

RESEARCH
PAPER



The role of niche overlap, environmental heterogeneity, landscape roughness and productivity in shaping species abundance distributions along the Amazon–Andes gradient

Gabriel Arellano^{1*}, María N. Umaña², Manuel J. Macía³, M. Isabel Loza^{4,5}, Alfredo Fuentes⁶, Victoria Cala⁷ and Peter M. Jørgensen⁸

¹Center for Tropical Forest Science – ForestGEO, Smithsonian Tropical Research Institute, Washington, DC, USA,

²Department of Biology, University of Maryland, College Park, MD 20742, USA,

³Departamento de Biología, Área de Botánica, Universidad Autónoma de Madrid, Calle Darwin 2, Madrid ES–28049, Spain,

⁴Department of Biology, University of Missouri, St Louis, MO 63121, USA,

⁵Herbario Nacional de Bolivia, Campus Universitario Cota-Cota Calle 27, Correo Central Cajon Postal 10077, La Paz, Bolivia,

⁶Herbario Nacional de Bolivia, Universidad Mayor de San Andrés, Casilla 10077 Correo-Central, La Paz, Bolivia,

⁷Departamento de Geología y Geoquímica, Facultad de Ciencias, Universidad Autónoma de Madrid, Madrid, ES-28049, Spain,

⁸Missouri Botanical Garden, St Louis, MO 63166-0299, USA

ABSTRACT

Aim Statistical and ecological mechanisms shape species abundance distributions (SADs). A lack of correlation between ecological gradients and SAD shape would suggest that SADs are caused by purely statistical reasons. We evaluated the variation in the shape of SADs for communities in landscapes of differing variable connectivity, environmental heterogeneity, species niches overlap and productivity.

Location Rainforests in the Madidi region (Bolivia).

Methods We compiled biological and environmental information on 65 sites (a site being a group of two to six 0.1-ha plots where woody plants of a diameter at breast height ≥ 2.5 cm were inventoried). We built unveiled (complete) SADs for each site and fitted Gambin models to those SADs. The Gambin α parameter served as a metric of SAD shape. Low α values characterize logseries-like SADs, while high α values characterize lognormal-like SADs. For each site, we estimated landscape roughness, environmental heterogeneity, species niche overlap and productivity. These variables were related to SAD shape by means of variation partitioning.

Results SADs changed from logseries-like to lognormal-like along the elevational gradient. Many of our predictor variables were correlated: 40.4% of the variation in SAD shape could not be attributed to specific factors. However, 50.62% of the variation in the SAD shape could be assigned to individual predictor matrices: 28.4% was explained exclusively by niche overlap, 15.41% exclusively by environmental heterogeneity, 5.20% exclusively by landscape roughness and 1.6% exclusively by productivity.

Main conclusions Ecological processes related to the topographical/environmental complexities that vary across the elevational gradient are correlated with the SAD shape. Purely statistical mechanisms are apparently not sufficient to explain the changes in SAD shape. The most important factor is the mean overlap of the niches of the species of an assemblage: avoiding competition with co-occurring species could be the most important mechanism driving species relative success at the ≤ 100 km² scale.

Keywords

Dispersal limitation, environmental heterogeneity, Madidi, niche partitioning, productivity, SAD, tropical rain forest.

*Correspondence: Gabriel Arellano, Center for Tropical Forest Science – ForestGEO, Smithsonian Tropical Research Institute, NMNH-MRC 166, West Loading Dock, 10th and Constitution Ave. NW, Washington, DC, USA.
E-mail: gabriel.arellano.torres@gmail.com

INTRODUCTION

In its essential form, the species abundance distribution (SAD) is a vector of 'comparable abundances for the species in a community' (McGill, 2011), and thus one of the most fundamental descriptors of any community. Many alternative theoretical models have been proposed, and many fit empirical SADs equally well, making it difficult or impossible to conclude from the shape of the SAD alone anything about underlying statistical or ecological mechanisms (McGill *et al.*, 2007; McGill, 2011). Because of the practical limitation of the curve fitting approach in particular, empirical evaluations of SAD hypotheses lag behind important theoretical developments (McGill *et al.*, 2007; McGill, 2011).

Although it is almost a universal law that most species are rare and few are common, communities can differ in the shape of SADs, from those where a few species are very abundant and many are very scarce, to those where abundances are more evenly distributed across species. Here, we will refer to those SADs where most species are very scarce as *logseries* SADs (Fisher *et al.*, 1943) and to those where abundances are more even as *lognormal* SADs (Preston, 1948). We use these as qualitative and purely descriptive terms which are widely recognized, without necessarily claiming that these two distributions exactly fit the data or even are the best possible fit. Both shapes (logseries and lognormal) constitute the extremes of a continuum of potential SAD shapes. Although previously unnoticed, changes in the SAD shape itself could be a more tractable phenomenon than changes in species abundances. Instead of asking why some species are more abundant than others, this work asks why some SADs are more lognormal or logseries than others.

Mechanisms shaping SADs can be statistical or ecological. Examples of statistical mechanisms are constraints imposed by the number of species and individuals (Locey & White, 2013), limited sampling from a pool of individuals (McGill, 2003; Green & Plotkin, 2007) or the central limit theorem (Conlisk *et al.*, 2012). Statistical mechanisms themselves could have an ecological basis: for example, SADs might be shaped by the number of species and individuals alone but these aspects, in turn, could be determined by elevation, nutrients or light availability. Therefore, empirical tests on the relationship between SADs and environmental or community properties are still required (McGill *et al.*, 2007; Yen *et al.*, 2013). A lack of correlation between the shape of a SAD and environmental covariates would indicate that SADs do not have an ecological basis and could be just a result of statistical mechanisms (Yen *et al.*, 2013).

There are several biological and ecological explanations for the changes in the shape of SADs (reviewed by McGill *et al.*, 2007). The reasoning underlying most of these alternative explanations is similar: communities where more species find more and better opportunities would present more lognormal SADs, whereas systems where only one or a few species find good opportunities would present logseries SADs, with most of the species occurring at low abundances. Both

neutral and niche-based views explain the SAD patterns by evoking this general mechanism. From a neutral perspective (Hubbell, 2001) space is considered the only resource for which species compete, and 'opportunity' means 'a place in the space in which to live'. When many species struggle to find available space in a local community, there are few opportunities for each species, and most species tend to occur at low abundances (reflected by logseries SADs). When only a few species occur at each point, these could have relatively more opportunities to be abundant, leading to more lognormal SADs at larger scales (Hubbell, 2001). Such a battle for space is determined by the rate of immigration to a given locality, a direct function of the quantity and effectiveness of barriers to dispersal in the surrounding landscape. Consequently, systems where dispersal is difficult would present more lognormal SADs and systems where dispersal is easy would present more logseries SADs (Hubbell, 2001; Carson & Schnitzer, 2009).

A second group of hypotheses develop a similar reasoning, with the niche concept as a cornerstone. Habitats that offer more niches to more species are expected to have more species that attain relatively large numbers of individuals, and thus more lognormal SADs. Therefore, more lognormal SADs are expected in more complex or heterogeneous environments and/or with less competitive overlap among species (Cotgreave & Harvey, 1994; Ugland *et al.*, 2007), as well as in more productive environments (Hurlbert, 2004; Chase, 2010; Brown, 2014).

The present work evaluates SAD shapes in the Madidi region (Bolivia), one of the longest elevational gradients with continuous forest cover in the world. This gradient from the Amazon to the Andean highlands implies more disconnected landscapes at higher elevations, where migration of individuals and dispersal of propagules is increasingly difficult (Young, 1995; Gradstein *et al.*, 2008; Graham & Fine, 2008; Parra-Olea *et al.*, 2012). Besides, we expect more complex and heterogeneous abiotic conditions in the montane areas than in the lowlands (Gentry, 1995; Vitousek, 1998; Gerold, 2008). Regarding the changes in the average overlap of species niches, montane floras have been hypothesized to be largely composed of climatic generalists colonizing from the lowlands (Stevens, 1992; Ghalambor *et al.*, 2006). However, both isolated localities and more complex habitats should favour the evolution of specialist species, even from a pool of generalist taxa. There are also important changes in productivity associated with temperature changes in the gradient, from luxuriant Amazonian forests at 300-m elevation to dwarf ridge forests above 3000-m elevation (Vitousek, 1998; Gerold, 2008; Gradstein *et al.*, 2008).

The objective of the present work is to test for causes underlying the variation in SAD shape at the ≤ 100 km² scale by characterizing the SADs of communities embedded in landscapes of differing properties. To do this, we performed variation partitioning to disentangle the effects of: (1) landscape roughness, (2) environmental heterogeneity, (3) niche overlap, and (4) productivity, on the shape of the SAD curve.

If SADs change from logseries to lognormal as topographical complexity increases, then ease of dispersal would determine the relative abundance of species (Hubbell, 2001; Carson & Schnitzer, 2009). If SADs change from logseries to lognormal as environmental heterogeneity increases, or in communities where species show less niche overlap, then niche-based processes determine the SAD shape (Cotgreave & Harvey, 1994; Ugland *et al.*, 2007). Finally, if SADs change from lognormal to logseries as productivity increases, then greater productivity would be correlated with more opportunities for more species (Hurlbert, 2004; Brown, 2014). Note that some of these hypotheses make opposite predictions about the direction of the logseries to lognormal change along the elevational gradient, given that the gradient in productivity is opposite to the gradient of habitat and topographical complexity.

METHODS

Study region and baseline floristic data

As part of the Madidi Project (<http://www.mobot.org/madidi>) we carried out extensive standardized floristic inventories in the Madidi region, located on the eastern slopes of the Bolivian Andes (latitude 12°25'48" S to 15°45'00" S; longitude 69°28'48" W to 66°39'36" W). For this study, we used data from rain forests between 250 and 3600 m in elevation. Dry forests and *Polypepis* stands were excluded from the analyses, as well as a few plots without environmental information (see below for the environmental variables considered). The rain forest dataset contains 94,322 woody plant individuals with a diameter at breast height (d.b.h., at 130 cm from the rooting point) equal to or greater than 2.5 cm and rooting within the limits of 308 0.1-ha plots (typically 20 m × 50 m, exceptionally 10 m × 100 m). This dataset contains 2233 species and morphospecies, each of which was vouchered at each visited locality, giving a total of 14,533 vouchers. All vouchers are deposited at the Herbario Nacional de Bolivia (LPB) and many duplicates at the Missouri Botanical Garden (MO). Our database is available for consultation at Tropicos® (<http://tropicos.org/PlotSearch.aspx?projectid=20>), which functions as a permanent data repository maintained by the Missouri Botanical Garden. Data availability and the terms of use are described in the data sharing and publication policy of the Madidi Project.

Measurement of the SAD shape of a given assemblage

SADs can be obtained from any set of samples surveying a given biological assemblage. In the present paper, we considered groups of two to six closely spaced plots ('sites'). The grouping of plots into sites took into account the goodness-of-fit of the SAD model employed. Hence, prior to explain the grouping of plots into sites, this section describes the process of fitting a model to any given SAD and estimating its shape.

We followed Chao *et al.* (2015), who characterized the relationship between the true SAD in the entire assemblage and the observed SAD in a sample using Good-Turing frequency estimation. This statistical technique estimates the probability of finding an object of an unseen class, given a set of past observations of objects from different classes. The unveiling process returns the relative abundances of all the species estimated to be present in the assemblage, enabling us to focus on the mechanisms driving the SAD shape and leave sampling issues aside. Once we had estimated the species abundances at each site using Chao *et al.*'s (2015) approach, we modelled the shape of unveiled SADs by fitting Gambin models, a purely statistical model demonstrated to be highly flexible to different SAD shapes (Ugland *et al.*, 2007; Matthews *et al.*, 2014). This model has a single parameter (α) that describes the shape of the SAD: low α values are characteristic of logseries SADs whereas high α values reflect lognormal SADs. Thus, the Gambin α parameter is a good diagnostic (and 'mechanism-agnostic') tool for any process that could affect the shape of the SAD curve (Ugland *et al.*, 2007; Matthews *et al.*, 2014). The accuracy of the estimation of each α parameter was assessed with bootstrapping (999 times).

The unveiling process was implemented using code from the 'JADE.R' appendix of Chao *et al.* (2015). To fit the models we employed customized functions in R that allowed us to fit models to relative abundances, borrowing code from the *gambin* R package (Ugland *et al.*, 2007; Matthews *et al.*, 2014). Appendix S1 in the Supporting Information contains methodological and technical details, all the R code needed and a worked example.

Definition of sites and final floristic dataset

We defined 'sites' as groups of plots fulfilling the following conditions: (1) they contain a minimum of 1000 individuals (McGill *et al.*, 2007; McGill, 2011); (2) they fit into a circle of radius ≤ 10 km, to limit the range of spatial scales evaluated; (3) they fit within an elevational band of ≤ 309 m, which is the average elevational change needed for a full floristic turnover in the rain forests of our study region, as measured by the first axis of a detrended correspondence analysis; and (4) the predictions from a Gambin model fitted to the SAD shape of a site must be statistically indistinguishable from the true SAD, according to a χ^2 test. The fulfilment of this last condition ensured that all our estimates of the SAD shapes of sites came from accurately fitted models.

The resulting 65 sites contained from two to six plots (mean = 3.88 plots), 1000 to 1838 individuals (mean = 1214), and 58 to 505 observed species (mean = 240). During this process 56 plots could not be assigned to any site and were excluded from the analyses. The final floristic dataset contained 252 plots with 79,950 individuals belonging to 2110 species. Only a few individuals (1343 = 1.67%) were excluded from the analyses because they could not be assigned to any species or morphospecies. The location of

the sites, the R code that conducted the assignment of plots to sites and a summary of the characteristics of the 65 sites (including Tropicos® plot name codes to allow the examination of our raw data) are presented in Appendix S2.

For the assemblage at each site, we calculated the Gambin α parameter characterizing the shape of its unveiled SAD. This was the response variable in subsequent analyses. The following sections describe the measurement of all the processes that we hypothesized to affect the SAD shape. In all cases, these predictors were measured at the site level ($n = 65$).

Estimation of the ease of dispersal

We estimated ease of dispersal by measuring topographical complexity as follows.

1. We obtained elevation data for the Madidi region from the ASTER Global Digital Elevation Map (DEM), version 2 (<http://asterweb.jpl.nasa.gov/gdem.asp>). This digital elevation map has a 1-sec resolution (~ 30 m), and is the most accurate topographic reference for the study area.
2. We extracted elevation data of concentric circles of radius r km centred at each plot, r in $\{0.1, 0.5, 1, 2, 5, 10\}$. This was done to assess the influence of the surrounding landscape at different scales.
3. We calculated the topographical complexity at each of these circles with the average surface roughness index

$$R_a = \frac{1}{n} \sum_{i=1}^n |h_i|$$

where n is the number of pixels touching each circle and $|h_i|$ is the absolute difference in elevation between the i th pixel in the circle and the mean value of all pixels. R_a is expressed in the same units as elevation, and is interpreted as the average deviation of each pixel from a horizontal plane at the mean elevation of the evaluated area ($R_a = 0$ in a perfectly plane horizontal surface, and no upper bound in theory).

4. For each r -value, the landscape roughness of a site was estimated as the simple mean of the R_a indices of the circles of that size surrounding each of the plots contained in the site. Therefore, the landscape roughness matrix (\mathbf{R}) contained 65 rows and 6 columns. All topographic data were handled with the *raster* R package (Hijmans, 2014).

Note that landscape roughness can also be interpreted as an environmental factor, and is probably correlated with environmental heterogeneity. In this study, we interpret landscape roughness as a purely dispersal-limiting agent only, once the effect of environmental heterogeneity is removed (see below).

Environmental variables considered and estimation of environmental heterogeneity

We measured superficial soil (0–30 cm below the litter layer) properties at each plot, after air-drying samples and sieving soils through a 2-mm sieve. The percentages of sand, silt and clay were measured using the hydrometer method; soil pH

was measured in a 1:2.5 soil:H₂O suspension; organic carbon (C) was determined with the Walkley and Black method; and total nitrogen (N) with the semi-micro Kjeldahl method (Reeuwijk, 2002). We also measured calcium (Ca), magnesium (Mg) and potassium (K) using two extractive methods for different sets of samples: (1) the 1 M ammonium acetate solution method, and (2) the Mellich-3 extraction method (Mehlich, 1984). The results from both methods are linearly correlated in our study region, so we standardized all soil measurements into comparable values (see Appendix 2 of Arellano *et al.*, 2016, for the data and calculations involved). We also calculated the C:N ratio as an additional edaphic variable. We estimated climatic conditions at each plot using the 19 bioclimatic variables in the WorldClim database (rasters of 30-arcsec resolution (~ 1 km); Hijmans *et al.*, 2005). These 19 climatic variables (plus elevation) were summarized into three principal components (PCs) of a principal components analysis (PCA). These three PCs accounted for over 95% of the variance of the climatic/elevational data. PC1 can be interpreted as temperature, PC2 as temperature seasonality, and PC3 as temperature seasonality mixed with precipitation (Fig. 1). Finally, we estimated slope, northness and eastness from the DEM and the geographical coordinates of the plots. This makes a total of 16 edaphic, climatic and topographic variables, which were centred and scaled to unit variance for further analyses.

We estimated the environmental heterogeneity of each site as: (1) the standard deviations along each of the 16 environmental variables for the plots contained in a site; (2) the mean environmental Euclidean distance between the plots contained in a site; (3) the maximum environmental Euclidean distance between the plots contained in a site; and (4) the mean environmental Euclidean distance along the minimum spanning tree connecting all the plots contained in a site. Therefore, the environmental heterogeneity matrix (\mathbf{H}) contained 65 rows and $16 + 3 = 19$ columns.

Estimation of overlap of species niches

For each pair of species present in our dataset we computed a metric of niche overlap along each of the 16 environmental variables. This implied several steps involving kernel density estimations. First we applied Silverman's rule of thumb (as in the *bw.nrd0* function of the R package *stats*) to estimate the bandwidth for a Gaussian kernel density estimator for each species i along the environmental variable k . We then averaged across all species to obtain a single bandwidth to be applied to all the species along k . This guaranteed that the underlying kernels were all the same for the density estimation of all the species in that environmental variable. Then we calculated the Gaussian kernel density estimator for each species i along k : D_{ik} (using the *density* function of the *stats* R package). Finally, we calculated the niche overlap along k for each pair of species (i, j) as the Bhattacharyya distance between D_{ik} and D_{jk} :

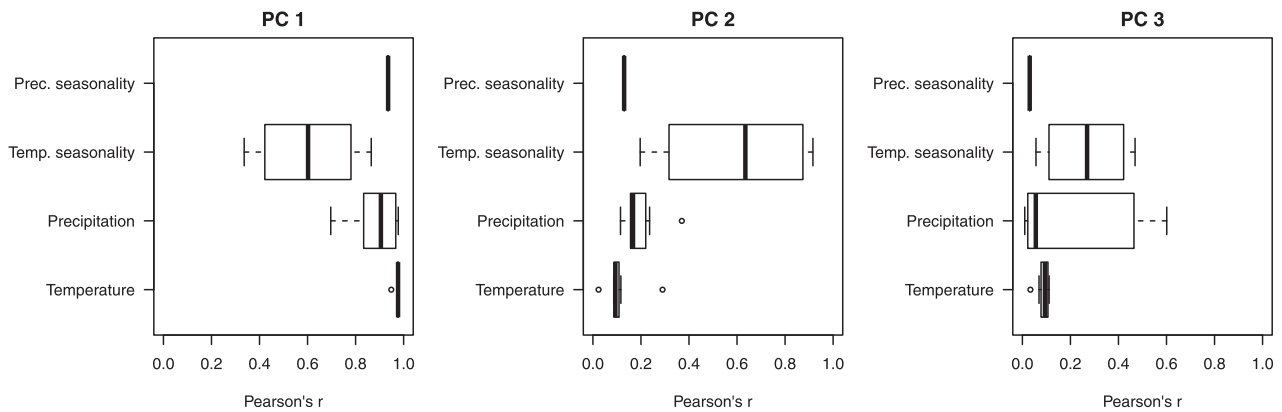


Figure 1 Correlation between original bioclimatic variables and the first three principal components of a climatic principal component analysis. ‘Prec. seasonality’, BIO15; ‘Temp. seasonality’, BIO2–BIO4 and BIO7; ‘Precipitation’, BIO12–BIO14 and BIO16–BIO19; ‘Temperature’, elevation, BIO1, BIO5, BIO6 and BIO8–BIO11. All codes as in WorldClim (<http://www.worldclim.org/bioclim>).

$$O_{ijk} = \int_{\min(k)}^{\max(k)} \sqrt{D_{ik} D_{jk}}$$

This is a measure of niche overlap based on the distribution of the species along environmental gradients. We considered it a good proxy of the competitive interactions between individuals of those species not only at the landscape scale but also at the plot and site scale.

Once we had obtained O_{ijk} for all species and environmental gradients, we calculated two versions of the average O_{ijk} for a given site with S species. First, by considering all the species in the site irrespectively of their co-occurrence within specific plots:

$$\text{site } \bar{O}_k = \frac{\sum_{i=1}^S \sum_{j=1}^S O_{ijk} N_i N_j}{\sum_{i=1}^S \sum_{j=1}^S N_i N_j}$$

where N_i and N_j are the total abundances of the species i and j , respectively, in the site. The expression above is the average niche overlap weighted by the expected number of interactions between individuals of species i and j , assuming perfectly random mixing of individuals within the site. This weighting gives such importance to the common species that in practice the exclusion of rare species (e.g. unseen species) does not influence the results. The second version of the average O_{ijk} for a given site took into account the apportioning of species within the m plots contained in the site:

$$\text{plot } \bar{O}_k = \frac{1}{m} \sum_{p=1}^m \frac{\sum_{i=1}^{S_p} \sum_{j=1}^{S_p} O_{ijk} n_{pi} n_{pj}}{\sum_{i=1}^{S_p} \sum_{j=1}^{S_p} n_{pi} n_{pj}}$$

where S_p , n_{pi} and n_{pj} are the species richness and the abundances of species i and j in plot p , respectively. The interpretation is similar to the first version above, but excludes from the calculations the niche overlap of those species that do not co-occur in the plots (even if they co-occur in the site). Both versions, along all environmental variables, constituted

the niche overlap matrix (\mathbf{O}), which contained 65 rows and $16 \times 2 = 32$ columns.

Both $\text{site } \bar{O}_k$ and $\text{plot } \bar{O}_k$ include in their calculation the abundances of the species in the site, so both can be influenced to some degree by the shape of the SAD in the site. This potential confounding effect is discussed in detail in Appendix S3, but was considered negligible in our study.

Estimation of productivity

We estimated productivity at plot level, which also served to characterize each site. To estimate the productivity of each plot, we first estimated its aboveground biomass (AGB) as the sum of the AGB of all the individuals in the plot. The AGB of each individual was estimated through the allometric equation of Chave *et al.* (2005) for tropical moist forests: $\text{AGB (Mg)} = 50.9 \times \text{wood density (g/cm}^3) \times \text{d.b.h.}^2 \text{ (cm}^2) \times \text{height (m)}$. The wood density data for each species was extracted from the Global Wood Density Database (Zanne *et al.*, 2009). If a species was not represented in the wood density database, or it was not fully identified in our dataset, we used the average for the genus (or family if we could not recover information at the genus level). In all cases we only employed data from the Neotropics. If a species belonged to a family that was not represented in the database, we used the average wood density of all the species found in the plots where the target species was present, weighted by the number of co-occurrences. Overall, 341 species received estimations at the species level, 1084 at the genus level, 480 at the family level and 180 at the community level. Then, we created a model to predict aboveground net primary productivity (ANPP) as a function of AGB. We employed the Michaelis–Menten asymptotic model fitted to the data on AGB and ANPP of tropical humid and tropical montane forests from Appendix 1 of Keeling & Phillips (2007). Only the most reliable data from this appendix were included (‘core data’, $n = 50$). The resulting model was

$$\text{ANPP} = \frac{36.31912 \times \text{AGB}}{201.14910 + \text{AGB}}$$

We calculated ANPP for each plot by applying this model to the previously obtained values of AGB. Finally, we estimated the ANPP for each site as the average ANPP of the plots contained in the site. Therefore, the productivity matrix (**P**) contained 65 rows and only one column.

Statistical analyses: variation partitioning

The variation in a response matrix (or variable) can be partitioned into components accounted for by two or more explanatory tables and their combined effects (Smith & Lundholm, 2010). In the case of two explanatory matrices (**A** and **B**), the variation in the response matrix can be decomposed into four fractions: (1) explained exclusively by **A**; (2) explained exclusively by **B**; (3) jointly explained by **A** and **B**; and (4) unexplained. Similarly, the variation partitioning analysis of four explanatory matrices (**R**, **H**, **O** and **P**) returns 16 fractions. Only four fractions of variation represent unequivocal support for underlying mechanisms (those exclusively explained by each of the four explanatory matrices). We conducted this analysis using function *varpart* of the R package *vegan* (Oksanen *et al.*, 2015).

All the calculations and analyses were conducted in R v. 3.2.2 and all packages were those contained in CRAN on 18 November 2015, as accessed through the *checkpoint* package (v. 0.3.15).

RESULTS

General results

Empirical SADs, unveiled SADs and fitted SADs were all more lognormal at higher elevations, ranging from invariably logseries curves below 1400 m to many symmetrical lognormal curves above 3000 m (Figs 2a & 3). Terrain roughness clearly covaried with elevation, with sites at higher elevations surrounded by rougher terrains (Fig. 2b). Sites were highly heterogeneous in general; for many sites and many variables, the standard deviation of a given parameter at the site scale was as large as the standard deviation of that same parameter in the whole study region (Fig. 2c, d). There were no elevational changes in the environmental heterogeneity of sites, regardless of the metric employed (Fig. 2d). Species co-occurring at plots and sites presented high overlap on their niches, at any elevation. Species overlapped 87–91% of their niches at site and plot scales for climatic PC1, PC2 and PC3, 92–95% of their niches for the edaphic variables, and 94–97% of their niches for the topographic variables (slope, northness and eastness). Species co-occurring at plots always presented more similar niches than those co-occurring at sites (Fig. 2e), but niche overlap at the plot and site scales were highly correlated (Pearson's $r = 0.92$, $P < 0.001$). From 1600 to 3500 m, the mean niche overlap decreased linearly, whereas trends below 1600 m in elevation were not clear. Finally, ANPP decreased linearly with elevation, from *c.*

20 Mg ha⁻¹ year⁻¹ at 1000 m to *c.* 5 Mg ha⁻¹ year⁻¹ at 3500 m (Fig. 2f). However, Amazonian and submontane sites below 1000 m presented similar (or slightly lower) ANPP values to lower montane sites at 1000–1500 m.

Results of the variation partitioning

The full model, including **R**, **H**, **O** and **P** as predictors, explained almost completely the observed variation in the SAD shape (adjusted $R^2 = 0.91$). Many of our predictor variables were correlated, making it impossible to determine the true underlying causes of 40.4% of the variation in SAD shape. However, 50.6% of the variation in the SAD shape could be assigned to individual predictor matrices: 28.4% was explained exclusively by niche overlap, 15.4% exclusively by environmental heterogeneity, 5.2% exclusively by landscape roughness and 1.6% exclusively by productivity (Fig. 4).

DISCUSSION

Neutral versus niche-based mechanisms explaining the variation in SAD shape

Landscape roughness (a proxy for neutral dispersal limitation, once the effect of environmental heterogeneity has been taken out), landscape heterogeneity (a proxy of the availability of niches), average niche overlap between species, and productivity, explained together 91% of the variation in SAD shape. More than half of the variation in SAD shape could be attributed to specific factors. Among the most important factors influencing the changes in SAD shape were variables linked to niche processes.

Niche overlap along different variables alone was enough to explain 28% of the variability in the Gambin α parameter across our study region (74% if considered jointly with the other factors analysed). Plant communities that exhibited lower niche overlap on average were characterized by more lognormal SADs. In places where species had high niche overlap, one or a few species outcompete the other species, and most species have very low abundances (reflected by logseries SADs). If more lognormal SADs imply more or better opportunities for more species, our results suggest that species attain such opportunities by having less niche overlap with co-occurring species. Thus, the Gambin α parameter would measure the ability of the species in the assemblage to efficiently fill the *n*-dimensional niche space available, as previously hypothesized (Ugland *et al.*, 2007; Matthews *et al.*, 2014).

Ugland *et al.* (2007) also suggested that the Gambin α could measure the dimensionality of the processes structuring a community, understood as environmental heterogeneity or complexity. Communities driven by many processes, or responding to multiple gradients (e.g. in highly heterogeneous or complex environments), would have a high α , and those driven by a few overriding processes (e.g. a major disturbance) would have a low α value (Ugland *et al.*, 2007; Matthews & Whittaker, 2014). In our system, however,

heterogeneity alone explained only 15% of the changes in SAD shape (23% if considered jointly with the other variables analysed). Thus, we found lower support for this interpretation of α in our study system. In addition, and contrary to our expectation, we did not find more complex and

heterogeneous abiotic conditions in the montane areas than in the lowlands. Consequently, the change in the SAD shape along the elevational gradient cannot exclusively be attributed to changes in sites' environmental heterogeneities or complexities. In any case, environmental heterogeneity should

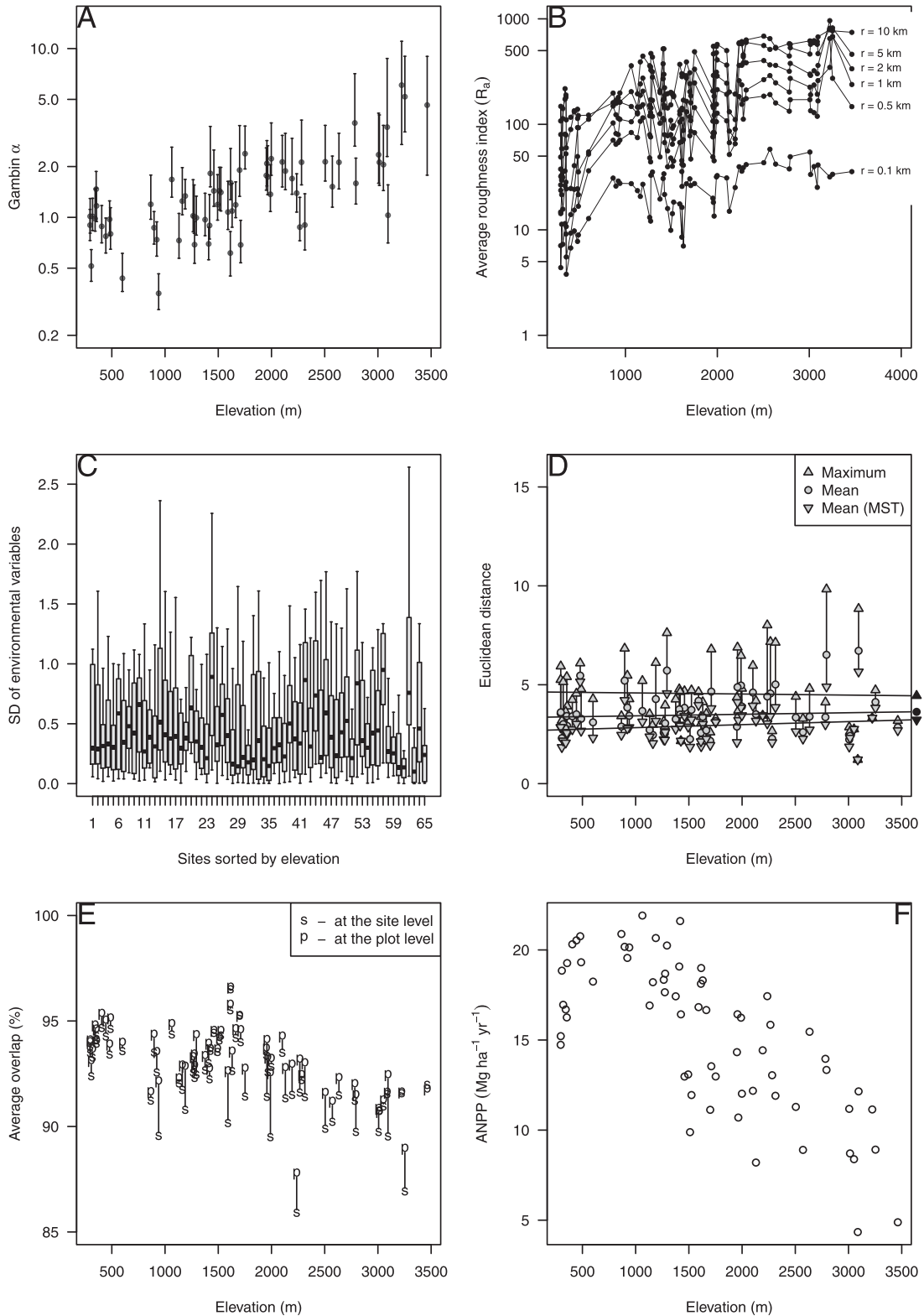


Figure 2 Elevational trends of forest characteristics along the elevational gradient in the Madidi region (Bolivia). (a) Shape of the species abundance distribution (SAD) curve, as summarized by the α parameter of a fitted Gambin model (lower α values characterize logseries SADs and high α values lognormal SADs). (b) Topographical complexity measured with the average surface roughness index (R_a). Different lines correspond to circles of increasing radius (R_a is always greater in larger circles). (c) Standard deviation of the plots at each site along each of the environmental variables considered in this study (three climatic axes and edaphic and topographic variables). A standard deviation of one in a given variable indicates that a site is as heterogeneous as the whole study region for that variable. (d) Environmental Euclidean distances between the plots of each site. 'Mean (MST)' is the mean distance between two plots along the minimum spanning tree connecting all the plots in a site. The horizontal lines are linear models fitted to each of the three measures of intra-site environmental distances along the elevational gradient. (e) Average niche overlap between the species co-occurring in each site ('s') and co-occurring in the plots of each site ('p'). Both measures are averaged across all the species and all the environmental variables considered. (f) Productivity measured as aboveground net primary productivity (ANPP) along the elevational gradient.

shape the SAD only if co-occurring species show non-overlapping niches. Thus, it is not surprising that we found average niche overlap to be the most important factor explaining the SAD shape, while environmental heterogeneity was secondary in importance.

Landscape roughness alone, interpreted as a proxy for neutral processes, explained only 5% of the variation in SAD shape. However, there is substantial correlation between landscape roughness and the other factors considered. In particular, 36% of the variation in SAD shape was explained jointly by landscape roughness and niche overlap (jointly including the other factors as well). As discussed extensively in the context of beta-diversity studies, joint fractions of explained variation cannot provide useful information in terms of causal mechanisms (Smith & Lundholm, 2010). Besides, landscape roughness could be correlated with unmeasured environmental heterogeneity. Overall, the results for

landscape roughness are not as informative as those for the other processes discussed.

One additional important limitation of the present study is the use of *realized* niches instead of fundamental niches. If only realized niches are measured, it is likely to be impossible to completely distinguish niche-based and neutral processes. This is because realized niches are calculated on the basis of the spatial distribution of the individuals of the species, without a real understanding of the preferences of the species. Consequently, realized niches might be reflecting, instead of causing, the spatial distribution of the species (Devictor *et al.*, 2010). Strong spatial clustering of the species might result in narrow and non-overlapping realized niches in sites where there is spatial autocorrelation of environmental conditions. Realized niches will tend to be narrow and non-overlapping where strong spatial clustering of species happens, even if such aggregation of conspecifics is caused by

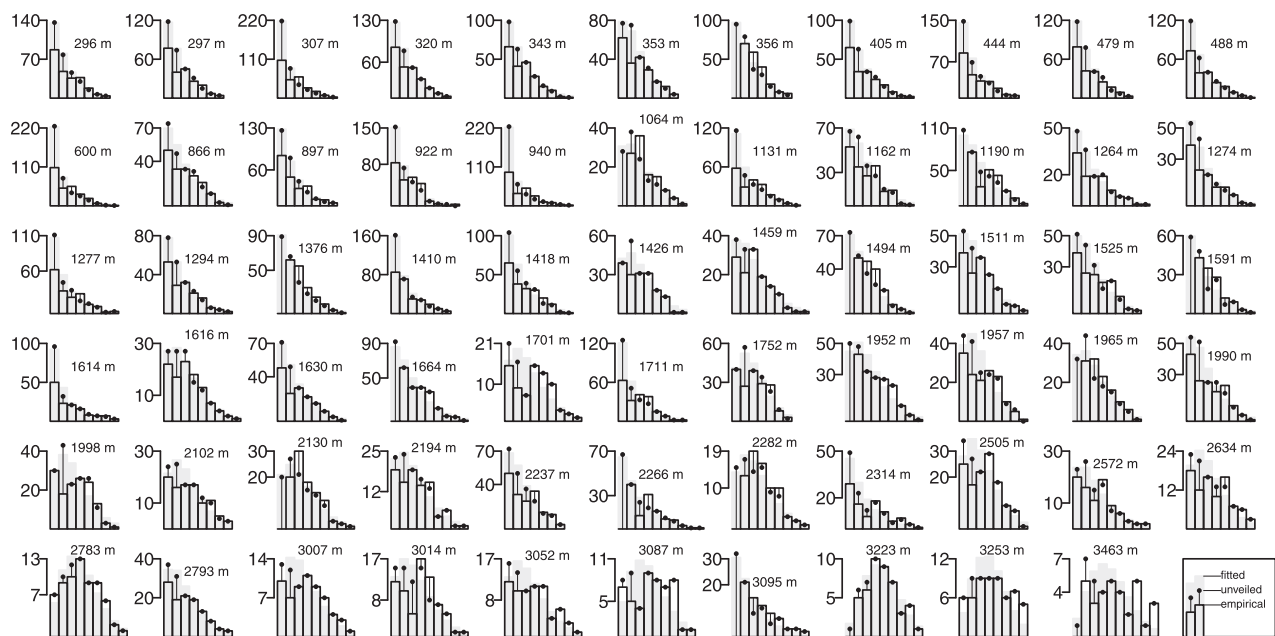


Figure 3 Frequency of species within doubling-abundance classes or Preston's octaves in 65 forest sites sorted by elevation in the Madidi region (Bolivia). Hollow bars are based on the empirical abundances; dots represent the frequency of species in abundance classes according to an unveiled SAD; grey backgrounds represent a Gambin model fitted to the distribution of unveiled abundances. Note the different scales of the y-axes.

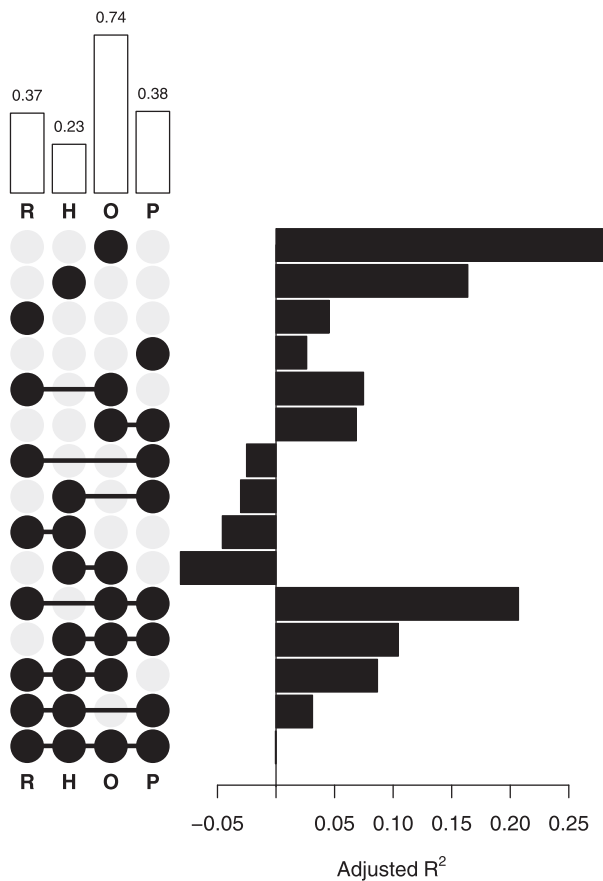


Figure 4 Results of the variation partitioning analysis. The main barplot with black bars represents the amount of variation in the shape of the species abundance distribution (SAD) that is explained by a given combination of factors, as indicated in the left panel. The small white barplot at the top of the left panel represents the amount of variation in the shape of the SAD explained by a given matrix (**R**, landscape roughness; **H**, environmental heterogeneity; **O**, mean niche overlap; **P**, productivity), regardless of it doing so alone or jointly with other matrices.

limited dispersion. In fact, any mechanism leading to strong autocorrelated samples of conspecific individuals (e.g. habitat preferences or dispersal limitation) would cause narrower and less overlapped realized niches, as well as more lognormal SADs (Green & Plotkin, 2007; Šizling *et al.*, 2009). The use of fundamental niches, albeit requiring experimental manipulation of the species, may be required to gain further insight into the processes driving the distribution and relative success of the species.

Higher productivity is not correlated with more even communities

Some researchers have reported more logseries SADs (i.e. less even communities) in less productive environments and with stronger environmental filtering (e.g. cold temperatures) (Hurlbert, 2004; Qiao *et al.*, 2015; Zhang *et al.*, 2015; but see Ulrich *et al.*, 2016). In our study system, however,

productivity decreases with elevation, implying a strong negative correlation between productivity and the Gambin α . This disagreement with previous literature may be caused because the other factors considered clearly override the expected effect of productivity on SAD shape (productivity alone explains only 1.6% of the variation in SAD shape). Furthermore, productivity can be thought of as the outcome of environmental factors (mainly elevation) and at least three community properties: (1) species richness; (2) SAD shape or evenness; and (3) differences between species (\approx trait divergence \approx niche overlap) (Nijs & Roy, 2000; Hillebrand *et al.*, 2008; Maire *et al.*, 2012). These three aspects of diversity are thought to cause changes in productivity through complementary effects (Loreau & Hector, 2001), but show very contrasting elevational trends. To understand the relationship between the SAD shape and productivity along the elevational gradient, these causal relationships and the covariation between the three factors, as well as the effect of environmental filtering, should be taken into account.

Is there an ecological basis for SADs?

The curve-fitting approach aims to extract meaningful ecological information from SADs. Examples are: (1) the dispersal rate (the parameter m of the zero sum multinomial model; Hubbell, 2001); (2) the relative probability with which available niches are colonized (the parameter k of the power fraction model; Tokeshi, 1996); and (3) the degree of conspecific frequency dependence in mortality rates (the parameter δ of the non-neutral model of Jabot & Chave, 2011). All these aspects are estimated by fitting specific models to species abundance data alone. However, and regardless of the goodness-of-fit of a given model to the SAD, it is important to check with external data how robust or plausible a given interpretation of the model parameters is. If fitted parameters are not confronted with external data (e.g. direct measures of seed dispersal or colonization rates in gaps), they cannot really inform us about the underlying ecological processes, and the link between a parameter and a process is necessarily just a subjective interpretation.

Despite its critical importance, studies relating ecological gradients to fitted parameters of SAD models are scarce, and have mostly focused on disturbance gradients (Dornelas *et al.*, 2009). The most complete study relating SAD parameters to external factors was that published by Yen *et al.* (2013). These authors evaluated the relationship between many habitat structure and landscape context covariates and the shape of multiple avifaunal SADs in south-eastern Australian forests. They did not find any relationship between the evaluated ecological covariates and SAD shape. Consequently, they concluded that there is no ecological basis for SADs, and that SAD shape is not useful for distinguishing among theories of assemblage structure. In contrast, as discussed above, we did find strong covariation between ecological and biological factors and SAD shape. We conclude that there are ecological bases for the SAD.

Two reasons could explain our disagreement with Yen *et al.* (2013). First, we considered a broader range of SAD shapes. The values for the metric of SAD shape employed by Yen *et al.* (2013) showed a coefficient of variation of 30% (mean = 0.885, SD = 0.267), while ours showed a coefficient of variation of 68% (mean = 1.56, SD = 1.06). This greatly increases our likelihood of finding correlations with external factors when such correlations exist. Second, and most importantly, we chose the covariates to check in the light of strong a priori hypotheses regarding SAD shape, whereas Yen *et al.* (2013) selected their variables based on previous work focused on species distributions in their study system (Yen *et al.*, 2011). It is important to realize that an environmental variable that influences the distribution of species does not need to influence SAD shape, and vice versa. On the one hand, two different SADs can be composed of totally different species in two contrasting communities and still show the same shape. On the other hand, a factor with no effect on compositional changes can cause changes in SAD shape. For example, landscape roughness predicts changes in SAD shape under a purely neutral perspective, i.e. even if the species living in flat areas are exactly the same as those living in rough areas the shape of the SAD itself is expected to change.

Nevertheless, describing statistical relationships between different aspects of community structure (e.g. species richness, mean niche overlap, SAD shape) is critical for establishing causal relationships among them and with the environment. In this regard, a major open question is the contribution of species richness to SAD shape (Locey & White, 2013). In our specific case it seems unlikely that statistical factors have greater weight than ecological factors (see Appendix S3 for an extensive discussion), but we found some covariation between different aspects of community structure and therefore our results need to be looked at with the appropriate caution.

CONCLUSION

Variation in SAD shape is often associated with purely statistical factors. However, we found a strong signature of ecological factors influencing the shape of the SAD curve. Lowland forests harbour many more species than montane forests while providing fewer opportunities for each species, probably because of more intense interspecific competition. Although montane forests show a similar diversity of microhabitats to lowland forests, montane species overlap less in their niche requirements with coexisting species than the species in the lowlands. This leads to better/more opportunities for the relatively few species that can withstand the stressful conditions of the montane forests. The Gambin α parameter, which roughly measures the ability of the species of the assemblage to efficiently fill the n -dimensional niche space available (Ugland *et al.*, 2007; Matthews *et al.*, 2014), could be useful when niche overlap or trait differences cannot be measured directly or reliably (Maire *et al.*, 2012). Finally, the

role of low migration and low niche overlap in topographically complex habitats, such as Andean forests, seem to be difficult to tell apart in both theory and practice. To disentangle the role of both processes in shaping plant communities in the Amazon–Andes system gradient represents a major open challenge, for which observational data might be not enough.

ACKNOWLEDGEMENTS

We are very grateful to L. E. Cayola, T. B. Miranda, M. Cornejo, V. W. Torrez, R. Seidel, N. Y. Paniagua, C. Maldonado, A. Araújo and J. Quisbert, who provided plot data for the present study. We thank the Dirección General de Biodiversidad, the Servicio Nacional de Áreas Protegidas, Madidi National Park, and local communities for permits, access and collaboration during fieldwork. Many students and volunteers provided indispensable assistance in the field and herbarium. Brian McGill, Alberto Jiménez-Valverde and three anonymous referees provided very valuable comments on the manuscript, for which we are grateful. We received financial support from the following institutions: Consejería de Educación (Comunidad de Madrid), National Geographic Society (8047-06, 7754-04), US National Science Foundation (DEB#0101775, DEB#0743457) and Universidad Autónoma de Madrid – Banco Santander.

REFERENCES

- Arellano, G., Jørgensen, P.M., Fuentes, A.F., Loza, M.I., Torrez, V. & Macía, M.J. (2016) Oligarchic patterns in tropical forests: role of the spatial extent, environmental heterogeneity and diversity. *Journal of Biogeography*, **43**, 616–626.
- Brown, J.H. (2014) Why are there so many species in the tropics? *Journal of Biogeography*, **41**, 8–22.
- Carson, W.P. & Schnitzer, S.A. (2009) *Tropical forest community ecology* (ed. by W.P. Carson and S.A. Schnitzer). Wiley-Blackwell, Oxford.
- Chao, A., Hsieh, T.C., Chazdon, R.L., Colwell, R.K. & Gotelli, N.J. (2015) Unveiling the species-rank abundance distribution by generalizing the Good–Turing sample coverage theory. *Ecology*, **96**, 1189–1201.
- Chase, J.M. (2010) Stochastic community assembly causes higher biodiversity in more productive environments. *Science*, **328**, 1388–1391.
- Chave, J., Andalo, C., Brown, S., Cairns, M.A., Chambers, J.Q., Eamus, D., Fölster, H., Fromard, F., Higuchi, N., Kira, T., Lescure, J.P.P., Nelson, B.W., Ogawa, H., Puig, H., Riéra, B. & Yamakura, T. (2005) Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia*, **145**, 87–99.
- Conlisk, J., Conlisk, E., Kassim, A.R., Billick, I. & Harte, J. (2012) The shape of a species' spatial abundance distribution. *Global Ecology and Biogeography*, **21**, 1167–1178.
- Cotgreave, P. & Harvey, P.H. (1994) Evenness of abundance in bird communities. *Journal of Animal Ecology*, **63**, 365–374.

- Devictor, V., Clavel, J., Julliard, R., Lavergne, S., Moullot, D., Thuiller, W., Venail, P., Villéger, S. & Mouquet, N. (2010) Defining and measuring ecological specialization. *Journal of Applied Ecology*, **47**, 15–25.
- Dornelas, M., Moonen, A.C., Magurran, A.E. & Bärberi, P. (2009) Species abundance distributions reveal environmental heterogeneity in modified landscapes. *Journal of Applied Ecology*, **46**, 666–672.
- Fisher, R.A., Corbet, S. & Williams, C. (1943) The relation between the number of species and the number of individual in a random sample of an animal population. *Journal of Animal Ecology*, **12**, 42–58.
- Gentry, A. (1995) Patterns of diversity and floristic composition in Neotropical montane forests. In *Biodiversity and conservation of Neotropical montane forests* (ed. by S. Churchill, H. Balslev, E. Forero and J.L. Luteyn). New York Botanical Garden, New York.
- Gerold, G. (2008) Soil, climate and vegetation of tropical montane forests – a case study from the Yungas, Bolivia. *Biodiversity and Ecology Series*, **2**, 137–162.
- Ghalambor, C.K., Huey, R.B., Martin, P.R., Tewksbury, J.J. & Wang, G. (2006) Are mountain passes higher in the tropics? Janzen's hypothesis revisited. *Integrative and Comparative Biology*, **46**, 5–17.
- Gradstein, S.R., Homeier, J. & Gansert, D. (2008) *The tropical mountain forest: patterns and processes in a biodiversity hotspot*. Göttingen Centre for Biodiversity and Ecology, Göttingen, Germany.
- Graham, C.H. & Fine, P.V.A. (2008) Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. *Ecology Letters*, **11**, 1265–1277.
- Green, J.J.L. & Plotkin, J.B.J. (2007) A statistical theory for sampling species abundances. *Ecology Letters*, **10**, 1037–1045.
- Hijmans, R.J. (2014) *raster: geographic data analysis and modeling*. R package version 2.2-31. Available at: <http://CRAN.R-project.org/package=raster>.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
- Hillebrand, H., Bennet, D.M. & Cadotte, M.W. (2008) Consequences of dominance: a review of evenness effects on local and regional ecosystem processes. *Ecology*, **89**, 1510–1520.
- Hubbell, S. (2001) *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, NJ.
- Hurlbert, A.H. (2004) Species–energy relationships and habitat complexity in bird communities. *Ecology Letters*, **7**, 714–720.
- Jabot, F. & Chave, J. (2011) Analyzing tropical forest tree species abundance distributions using a nonneutral model and through approximate Bayesian inference. *The American Naturalist*, **178**, E37–E47.
- Keeling, H.C. & Phillips, O.L. (2007) The global relationship between forest productivity and biomass. *Global Ecology and Biogeography*, **16**, 618–631.
- Locey, K.J. & White, E.P. (2013) How species richness and total abundance constrain the distribution of abundance. *Ecology Letters*, **16**, 1177–1185.
- Loreau, M. & Hector, A. (2001) Partitioning selection and complementarity in biodiversity experiments. *Nature*, **412**, 72–76.
- McGill, B.J. (2003) Does Mother Nature really prefer rare species or are log-left-skewed SADs a sampling artefact? *Ecology Letters*, **6**, 766–773.
- McGill, B.J. (2011) Species abundance distributions. In *Biological diversity: frontiers in measurement and assessment* (ed. by A.E. Magurran and B.J. McGill). Oxford University Press, Oxford.
- McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas, M., Enquist, B.J., Green, J.L., He, F., Hurlbert, A.H., Magurran, A.E., Marquet, P.A., Maurer, B.A., Ostling, A., Soykan, C.U., Ugland, K.I. & White, E.P. (2007) Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, **10**, 995–1015.
- Maire, V., Gross, N., Börger, L., Proulx, R., Wirth, C., Pontes, L.D.S., Soussana, J.F. & Louault, F. (2012) Habitat filtering and niche differentiation jointly explain species relative abundance within grassland communities along fertility and disturbance gradients. *New Phytologist*, **196**, 497–509.
- Matthews, T.J. & Whittaker, R.J. (2014) On the species abundance distribution in applied ecology and biodiversity management. *Journal of Applied Ecology*, **52**, 443–454.
- Matthews, T.J., Borregaard, M.K., Ugland, K.I., Borges, P. A V., Rigal, F., Cardoso, P. & Whittaker, R.J. (2014) The Gambin model provides a superior fit to species abundance distributions with a single free parameter: evidence, implementation and interpretation. *Ecography*, **37**, 1002–1011.
- Mehlich, A. (1984) Mehlich 3 soil test extractant: a modification of Mehlich 2 extractant. *Communications in Soil Science and Plant Analysis*, **15**, 1409–1416.
- Nijs, I. & Roy, J. (2000) How important are species richness, species evenness and interspecific differences to productivity? *Oikos*, **88**, 57–66.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H. & Wagner, H. (2015) *vegan: community ecology package*. R package version 2.3-1. Available at: <http://CRAN.R-project.org/package=vegan>.
- Parra-Olea, G., Windfield, J.C., Velo-Antón, G. & Zamudio, K.R. (2012) Isolation in habitat refugia promotes rapid diversification in a montane tropical salamander. *Journal of Biogeography*, **39**, 353–370.
- Preston, F.W. (1948) The commonness, and rarity, of species. *Ecology*, **29**, 254–283.
- Qiao, X., Jabot, F., Tang, Z., Jiang, M. & Fang, J. (2015) A latitudinal gradient in tree community assembly processes evidenced in Chinese forests. *Global Ecology and Biogeography*, **24**, 314–323.

- Reeuwijk, L.P. van (2002) *Procedures for soil analysis*. Technical paper 9. International Soil Reference and Information Centre, Wageningen, The Netherlands.
- Šizling, A., Storch, D., Sizlingova, E., Reif, J. & Gaston, K.J. (2009) Species abundance distribution results from a spatial analogy of central limit theorem. *Proceedings of the National Academy of Sciences USA*, **106**, 6691–6695.
- Smith, T.W. & Lundholm, J.T. (2010) Variation partitioning as a tool to distinguish between niche and neutral processes. *Ecography*, **33**, 655.
- Stevens, G.C. (1992) The elevational gradient in altitudinal range: an extension of Rapoport's latitudinal rule to altitude. *The American Naturalist*, **140**, 893–911.
- Tokeshi, M. (1996) Power Fraction: a new explanation for species abundance patterns in species-rich assemblages. *Oikos*, **75**, 543–550.
- Ugland, K.I., Lambshead, P.J.D., McGill, B., Gray, J.S., Dea, N.O., Ladle, R.J. & Whittaker, R.J. (2007) Modelling dimensionality in species abundance distributions: description and evaluation of the Gambin model. *Evolutionary Ecology Research*, **9**, 313–324.
- Ulrich, U., Kusumoto, B., Shiono, T. & Kubota, Y. (2016) Climatic and geographic correlates of global forest tree species–abundance distributions and community evenness. *Journal of Vegetation Science*, **27**, 295–305.
- Vitousek, P.P.M. (1998) The structure and functioning of montane tropical forests: control by climate, soils, and disturbance. *Ecology*, **79**, 1–2.
- Yen, J.D.L., Thomson, J.R., Veski, P.A. & Mac Nally, R. (2011) To what are woodland birds responding? Inference on relative importance of in-site habitat variables using several ensemble habitat modelling techniques. *Ecography*, **34**, 946–954.
- Yen, J.D.L., Thomson, J.R. & Mac Nally, R. (2013) Is there an ecological basis for species abundance distributions? *Oecologia*, **171**, 517–525.
- Young, K.R. (1995) Biogeographical paradigms useful for the study of tropical montane forests and their biota. In *Biodiversity and conservation of Neotropical montane forests* (ed. by S. Churchill, H. Balslev, E. Forero and J.L. Luteyn). New York Botanical Garden, New York.
- Zanne, A.E., Lopez-Gonzalez, G., Coomes, D.A., Ilic, J., Jansen, S., Lewis, S.L., Miller, R.B., Swenson, N.G., Wiemann, M.C. & Chave, J. (2009) *Global wood density database*. Dryad Digital Repository. Available at: <http://dx.doi.org/10.5061/dryad.234>.
- Zhang, J., Qiao, X., Liu, Y., Lu, J., Jiang, M., Tang, Z. & Fang, J. (2015) Species–abundance distributions of tree species varies along climatic gradients in China's forests. *Journal of Plant Ecology*, doi: 10.1093/jpe/rtv055. [This article has been retracted because of plagiarized sentences; see *Journal of Plant Ecology* **9**, 240.]

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site:

Appendix S1 Description and R code for the unveiling and model fitting of species abundance distributions.

Appendix S2 Grouping of plots into sites and characteristics and location of the sites.

Appendix S3 Statistical relationships between species richness, niche overlap and shape of the species abundance distribution.

BIOSKETCH

G. Arellano is a community ecologist who focuses on the study of patterns of commonness and rarity, community assembly, and the determinants of beta diversity. He is part of the Madidi Project, a long-term study of plant biodiversity in the Bolivian Andes (<http://www.mobot.org/madidi/>).

Editor: Alberto Jiménez-Valverde