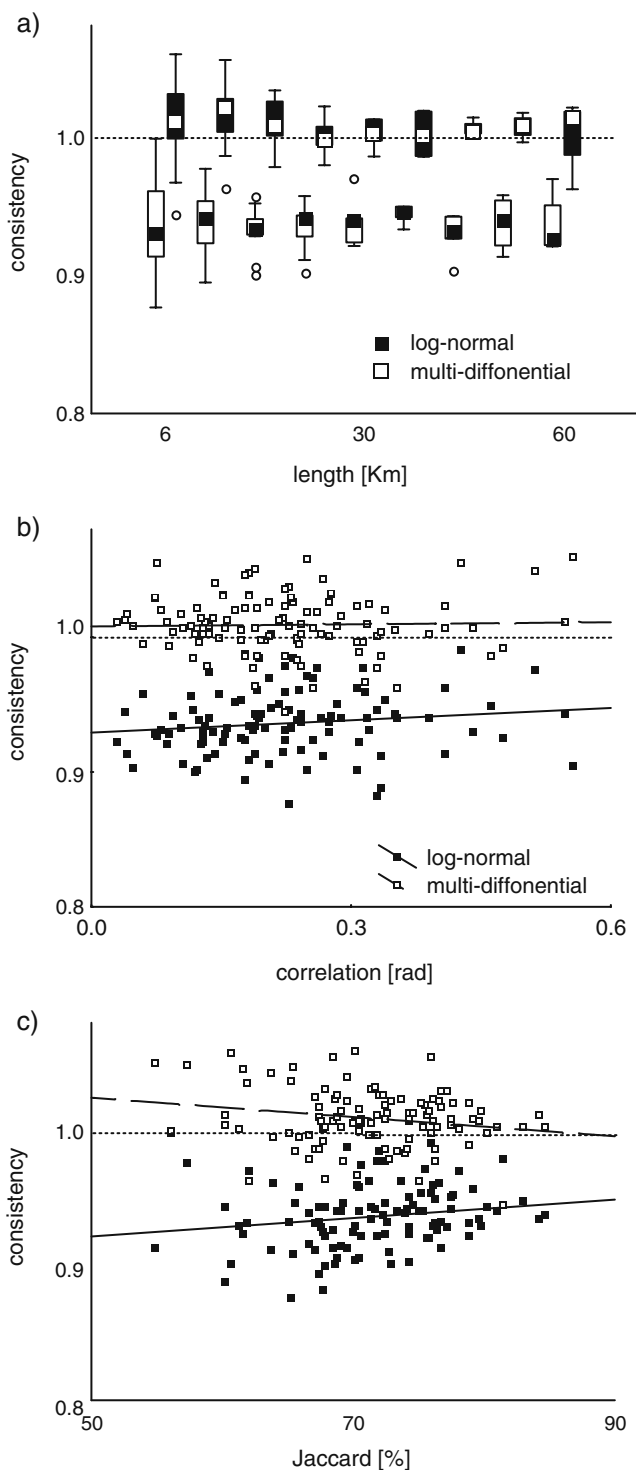


Fig. 5 Test of consistency for log-normal (black squares) and multi-diffonential (white squares) distributions. Consistency refers to the difference between K–S distances of component and composed distributions. Consistency of one (dotted lines) refers to the full consistency between composed and fitted distributions when adding assemblages of new plots. **a** Consistency against area measured as the length of the transect in km. Boxes and whiskers show minimum and maximum estimates of 95% CIs for expected and observed values, respectively (t -distribution); circles show the outliers. **b** The dependence of consistency on the correlation of species abundances between two adjacent plots. The full increasing line shows the regression for the log-normal distribution, the dashed constant line shows the regression for the multi-diffonential distribution. Note that the point of absolute correlation is at $\pi/2 \cong 1.57$. **c** The dependence of consistency on Jaccard index (J). The full increasing line shows the regression for the log-normal distribution, the dashed decreasing line shows the regression for the multi-diffonential distribution

Table 2 Tests of consistency of log-normal and multi-diffonential SADs

Area [<i>n</i> of points]	Log-normal			Multi-diffonential		
	Below one	Above one	<i>p</i> value	Below one	Above one	<i>p</i> value
20	38	0	3.6–10 ⁻¹²	11	27	0.99
40	19	0	1.2–10 ⁻⁶	3	16	0.99
60	12	0	2.5–10 ⁻⁴	3	9	0.98
80	9	0	0.002	5	4	0.5
100	7	0	0.008	2	5	0.94
120	6	0	0.015	3	3	0.66
140	5	0	0.03	1	4	0.97
160	4	0	0.06	1	3	0.94
180	4	0	0.06	1	3	0.94
Overall	104	0	5–10 ⁻³²	30	74	0.999

Numbers of pairs of adjacent plots that reveal inconsistency (columns ‘below one’; consistency <1) and high consistency (columns ‘above one’; consistency >1), for the log-normal and multi-diffonential distributions. The significance *p* is the probability that less or equal numbers fall above one by chance assuming equal probability to be in both groups, calculated as $p = \sum_{i=0}^n C_i^N 0.5^i$ (*N* is the total number of points and *n* is number of points that fell above one; *C* is the combination number)

correlation between abundances in neighboring plots and on species turnover (based on the knowledge that all possible distributions reveal consistency in the case of strongly correlated abundances and low species turnover; see the **Theory** section). The log-normal distribution revealed low consistency regardless of the level of correlation between abundances ($p \approx 0.15$, $r \approx 0.14$, $N = 104$), indicating that the range of measured correlations did not approach the level at which it would reveal consistency due to the similarity of component SADs. However, the consistency of log-normal SADs increased with similarity of assemblage composition as measured by the Jaccard index ($p \approx 0.08$, $r \approx 0.17$, $N = 104$) (Fig. 5b,c). The consistency of the multi-diffonential distribution was invariably high and independent of the correlation of abundances ($p \approx 0.79$, $r \approx 0.03$, $N = 104$), although it slightly decreased with assemblage similarity ($p \approx 0.05$, $r \approx -0.19$, $N = 104$). The plot actually showed better agreement with data after composition (i.e. consistency higher than 1) than before when species turnover between assemblages was high (and when the effect of rare species should rather lead to inconsistency). Conversely to the multi-diffonential distribution, the log-normal distribution not only failed in producing the same distributional form after the composition, but the distribution composed from two log-normal distributions which fitted sub-assemblage SADs itself fitted poorly to the merged data at all tested scales.

Discussion

We have shown that the functional form of the multi-diffonential distribution is much more consistent with observed abundance data than is the log-normal, implying that

the observed SAD is likely to be area-and-taxon invariant. Moreover, the log-normal model of SAD was shown to be inconsistent when enlarging area, as the composition of two log-normal SADs does not fit the SAD of the joined sub-assemblages for any of the sizes of area tested. If SADs were log-normal for some area, then we should see (in Fig. 5a) high consistency when joining plots of that area regardless of any lack of area invariance (i.e. regardless of whether the composite distribution is also log-normal). Our finding thus highlights the failure of the log-normal distribution as a consistent descriptor of species abundance patterns, although the log-normal SAD may still serve as a useful approximation for a particular assemblage for some practical purposes.

This has important implications for the mechanisms possibly producing observed SADs. The inconsistency means that no underlying process producing log-normal SADs (for reviews see May 1975; Sugihara 1980; Engen 2001; May et al. 2007) can be considered as universal and satisfactory (see also Williamson and Gaston 2005). Whereas log-normal-like distributions emerge due to multiplicative processes (May 1975), the processes generating overall SADs are naturally at least partially additive. This is because every SAD can be considered as composed from SADs of sub-assemblages occupying only part of the whole area, which share several species. This also means that the models based on niche division (the broken stick and power-fraction models, see Tokeshi 1999), are equally inappropriate as all of these models are implicitly based on purely multiplicative processes (Williamson and Gaston 2005). It would still be possible that multiplicative processes could determine SADs at the largest scales (i.e. continental or global), whereas SADs at local scales would emerge by sampling the regional SAD, but even in this case the log-normal SAD would not be an appropriate charac-

terization of observed local SADs due to its spatial and taxonomical inconsistency.

Some authors have highlighted the possible scale (sensu area) dependency of the SAD. For instance, Loehle and Hansen (2005) showed that whereas local SADs for North American birds were best represented by a log-series model, regional SADs revealed more complex shapes corresponding neither to a log-normal nor a log-series, but which could be reconstructed using a complex sampling model. Likewise, neutral models of metacommunity dynamics (Hubbell 2001, see also Chave 2004) predict the zero-sum multinomial SAD for the whole metacommunity, which is modified by the effect of dispersal limitation for local communities (Volkov et al. 2003; He 2005), so that local SADs are characterized by a higher prevalence of abundant species (see also Borda-de-Água et al. 2007). Such empirical findings and theoretical considerations concur with our main point: the SAD need not necessarily be area and/or taxon invariant, but if it is not invariant, then its functional form must be translatable from one scale to another in a statistically consistent way, and one particular scale has to be considered as fundamental, the SADs on different scales being derived from this fundamental level.

It seems that the commonly used log-normal distribution in general fits data reasonably well because both log-normal and multi-diffonential distributions can follow similar shapes. For our data the log-normal distribution fits quite well over a relatively wide range of areas (length, 20–270 km which corresponds to areas between 5–80 km²; Fig. 4). However, such fitting to larger areas is independent from those for smaller areas and thus is not constrained by parameters of component distributions. This explains why the dependence of the log-normal distribution on area has not yet been recognized although the distribution has been used for various sizes of areas (e.g. Preston 1948; Etienne and Olff 2004; Walla et al. 2004; Connolly et al. 2005; Ulrich and Ollik 2005; Yin et al. 2005; Syrek et al. 2006).

Indeed, Preston (1981) suggested the diffonential distribution (i.e. a single additive term of our area-and-taxon invariant distribution) as an appropriate model of the SAD. He noticed the similarity of this distribution to the log-normal distribution and also described its slight left-skewness in logarithmic space, which agrees well with many empirical SADs (e.g. Nee et al. 1991; Gregory 2000). The only problem discussed by Preston (1981) was that the diffonential distribution had small variance in comparison with data. The area-and-taxon invariant distribution (Eq. 3) is composed from several “diffonentials”, which increases its variance and thus solves the problem. A vital consequence of the complexity of the distribution is that it can potentially be multimodal, in agreement with some data (Gray et al. 2005) and in striking contrast to all other models of the SAD (McGill et al. 2007).

Importantly, it is not only the log-normal distribution which may approximately follow an area and taxon invariant shape. When increasing αs (Eq. 3), the multi-diffonential distribution approaches the sum of functions (ce^{-Aa}), each representing a continuous version of the discrete log-series models (bq^a/a) introduced by Fisher et al. (1943) (after logarithmic transformation each of the distribution functions closely approach a line), and which has been found to be a good descriptor of SADs for successional stands (Visser 1995; Death 1996; Kammesheid 1998; Salvador-Van-Eysendore et al. 2003).

Baker and Preston (1946) presented the diffonential distribution as a result of two counteracting effects exponentially decreasing in their frequency (for a particular mechanism in the case of species-abundance distribution see Preston 1981). However, this would require some fundamental area and/or taxon for which the effects apply. In other cases (i.e. areas larger than the area of the fundamental level), a SAD must be composed from diffonentials for many areas and taxa (Eq. 2), which results in a sum of more diffonentials (see also the reasoning in Mandelbrot 1963). Since the number of component areas and taxa is always finite, we also assume a finite number of diffonentials to be summed.

Besides the mechanism suggested by Preston (1981), a purely formal reason might play a role. A large variety of functions can be approached using a sum of exponential distributions, $\sum \delta_i \lambda_i e^{-\lambda_i a}$ ($\sum \delta_i = 1$). A composition of two such forms over areas and taxa (Eq. 2) then produces diffonential terms which are likely to become dominant over exponential terms if the similarity of species composition (measured by Jaccard index) is high (for details see Appendix III). Because each SAD necessarily originates as the composition of SADs of many local subassemblages, it is likely that the multi-diffonential distribution prevails over other forms. This process constraints the resulting distributions (e.g. $f(0) \rightarrow 0$, which was indeed an assumption imposed on the SAD; McGill et al. 2007). This mechanism depends on the particular model of c-convolution, so it must be taken only as a suggestion for further research (which is the case also of Preston's suggestion).

Our theory sheds fresh light on the evidence of decreasing log-skewness with area (McGill et al. 2007). Numerical simulation suggests that this effect might be attributed to increasing community similarity (Jaccard index). If each subsequent composition is parameterized by higher Jaccard index than the previous one, the log-skewness decreases. Since our data show that the Jaccard index between two adjacent subplots of the same size indeed increases with area, the effect might be a consequence of that increase.

In conclusion, we have shown that although the log-normal distribution may often be a good approximation to

observed SADs, it does not follow data when applying probability rules to study its changes with study area or taxonomic composition. It thus does not represent a consistent statistical description of the SAD which would have a potential to reveal underlying mechanisms. An appropriate description of the SAD would then be properly represented either by some area-and-taxon invariant distribution, such as our multi-diffonential distribution, or by a multitude of distributions whose functional forms would change when changing ‘scale’ or taxonomic delimitation. The latter approach would require, however, a specification of the fundamental spatial ‘scale’ and taxonomic level at which the respective SAD-generating mechanism would work, and from which the functional forms of other distributions would be derived. To find or dismiss such a fundamental scale is a great challenge for ecology.

Acknowledgement We thank J. Mourková and J. Cepák for their help with the fieldwork. We are grateful to M. Williamson, J. Nekola, B. J. McGill, F. He and anonymous referees for their comments. A.L. Š. was supported by the Marie Curie Fellowship no. 039576-RTBP-EIF. K.J.G. holds a Royal Society-Wolfson Research Merit Award. This work was supported by grants from the Czech Ministry of Education No. LC06073 and MSM0021620845, and Grant agency of the AS CR (IAA601970801).

Appendix I: Composition of two SADs with correlated abundances

If abundances of the assemblages to be composed are correlated with each other, the SADs of the two assemblages interact in a particular way, which must be reflected in the definition of the mathematical operation that models a composition of the SADs. This operation was marked as $*_c$ in Eq. 2, which refers to convolution (*) accounting for a correlation (index c).

If there is no correlation between abundances, the SADs $f_1(a)$ and $f_2(a)$ are convoluted as

$$f_1(a) * f_2(a) \approx \int_0^a f_1(a_1)f_2(a - a_1)da_1. \tag{4}$$

This formula is based on the fact that the probability

density of each abundance, a , is given by the sum (integral) of densities for all combinations of abundances, $\{a_1;a_2\}$, along a line of possible combinations of abundances, $a_2=a - a_1$ (Fig. 1a). The reason is that (1) each point along the line represents one independent combination of abundances which gives the resulting abundance, a , and (2) the product of $f_1(a_1)$ and $f_2(a - a_1)$ is the probability (density) that the abundances a_1 and $a_2 (=a - a_1)$, which give the abundance a , occur simultaneously (i.e. each abundance applies to one assemblage). If abundances are not correlated, the line of the combinations is bounded by abundances of zero and a ($0 < a_1 < a$ and $0 < a_2 < a$; full line in Fig. 1a).

The simplest way to model correlation of abundances is by constraining the lines representing the possible combination of abundances which produce resulting abundance a by two increasing lines ($a_2 = \sigma_{\min}a_1$, $a_2 = \sigma_{\max}a_1$) intersecting the origin (Fig. 1b,c). If abundances are perfectly correlated, the abundances of the plots are proportional to each other — i.e. both lines approach each other.

In this case, we integrate probability densities (simply said, we sum the probabilities) not along the whole line between 0 and a (Fig. 1a), but only between two extreme abundances ($a/(\sigma_{\max} + 1)$ and $a/(\sigma_{\min} + 1)$), which are imposed by the correlation (Fig. 1b). The composition of two SADs for species common to both subassemblages then obeys

$$f_1(a) *_c f_2(a) = \nu \int_{\frac{a}{\sigma_{\max} + 1}}^{\frac{a}{\sigma_{\min} + 1}} f_1(a_1)f_2(a - a_1)da_1, \tag{5}$$

where ν is a normalization constant.

Appendix II: A solution of the Eq. 2

Having the analytical form of the composition of two functional forms of SADs (Eq. 2 where $*_c$ is defined by Eq. 5), we can check whether a multi-diffonential distribution (Eq. 3) is area-and-taxon invariant, i.e. whether the form composed from two multi-diffonentials is a multi-diffonential distribution. The evidence has been done by substituting the solution to Eq. 5. which gives

$$f_1 *_c f_2 \approx \sum_{i \neq j} \left(\frac{c_i c_j}{A_i - A_j} \left(e^{\frac{A_i + A_j \sigma_{\min}}{\sigma_{\min} + 1} a} - e^{\frac{A_i + A_j \sigma_{\max}}{\sigma_{\max} + 1} a} \right) + \frac{c_i c_j}{\alpha_i - \alpha_j} \left(e^{\frac{\alpha_i + \alpha_j \sigma_{\min}}{\sigma_{\min} + 1} a} - e^{\frac{\alpha_i + \alpha_j \sigma_{\max}}{\sigma_{\max} + 1} a} \right) + \frac{c_i c_j}{A_j - \alpha_i} \left(e^{\frac{\alpha_i + A_j \sigma_{\min}}{\sigma_{\min} + 1} a} - e^{-\frac{\alpha_i + A_j \sigma_{\max}}{\sigma_{\max} + 1} a} \right) + \frac{c_i c_j}{\alpha_j - A_i} \left(e^{\frac{A_j + \alpha_j \sigma_{\min}}{\sigma_{\min} + 1} a} - e^{\frac{A_j + \alpha_j \sigma_{\max}}{\sigma_{\max} + 1} a} \right) \right), \tag{6}$$

which is, after substitution and reindexation, the multi-diffonential form (Eq. 3) as well (with different parameter N). The multi-diffonential distribution is obviously robust against proportional summation (linear combination), and thus follows the entire composition (Eq. 2) and is both taxon and area invariant. Note that in the cases in which the differences between respective A s and/or α s approach zero the additive terms, whose denominators (see Eq. 6) are affected by these small values, turn into gamma distributions ($\gamma_i a^{v_i} e^{-\beta_i a}$; $v_i \in \mathbb{N}$), and the area-and-taxon invariant distribution becomes more complicated. However, there is only a very low probability that this happens by chance, and thus here we present only the simpler solution. The cumulative distribution function of the multi-diffonential distribution obeys

$$P(x < a) = \sum_{i=1}^N c_i \left(\frac{e^{-\alpha_i a} - 1}{\alpha_i} - \frac{e^{-A_i a} - 1}{A_i} \right), \tag{7}$$

and the expectation is

$$E(a) = \sum_{i=1}^N c_i (A_i^{-2} - \alpha_i^{-2}). \tag{8}$$

For the range of shapes see Fig. 2, for fitting procedure see Šizling and Storch (2007) or supplement SIII, and for fitting utility <http://www.cts.cuni.cz/wiki/ecology:start>.

Appendix III: Possible formal mechanism producing multi-diffonential SADs

First we show that a variety of distributions can be modelled as a sum of exponential distributions. According to the Taylor theorem almost any function f can be approached by $f(y) = \sum_{i=0}^N c_i y^i$. Denote abundance a as $\ln(1/y)$ ($y \in (0; 1)$). Then a variety of shapes can be expressed as $f(e^{-a}) = \sum_{i=0}^N c_i e^{-ia}$. Because the result of the transformation is assumed to be a distribution (i.e. with finite integral between zero and infinity), we exclude all functions with $i=0$. As a result, a variety of distributions can be approached by a sum of exponential forms

$$\phi(a) = \sum_{j=1}^N c_j e^{-\lambda_j a}, \tag{9}$$

where $\lambda_{\min} \leq \lambda_j (0 < \lambda_{\min} \leq 1)$ for all j .

Now we show that the composition of two sums of exponential forms produces diffonential terms, which imposes additional constraints on a resulting distribution (e.g. an existence of indices m, n so that $c_m = -c_n$, which cannot happen by chance). Assume two distributions

expressed as $\phi_{10}(a) = \sum c_j e^{-\lambda_j a}$ and $\phi_{01}(a) = \sum c_l e^{-\lambda_l a}$. Their composition follows

$$\begin{aligned} \phi_{10} \text{ composed with } \phi_{01} &= \pi_{10} \sum c_j e^{-\lambda_j a} \\ &+ \pi_{01} \sum c_l e^{-\lambda_l a} \\ &+ \pi_{11} \sum \frac{c_j c_l}{\lambda_l - \lambda_j} (e^{-\lambda_j a} - e^{-\lambda_l a}). \end{aligned} \tag{10}$$

(Eq. 2; c-convolution is defined as in Appendix I). If either of the component distributions comprises diffonential terms, the evidence does not change, because composition of two diffonentials remains diffonential (Appendix II). A gamma distribution results if both parameters λ happen to be equal to each other (Appendix II), which we consider unlikely.

If we express each additive term as a normalized distribution (terms in brackets ‘[]’ in Eq. 11) multiplied with its dominance δ , we get

$$\begin{aligned} \phi_{10} \text{ cmp with } \phi_{01} &= \pi_{10} \sum \delta_j [\lambda_j e^{-\lambda_j a}] \\ &+ \pi_{01} \sum \delta_l [\lambda_l e^{-\lambda_l a}] \\ &+ \pi_{11} \sum \delta_j \delta_l \left[\frac{\lambda_j \lambda_l}{\lambda_l - \lambda_j} (e^{-\lambda_j a} - e^{-\lambda_l a}) \right], \end{aligned} \tag{11}$$

where $\delta = c/\lambda$. This suggests that a high Jaccard index (the proportion of species common to both plots π_{11} ; note that $\pi_{11} = 1 - \pi_{10} - \pi_{01}$ and all $\pi \geq 0$) can make the diffonential terms dominant when composing SADs of several subplots.

References

Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19:716–723. doi:10.1109/TAC.1974.1100705

Baker TC, Preston FW (1946) Fatigue of glass under statistic loads. *J Appl Phys* 17:170–178. doi:10.1063/1.1707702

Bibby CJ, Burgess ND, Hill DA (1992) *Bird census techniques*. Academic, London

Borda-de-Água L, Hubbell SP, He F (2007) Scaling biodiversity under neutrality. In: Storch D, Marquet PA, Brown JH (eds) *Scaling biodiversity*. Cambridge University Press, Cambridge, pp 347–375

Chave J (2004) Neutral theory and community ecology. *Ecol Lett* 7:241–253. doi:10.1111/j.1461-0248.2003.00566.x

Connolly SR, Hughes TP, Belwood DR et al (2005) Community structure of corals and reef fishes at multiple scales. *Science* 309:1363–1365. doi:10.1126/science.1113281

Death RG (1996) The effect of habitat stability on benthic invertebrate communities: the utility of species abundance distributions. *Hydrobiologia* 317:97–107. doi:10.1007/BF00018733

- Dewdney AK (2000) A dynamical model of communities and a new species-abundance distribution. *Biol Bull* 198:152–165. doi:10.2307/1542811
- Engen S (2001) A dynamic and spatial model with migration generating the log-Gaussian field of population densities. *Math Biosci* 173:85–102. doi:10.1016/S0025-5564(01)00077-3
- Engen S, Lande R (1996) Population dynamic models generating the lognormal species abundance distribution. *Math Biosci* 132:169–183. doi:10.1016/0025-5564(95)00054-2
- Etienne RS, Olff H (2004) A novel genealogical approach to neutral biodiversity theory. *Ecol Lett* 7:170–175. doi:10.1111/j.1461-0248.2004.00572.x
- Fisher RA, Corbet AS, Williams CB (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. *J Anim Ecol* 12:42–58. doi:10.2307/1411
- Golicher DJ, O'Hara RB, Ruiz-Montoya L et al (2006) Lifting a veil on diversity: a bayesian approach to fitting relative abundance models. *Ecol Appl* 16:202–212. doi:10.1890/04-1599
- Gray JS, Björnsaeter A, Ugland KI (2005) The impact of rare species on natural assemblages. *J Anim Ecol* 74:1131–1139. doi:10.1111/j.1365-2656.2005.01011.x
- Green JL, Plotkin JB (2007) A statistical theory for sampling species abundances. *Ecol Lett* 10:1037–1045. doi:10.1111/j.1461-0248.2007.01101.x
- Green J, Harte J, Ostling A (2003) Species richness, endemism and abundance patterns: test of two fractal models in a serpentine grassland. *Ecol Lett* 6:919–928. doi:10.1046/j.1461-0248.2003.00519.x
- Gregory RD (2000) Abundance patterns of European breeding birds. *Ecography* 23:201–208. doi:10.1111/j.1600-0587.2000.tb00276.x
- Harte J, Conlisk E, Ostling A et al (2005) A theory of spatial structure in ecological communities at multiple spatial scales. *Ecol Monogr* 75:179–197. doi:10.1890/04-1388
- He F (2005) Deriving a neutral model of species abundance from fundamental mechanisms of population dynamics. *Funct Ecol* 19:187–193. doi:10.1111/j.0269-8463.2005.00944.x
- Hubbell SP (2001) The unified theory of biodiversity and biogeography. Princeton University Press, Princeton
- Kammesheidt L (1998) The role of tree sprouts in the restorations of stand structure and species diversity in tropical moist forest after slash-and-burn agriculture in Eastern Paraguay. *Plant Ecol* 139:155–165. doi:10.1023/A:1009763402998
- Kempton RA, Taylor LR (1974) Log-series and log-normal parameters as diversity discriminants for the lepidoptera. *J Anim Ecol* 43:381–399. doi:10.2307/3371
- Loehle C, Hansen A (2005) Community structure and scaling relations for the avifauna of the US pacific and inland northwest. *Ecol Complex* 2:59–70. doi:10.1016/j.ecocom.2004.09.001
- Magnussen S, Boyle TJB (1995) Estimating sample-size for inference about the Shannon-Weaver and the Simpson index of species-diversity. *For Ecol Manage* 78:71–84. doi:10.1016/0378-1127(95)03596-1
- Magurran AE (1988) Ecological diversity and its measurement. Croom Helm Australia, New South Wales
- Magurran AE, Henderson PA (2003) Explaining the excess of rare species in natural species abundance distributions. *Nature* 422:714–716. doi:10.1038/nature01547
- Mandelbrot BB (1963) New methods in statistical economics. *J Polit Econ* 71:421–440. doi:10.1086/258792
- Marquet PA, Keymer JE, Cofré H (2003) Breaking the stick in space: of niche models, metacommunities and patterns in the relative abundance of species. In: Blackburn TM, Gaston KJ (eds) *Macroecology: Concepts and consequences*. British Ecological Society and Blackwell Science, Oxford, pp 64–81
- May R (1975) Patterns of species abundance and diversity. In: Cody ML, Diamond JM (eds) *Ecology and evolution of communities*. The Belknap Press of Harvard University Press, Cambridge, pp 81–120
- May RM, Crawley JM, Sugihara G (2007) Communities: Patterns. In: May RM, McLean A (eds) *Theoretical ecology*. Oxford University Press, Oxford, pp 111–131
- McGill BJ (2003) Strong and weak tests of macroecological theory. *Oikos* 102:679–685. doi:10.1034/j.1600-0706.2003.12617.x
- McGill BJ, Etienne RS, Gray JS et al (2007) Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol Lett* 10:995–1015. doi:10.1111/j.1461-0248.2007.01094.x
- Motomura I (1932) A statistical treatment of associations. *Zool Mag Tokyo* 44:379–383 in Japanese
- Nee S, Harvey PH, May RM (1991) Lifting the veil on abundance patterns. *Proc R Soc Lond B Biol Sci* 243:161–163. doi:10.1098/rspb.1991.0026
- Preston FW (1948) The commonness, and rarity, of species. *Ecology* 29:254–283. doi:10.2307/1930989
- Preston FW (1981) Pseudo-lognormal distributions. *Ecology* 62:355–364. doi:10.2307/1936710
- Pueyo S (2006) Diversity: between neutrality and structure. *Oikos* 112:392–405. doi:10.1111/j.0030-1299.2006.14188.x
- Pueyo S, He F, Zillio T (2007) The maximum entropy formalism and the idiosyncratic theory of biodiversity. *Ecol Lett* 10:1017–1028. doi:10.1111/j.1461-0248.2007.01096.x
- Salvador-Van-Eysendore D, Bogaert J, Zak-Mnacek V et al (2003) Sapling diversity in canopy gaps in an Equadorian rain forest. *For Sci* 49:909–917
- Sichel HS (1997) Modelling species-abundance frequencies and species-individual functions with the generalized inverse Gaussian-Poisson distribution. *South African Statist J* 31:13–37
- Šizling AL, Storch D (2007) Geometry of species distributions: Random clustering and scale invariance. In: Storch D, Marquet PA, Brown JH (eds) *Scaling biodiversity*. Cambridge University Press, Cambridge, pp 77–100
- Storch D, Šizling AL (2008) The concept of taxon invariance in ecology: Do diversity patterns vary with changes in taxonomic resolution? *Folia Geobot*. doi:10.1007/s12224-008-9015-8
- Storch D, Gaston KJ, Cepák J (2002) Pink landscapes: 1/f spectra of spatial environmental variability and bird community composition. *Proc R Soc Lond B Biol Sci* 269:1791–1796. doi:10.1098/rspb.2002.2076
- Sugihara G (1980) Minimal community structure: an explanation of species abundance patterns. *Am Nat* 116:770–787. doi:10.1086/283669
- Syrek D, Weiner WM, Wojtylak M et al (2006) Species abundance distribution of collembolan communities in forest soils polluted with heavy metals. *Appl Soil Ecol* 31:239–250. doi:10.1016/j.apsoil.2005.05.002
- Tokeshi M (1999) Species coexistence: Ecological and Evolutionary perspectives. Blackwell, Oxford
- Ulrich W, Ollik M (2005) Limits to the estimation of species richness: the use of relative abundance distributions. *Divers Distrib* 11:265–273. doi:10.1111/j.1366-9516.2005.00127.x
- Visser S (1995) Ectomycorrhizal fungal succession in jack pine stands following wildfire. *New Phytol* 129:389–401. doi:10.1111/j.1469-8137.1995.tb04309.x

- Volkov I, Banavar JR, Hubbell SP et al (2003) Neutral theory and relative species abundance in ecology. *Nature* 424:1035–1037. doi:[10.1038/nature01883](https://doi.org/10.1038/nature01883)
- Walla TR, Engen S, DeVries PJ et al (2004) Modeling vertical beta-diversity in tropical butterfly communities. *Oikos* 107:610–618. doi:[10.1111/j.0030-1299.2004.13371.x](https://doi.org/10.1111/j.0030-1299.2004.13371.x)
- Wasserman LA (2004) *All of statistics: A concise course in statistical inference*. Springer, Berlin
- Whittaker RH (1970) *Communities and Ecosystems*. Macmillan, New York
- Wilson JB (1993) Would we recognise a broken-stick community if we found one? *Oikos* 67:181–183. doi:[10.2307/3545108](https://doi.org/10.2307/3545108)
- Williamson M, Gaston KJ (2005) The lognormal distribution is not an appropriate null hypothesis for the species-abundance distribution. *J Anim Ecol* 74:409–422. doi:[10.1111/j.1365-2656.2005.00936.x](https://doi.org/10.1111/j.1365-2656.2005.00936.x)
- Yin ZY, Peng SL, Ren H et al (2005) LogCauchy, log-sech and lognormal distributions of species abundances in forest communities. *Ecol Modell* 184:329–340. doi:[10.1016/j.ecolmodel.2004.10.011](https://doi.org/10.1016/j.ecolmodel.2004.10.011)