Using species abundance distribution models and diversity indices for biogeographical analyses

Simone Fattorini[1,2], François Rigal[1], Pedro Cardoso[1,3], Paulo A.V. Borges[1]

[1]CE3C – Centre for Ecology, Evolution and Environmental Changes / Azorean Biodiversity Group and Universidade dos Açores - Departamento de Ciências Agrárias, 9700-042 Angra do Heroísmo, Açores, Portugal
[2]Water Ecology Team, Department of Biotechnology and Biosciences, University of Milano Bicocca, Milan, Italy.
[3]Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland

* corresponding author, E-mail: *simone.fattorini@gmail.com*

Abstract

We examine whether Species Abundance Distribution models (SADs) and diversity indices can describe how species colonization status influences species community assembly on oceanic islands. Our hypothesis is that, because of the lack of source-sink dynamics at the archipelago scale, Single Island Endemics (SIEs), i.e. endemic species restricted to only one island, should be represented by few rare species and consequently have abundance patterns that differ from those of more widespread species. To test our hypothesis, we used arthropod data from the Azorean archipelago (North Atlantic). We divided the species into three colonization categories: SIEs, archipelagic endemics (AZEs, present in at least two islands) and native non-endemics (NATs). For each category, we modelled rank-abundance plots using both the geometric series and the Gambin model, a measure of distributional amplitude. We also calculated Shannon entropy and Buzas and Gibson’s evenness. We show that the slopes of the regression lines modelling SADs were significantly higher for SIEs, which indicates a relative predominance of a few highly abundant species and a lack of rare species, which also depresses diversity indices. This may be a consequence of two factors: (i) some forest specialist SIEs may be at advantage over other, less adapted species; (ii) the entire populations of SIEs are by definition concentrated on a single island, without possibility for inter-island source-sink dynamics; hence all populations must have a minimum number of individuals to survive natural, often unpredictable, fluctuations. These findings are supported by higher values of the α parameter of the Gambin mode for SIEs. In contrast, AZEs and NATs had lower regression slopes, lower α but and higher diversity indices, resulting from their widespread distribution over several islands. We conclude that these differences in the SAD models and diversity indices demonstrate that the study of these metrics is useful for biogeographical purposes.

Key words: Arthropods; diversity; endemism; Gambin model; macroecology; rank-abundance curves

1. Introduction

Community ecology is dominated by idiosyncratic results and few general laws (Lawton 1996, Storch et al. 2008, Beck et al. 2012, Passy 2012). Many ecological processes are contingent on the temporal and spatial scale in which they operate (Gaston and Lawton 1990, Gaston and Blackburn 1996, Fraterrigo and Rusak 2008), which makes it difficult to identify recurrent patterns and processes. Moreover, completely different processes can lead to very similar patterns (Gaston et al. 2000, Starzomski et al. 2008), which can impede inferring processes from patterns.

Most work that has been done in community ecology to characterize species assemblages according to their abundance relied on the study of patterns of species abundance distributions (SADs), because they are considered one of the most basic descriptors of a community (Magurran 2003, McGill et al. 2007). In general, local communities tend to have mostly rare and only few common species (Magurran 2003, McGill et al. 2007), and there are many proposed explanations for the underlying processes, ranging from niche-based competition to neutral processes (Hubbell 2001, Gravel et al. 2006, Vergon et al. 2009, Matthews and Whittaker 2014a).

The study of species abundances in community ecology has led to the development of two main approaches that are now routinely used to describe communities: species abundance distribution models, typically represented by rank-abundance plots, and diversity indices (Hayek and Buzas 2010). Although the mechanisms leading to particular distribution models or diversity values are difficult to be established, the study of community structure by species abundance distribution models and diversity indices has proven useful in shedding light on a variety of community characteristics, including how they are affected by environmental disturbances (Magurran 1988, 2004, Hayek and Buzas 2010, Dornelas 2010).

The same statistical tools used in community ecology could be profitably used to investigate SADs of assemblages defined by criteria different from those used to define communities, i.e. groups of organisms occupying a particular area, usually interacting with each other and their environment. For example, species can be grouped according to their persistence (such as core versus satellite, or native versus vagrant species; see Holloway 1996, Magurran and Henderson 2003, Matthews et al. 2014a) or colonization history (e.g., endemic versus non endemic), and one can ask how species abundances change within and among these categories. This may be important to develop hypotheses about the effect of species' history, dispersal and colonization ability on community structure. However, as far as we know, no research has attempted to use species abundance distribution models and diversity indices to compare species groups defined on the basis of broad biogeographical categories such as those that express their level of endemicity.

Oceanic islands are good models to examine whether SADs and diversity indices can clarify how biogeographical categories can influence relative species abundances. Indeed, oceanic islands form discrete isolated spaces where the biogeographical status of their species can be established easier than in continental systems, where continuous ecological gradients make biogeographical characterization more subjective (after all, any species can be considered as endemic to a certain area, depending on the way the area is defined, i.e. it is a question of scale, Laffan and Crisp 2003). In general, oceanic islands contain a large number of endemic species, either through *in situ* speciation (neoendemics), or through the extinction of species outside the islands (palaeoendemics) (Lomolino et al. 2010). These endemic species may occupy several islands (Archipelagic Endemics, hereafter AREs) or be restricted to one particular island (Single Island Endemics, hereafter SIEs). Such species, together with other native but non-endemics (hereafter NATs) form the three distinct biogeographical categories into which the species inhabiting an archipelago can be grouped. A further category may be represented by exotic (introduced) species, which are however not considered here because their presence is due to human introduction.

Our basic hypothesis is that these three biogeographical categories (SIEs, AREs, and NATs) are characterized by different patterns of relative abundance of individuals. In particular, we hypothesize that rare species (i.e. species having small size populations) should be less common among the SIEs than in other categories, because their distribution restricted to single islands determines the lack of source-sink dynamics necessary for maintaining viable populations of species with reduced population size (i.e. rescue effects, see Hanski 1994, Sutherland et al. 2012)., Because of the small number of individuals that form their populations, rare species are more sensitive to demographic accidents that increase the likelihood of local extinction (e.g. Allee's effect; Landle et al. 2003, Freckleton et al. 2005), which leads, in turn, to a lower probability of persistence on single islands, i.e. as SIEs. Therefore, the SADs of SIEs should be characterized mainly by a predominance of abundant species, well adapted to specific island environmental conditions, and few rare species (see, for example, Borges et al. 2006). The few rare SIEs species may be signatures of past extinctions of archipelagic endemics formerly distributed on more islands but now restricted to only one island, or SIEs really evolved on single islands (see Borges et al. 2006 and Gaston et al. 2006). By contrast, since AREs and most NATs are distributed on different islands of an archipelago, inter-island source-sink dynamics should maintain a substantial amount of rare species (Freckelon et al. 2005, Matthews et al. 2014a). In addition, it is well known that there is a correlation between species range size and abundance, with widespread species being also more abundant (a pattern known as the positive interspecific abundance-occupancy relationship, Gaston and Lawton 1990, Gaston et al. 2000, 2006). Therefore, AREs and NATs should include both rare species, which occur on few islands and have low mean abundance, and common species, which occur on most islands and are abundant on most of them. Following Taylor's Power Law (see Gaston et al. 2006), this should generate, for AREs and NATs, SADs less strongly affected by the dominance of few, extremely abundant species and, hence, less steep rank abundance patterns.

To test these predictions, we analysed the rank-abundance distribution and diversity patterns of the arthropods inhabiting the Azorean Islands, a volcanic archipelago located in the North Atlantic. Using the Azores as a model system, we expect to demonstrate the utility of SADs and diversity indices to disentangle fundamental ecological processes among groups of species having different biogeographical origins.

2. Materials and methods

2.1. Fieldwork

The Azorean archipelago is located in the North Atlantic, 37° to 40° N and 25° to 31° W. It comprises nine main islands and some small islets aligned along a roughly WNW-ESE trend. All islands are of volcanic origin (ranging from 250,000 years B.P. in Pico - Demand et al. 1982 - to 8.12Myr B.P. in Santa Maria - Abdel et al. 1975). Documents from the 15th century suggest that native vegetation almost completely covered all of the islands until when the first human settlements were established. Clearing for wood, agriculture and pasture have markedly reduced the native forests, which are now largely restricted to high and steep areas (Martins 1993, Gaspar et al. 2008). Here, we focused on arthropod species, which are the most diverse animal group in the Azores (Borges et al. 2010) and for which we have large sets of species standardized abundance data. Species abundance data used in the present study were collected during several field surveys carried out from 1999 to 2007. Arthropods were collected using standardized protocols targeting both soil and canopy communities. Sampling was conducted in 100 sites in 18 native forest fragments in seven islands (BALA project; Borges et al. 2005, 2006, Ribeiro et al. 2005, Gaston et al. 2006, Gaspar et al. 2008). We laid out a minimum of four independent 150-m long, 5-m wide transects in each forest fragment, with more transects in larger fragments. We sampled the epigean arthropod fauna using 30 pitfall traps (100 ml each) per transect for at least a two-week period during summer months. Half of the pitfall traps contained an attractive solution (Turquin solution, Turquin 1973), while the remaining had a non-attractive solution with a small proportion of ethylene glycol. The two types of traps were placed alternately. Canopy arthropods from woody shrub and tree species were sampled using a beating tray in the same period of trap functioning. Ten beating samples were taken at 10-m intervals from each of the three most dominant woody plant species in the site (Ribeiro et al. 2005) covering part of the original 150-m transect.

Arthropod sampling in non-native habitats was based on pitfall trapping only, following a design similar to the one described for native forests, and involved the following land use types (see also Florencio et al. 2013): natural grasslands (20 sites in five islands), peat bogs (4 sites in Terceira Island only), exotic forests (37 sites in four islands), semi-natural pastures (29 sites in four islands) and intensively managed pastures (38 sites in four islands). Eight out of nine islands were sampled (i.e., all except Corvo). A total of 454 species (230 species - 51% - of which endemic or native) belonging to 21 arthropod orders were collected. All arthropods were initially sorted into morphospecies by a team of trained parataxonomists and later identified to species (with few exceptions) by expert taxonomists. Further information on the sampling methodology adopted can be found in Borges et al. (2005), Ribeiro et al. (2005) and Gaspar et al. (2008).

2.2. Datasets

Based on their geographical distributions, the arthropods of the Azorean Islands were grouped into the three aforementioned categories: single island endemics (SIEs), archipelagic endemics (AREs) and native, non-endemics (NATs). For consistence with previous papers, we used the acronym AZE (Azorean endemics) for ARE species. Our basic approach was to test if the three colonization categories had different SADs. Because we collected arthropods from different habitats, we conducted two sets of analyses to investigate the possible confounding effects of mixing species that are present in different habitats. Firstly, we analysed data from native forests only (100 sites in 18 native forest fragments in seven islands). For this single habitat dataset, we analysed data from both epigean and arboreal species separately (for which we have 99 sites). Secondly, we analysed a more comprehensive dataset for epigean arthropods pooling samples from the different habitats described above, using pitfall data only. Therefore, all the following analyses were conducted for native forest soil and canopy samples and for soil samples across different habitats separately (three sets of analyses).

2.3. Species abundance distribution

We analysed SADs using rank-abundance curves (Magurran 1988, 2004). In the rank-abundance curves all the species in a sample are ranked from the most abundant to the least abundant. Each species has a rank plotted on the *x*-axis, and an abundance plotted on the *y*-axis. Thus the abundance for the most abundant species is plotted first, then the next most common and so on.

Several *a priori* established distributions can be used to model empirical rank-abundance curves (e.g. Ulrich et al. 2010). Most of these theoretical curves are in fact very similar and more than one theoretical model may provide a good fit to the same empirical data. Therefore, we decided to compare our data against two commonly used theoretical SAD models: the geometric series (GS) and the MacArthur broken stick (BS) (Büssenschütt and Pahl-Wostl 1999, Fattorini, 2005, He and Tang 2008) using the regression approach described by Fattorini (2005). Among all proposed SAD models, the GS represents the least equitable distribution (i.e. the distribution with lower evenness among species) and it is known to provide a good fit to simple communities characterized by the high dominance of a few species (Giller 1984, Magurran 1988, 2004). On the opposite, most equitable empirical distributions should be modelled by the broken stick model (Higgins and Strauss 2008). It is well known that the BS model is theoretically questionable and communities rarely are correctly characterized by such model (Wilson 1993, Hayek and Buzas 2010). Yet, the BS model is useful in comparative analyses because it represents a simple benchmark in opposition to the GS. Thus, we fitted the GS and BS models to each colonization category to assess if community structure followed a simple, highly dominated (geometric series), or a more complex and balanced (broken stick) abundance distribution. Note that another species abundance distribution model widely used in community ecology for communities dominated by few species is the log-series, a probability distribution which results from the Poisson sampling of a gamma distribution after a certain relevant limit is taken (May 1975, Matthews and Whittaker 2014a). However, the geometric series and the log-series abundance distributions are interrelated and are two representations of, essentially, the same underlying abundance distribution, either as a rank-abundance curve or as the typical frequency versus abundance curve (May 1975, Solé et al. 2004). The ubiquity of the log-series across scales, but particularly for small samples, is well recognized (Magurran 2005). However, we preferred to model the geometric series (instead of the log-series) because it can be easily expressed by a regression line in rank abundance plots, which facilitates comparisons (see Fattorini 2005).

GS and BS models have long been used in community ecology to test hypotheses of niche partitioning (see Fattorini 2005). In our analyses, we used both only as statistical approaches to describe and compare SADs among species belonging to different biogeographical categories (i.e. in a macroecological perspective), with no direct implications regarding niche partitioning. GS and BS were fitted to the data (raw abundance) using regression analyses as described in Fattorini (2005). With the geometric series, if a log scale is used for abundance, the species exactly fall along a straight line, according to the equation $\log A = b_0 + b_1 R$, where A is the species abundance, R is the respective rank, and b_0 and b_1 are optimized fitting parameters. In the broken stick, the distribution follows almost exactly a straight line if a log scale is used for the rank axis only. Using this approach, it is possible to use the regression slope to compare different species assemblages that follow the same rank-abundance distribution (see Fattorini 2005). In regression analyses, species with the same abundance received the same rank. Then, slope comparisons were assessed by performing analysis of covariance (ANCOVA). If significant differences were detected, multiple comparisons tests were implemented by using the R package multcomp (Hothorn et al. 2008). We compared both fits with the corrected Akaike Information Criterion (AICc) and the R^2 goodness-of-fit statistics using the formula of Kvalseth (1985). We found that, in all cases, the geometric series provided a much better fit than the broken-stick (See Supplementary material Appendix 1). Then, only GS was considered for the subsequent analyses.

SADs are known to change as a function of sample size (Preston 1962). Because sample size (number of individuals) varied among categories, we built rarefied models by resampling the data of SIEs, AZEs and NATs to a fixed number of 1000 individuals. This was done 1000 times per dataset. Therefore, the linear model for the GS (see above) was fitted for each rarefied run in order to build the 95% confidence limits for the slopes of SIEs, AZEs and NATs, defined as the 0.025 and 0.975 percentiles. We used these confidence limits to assess if the curves were distinct from one another and to test whether these results were consistent with those obtained by the ANCOVA.

In addition to these traditional models, we also applied the Gambin model proposed by Ugland et al. (2007) and which has proved to provide good fits to SADs (Matthews et al. 2014a). In this model, which is based on the gamma distribution, a single variable, α, determines the shape of the gamma distribution and the 'dimensionality' of the sampled community. A small α indicates a unimodal distribution with a positive skew, i.e. a high density at small abundance values, whereas a high α indicates a distribution closer to normal on a log scale of abundances (Ugland et al. 2007, Matthews et al. 2014b, Matthews and Whittaker 2014b). The α parameter has also been argued to reflect the complexity of a community's interactions with its environment (Ugland et al. 2007). The goodness of fit was evaluated using Pearson's chi-square test. As it is mentioned for the log-series analysis, rigorous comparisons of α values across samples should be based on keeping sample size constant across samples (Matthews et al. 2014b). We then re-calculated α for SIEs, AZEs and NATs by resampling the data to a fixed number of 1000 individuals. This was made 1000 times per dataset. Thus, we used the 95% confidence limits computed from 1000 resampled α to assess whether α differed between the three biogeographical categories.

SAD analyses were implemented within the R programming environment (R Development Core Team, 2014) using the package *sads* (Prado and Miranda, 2014) and *gambin* (Matthews et al. 2014b). Authors wrote the code to perform rarefactions for the GS analysis.

2.4. Diversity indices

On the basis of the extensive review of diversity indices performed by Hayek and Buzas (2010), both Shannon entropy and Buzas and Gibson's evenness were selected to express diversity of the three biogeographical categories (SIE, AZE and NAT). Shannon entropy was calculated as:

$H = -\sum \frac{n_i}{n} \ln\left(\frac{n_i}{n}\right)$, where n_i was number of individuals of taxon i, and n was the total number of individuals across all taxa. *H* ranges from 0 (one taxon dominates the community completely) to high values for communities with many taxa, each with similar abundance. Buzas and Gibson's evenness is expressed as $E = e^H/S$ (where S is the number of species, and H is Shannon entropy). This index varies from 0 (highest dominance by a single species) to 1 (all species have the same abundance). These two measures are particularly effective in encapsulating many aspects of diversity into a single value (Hayek and Buzas, 2010). To compare diversity indices of the three categories in pairwise combinations, we applied a bootstrapping procedure (9999 random samples) to calculate 95% confidence intervals for each category. Diversity indices and the respective bootstrapping procedures were computed using PAST 3.0 (Hammer et al. 2013).

3. Results

A total of 124 (N = 41,001 individuals) and 119 (N = 10,702) species from the canopy and soil of native forests respectively, and 230 (N = 11,884) species from soil of all habitats were considered in our study. In native forests, for canopy species, 14 (N = 4,644) species were SIEs, 46 (N = 21,550) were AZEs and 64 (N = 14,807) were NATs, while for epigean species, 30 (N = 1,361) species were SIEs, 34 (N = 2,073) were AZEs and 55 (N = 7,268) were NATs. In all habitats, 33 (N = 1,145) were SIEs, 54 (N = 3,666) were AZEs and 143 (N = 7,073) were NATs.

We found that, in all cases, SADs were best fitted by a GS model (Table S1). In general, slopes of GS regression lines for both native forest and all habitats decreased with increasing geographical distribution. For soil samples in native forest, difference between slopes was marginally non-significant (ANCOVA interaction categories × species rank: $F_{2,113} = 2.914$, p = 0.058) while for both canopy samples in native forest and soil samples in all habitats, slopes varied among biogeographical categories (ANCOVA interaction categories × species rank: $F_{2,118} = 76.133$,

$P < 0.001$ and $F_{2,128} = 208.730$, $p < 0.001$, respectively). For canopy samples in native forest, slope of SIEs was significantly steeper than the slopes of both AZEs and NATs, while no significant difference was found between the two latter categories. For soil samples in all habitats, the three slopes were significantly different from each other (multiple comparisons $p < 0.05$) with AZEs having the steepest slope and NATs the lowest. Overall, results of the resampling analysis (Figure 1) were consistent with the results of ANCOVAs. The slopes of AZEs and NATs in the canopy samples of native forests did not show any significant difference (AZE CI95% = [-0.176;- 0.133]; NAT CI95% = [-0.164 ; - 0.121]) but were significantly less steep than the slope for SIEs (SIE CI95% = [-0.277; -0.195]). For the soil samples in native forest, slopes of the SIEs, AZEs and NATs did not show any difference (SIEs CI95% = [-0.141; - 0.100], AZEs CI95% = [-0.176; -0. 0.133], NATs CI95% = [-0.191; - 0.128]). However, in soil samples in all habitats, the slope between SIEs and AZEs did not differ (SIE CI95% = [-0.217; - 0.166], AZE CI95% = [-0.315; - 0.201]) but both were significantly steeper than NATs (NAT CI95% = [-0.128; - 0.095]).

The Gambin model provided very robust fits (Pearson's chi-square $p > 0.05$) in all cases. For canopy and soil samples in native forest, the Gambin α parameter increased in the order NATs < AZEs < SIEs with NATs having the most left skewed abundance distribution (Table 2 and Fig. 1). For soil samples with all habitats merged, the Gambin α parameter increased in the order AZEs < NATs < SIEs (Table 2 and Fig. 1). Moreover, an analogous pattern was found after standardizing the Gambin α parameter to the same sample size (N=1,000) (Table 2 and Fig.1). For canopy samples in native forests, all categories differed from each other, whereas in soil samples from native forests and all habitats, SIEs differed from AZEs and NATs, but the two latter categories did not differ (Table 2 and Fig. 1).

The Shannon diversity of canopy arthropods in native forests increased with range size: SIEs were the category with the lowest diversity, whereas NATs formed the category with the highest diversity (Table 2). Evenness followed a reversed pattern, with SIEs having the highest evenness and NATs and AZEs similarly low values (Table 2). For soil epigean species in native forests, AZEs and NATs had similarly higher Shannon values than SIEs, whereas evenness decreased from SIEs to NATs. When soil arthropods from all habitats were considered, diversity was highest for NATs and lowest for the AZEs (Table 2). SIEs were the category with highest evenness, whereas AZEs formed the category with the lowest evenness, with NATs attaining an intermediate value (Table 2).

4. Discussion

The study of SADs has a long history in community ecology (e.g. Preston 1948, Tokeshi 1993, McGill et al. 2007). Recent theoretical and computational advancements have largely broadened the scope of SAD applications, thus encouraging a more widespread use of SADs in ecological research (Mattthews et al. 2014a, 2014b, Matthews and Whittaker 2014a, 2014b). However, as far as we are aware, no previous study analyzed SAD patterns by grouping species into biogeographical categories. We believe that SAD analysis can be a useful tool to investigate biogeographical patterns if species are grouped according to their endemicity levels and to illustrate this possibility we analyzed a large and standardized island dataset, the arthropods of the Azorean archipelago. Confirming our predictions, the three biogeographical categories (single island endemics SIEs, Azorean endemics AZEs, and native non-endemics NATs) showed different SAD patterns. Regression lines modelling geometric series had the highest slopes for the SIEs of canopy arthropods. This indicates that SIEs patterns are strongly influenced by the high abundance of a few species, whereas species with extremely low abundance are virtually absent. This is consistent with the low diversity (Shannon entropy) value but higher evenness of SIEs (i.e., SIE species abundances vary “regularly” from a species to another, as rare species are few). For NATs, abundances varied in a less “even” manner (low evenness), but the influence of the most abundant species was less marked (lower slopes of regression lines). Our results concerning the arthropods sampled in the canopy habitat are particularly interesting, since this is a habitat characterized by a dominance of indigenous species and very few exotic species (Borges et al. 2008, Borges and Wunderlich 2008), whereas the soil fauna in Azores has a high component of exotic species (Borges et al. 2006, Cardoso et al. 2009). Therefore, we may interpret the most obvious differences among the three species categories found within canopy samples as a true natural phenomenon, whereas the fact that SAD slopes for soil samples did not show any difference among the three categories can be interpreted as a consequence of high levels of disturbance and homogenization effects in the soil fauna (Florencio et al. 2013). Results for soil samples in all habitats showed an intermediate pattern, with the slopes of both SIEs and AZEs being not differentiated after resampling, but both were steeper than the slope of NATs (Table 2, 3 and Fig. 1). To explain these results we must consider how human disturbance may have very different impacts on mainland and island ecosystems. In general, on mainland ecosystems, creation of new habitat patches as a result of human disturbance tends to increase the species richness (up to a point). This is due to the fact that species originally not present in a certain area can enter the original community from adjacent ecosystems when new habitats are created (Hobbs 2000). In the case of the Azorean arthropods, the SIE, AZE and NAT species represent the original fauna of these islands, thus the creation of new, anthropogenic habitats cannot add new species to the pools (except exotic species, which were not considered in our study). However, anthropogenic alterations can alter the proportion with which the various species are present in a community, for example by increasing the abundance of certain epigean species that are able to survive outside native forests. As most of the AZEs persisting in non-native habitats are the most abundant ones in the native forests, summing all the habitats will mechanically increase the abundance of those species making the slope of the rank-abundance curve steeper.

Overall, we hypothesize that NAT species, being by definition more widely distributed than endemics, are those that tend to be also ecologically more varied, and therefore with higher variability in their ecological tolerance and abundances. By contrast, SIEs are a more idiosyncratic assortment of species, because of their always non-overlapping island distributions. In our samples, we found that some SIE species were locally very abundant, thus increasing the slope of the regression line and depressing general diversity indices. This can also be interpreted in terms of inter-island species exchanges and possible source-sink dynamics. Inter-island species exchanges is null among single island endemics (because species do not occur in more than one island) and could be maximum in NAT for the ones that are distributed in most islands.

These interpretations are supported by higher values of the Gambin α parameter found in the SIEs and AZEs in all cases (Fig.1). According to Ugland et al. (2007), the Gambin α parameter provides a measurement of the complexity or dimensionality of the 'bundles' of processes structuring a community, and hence of the relative ability of the species making up the sample to 'reach' into the axes of the niche-space hypervolume characterizing the sampling space. Thus, communities dominated by rare species should have low α values (which is the case of the native species), whereas communities with relatively few rare species, and which are regulated by multiple processes (or that are responsive to multiple gradients) should have higher α values (which is the case with SIEs and AZEs). A high α value means an approximation to a log-normal distribution with few rare species, many species of intermediate abundance and some additional abundant species (see also Matthews et al. 2014b).

The patterns found in this study may be greatly exaggerated by the loss of the least abundant SIEs and the consequent lack of a "right-tail" in the SAD curves due to both natural and human induced factors. SIEs have low persistence probabilities for long periods of time if they are rare in the single island they occupy. No rescue effects are possible and they are eventually driven to extinction both naturally, because of small population fluctuations, and as a result of human disturbance. Thus, very rare SIEs may have disappeared early after human island colonization, as is the case of several specialized forest species never collected in the Azores after their original description early in 20th century (Borges et al. 2000, Terzopoulou et al. 2015) or the unknown epigean relatives of the SIE cave adapted *Trechus* (Borges et al. 2007). This is also confirmed by the past and future extinctions of endemic species in the Azores recently hypothesized for a number of arthropod taxa (Cardoso et al. 2010, Triantis et al. 2010, Terzopoulou et al. 2015). This is especially critical for native forest species, as seen in our results (Fig. 1). In contrast, AZEs and especially NATs may be able to persist in very low abundances for long periods of time due to rescue effects by emigrants coming from different habitats (Borges et al., 2008), islands or even outside the archipelago.

Conspecific spatial aggregation is an important factor in SADs (Storch et al. 2008), and, in relation to our study, it is clear that SIEs are a case of extreme aggregation, because the rarest species are spatially aggregated (see Borges et al. 2006, Gaston et al. 2006). Dispersal ability is also known to determine the shape of SADs (see Borda-de-Água et al. 2007). Thus, as it is the case of many native species, SADs of species with high dispersal ability tend to follow approximately a log-normal distribution for the entire range of areas in the dataset (Borda-de-Água et al., 2007). Here we also show that when a regional pool of species is decomposed into different colonization assemblages, different patterns arise in the SADs. The fact that different assemblages present different SADs is not new. Dividing species into guilds also allows the possibility of distinguishing contrasting patterns. Marquet et al. (2004) suggest that at some scales SADs have several local maxima, which could be explained by mixing different guilds (see also Matthews et al., 2014a). For instance, Borges et al. (2008) showed for the Azorean arthropods that four functional groups (suckers, chewers, spiders and other predatory arthropods) differ substantially in their SADs when the community was decomposed into three abundance groups (rare, intermediate abundance, common).

5. Conclusions

In community ecology, the study of SADs is considered critical for the understanding of species community structure and has great potential for comparison among completely different communities (McGill et al. 2007). Using data on a biogeographical scale with three disparate groups of species defined by colonization stories (i.e. single island endemics, archipelagic endemics and native non-endemics), we have found consistent differences in the shapes and parameters of SADs models.

In both ecological and biogeographical studies, pooling all species can obscure important patterns (see, for example, Borges et al. 2008). Island communities are often comprised of biogeographically and ecologically distinctive groups of species with potential differential relative accumulation of individuals in space and time, for example, species with different colonization abilities / distribution ranges. Other examples of biogeographical groups not investigated in the current study are exotic and invasive species. We suggest that when standardized abundance is available, abundance profiles (e.g. SAD analyses) be more widely used in biogeographical studies, and that existing island data be meta-analysed to understand differences in SADs of species with different colonization stories.

Although our study may be illustrative of general patterns, there is a major need for more studies combining standardized measures of species abundance in islands across different taxa and archipelagos. These analyses will help researchers to formulate hypotheses concerning relative abundance of species, explain regional differences in the abundance profile of different groups of species, and contribute to the development of biogeographical theory more generally.

Acknowledgments

We are grateful to L. Borda-de-Água and Tom Matthews for useful discussions during the development of this research. D. Alonso provided many useful comments on a first version of the manuscript. This study was partly supported by the project FCT-PTDC/BIA-BIC/119255/2010 – “Biodiversity on oceanic islands: towards a unified theory”.

References

Abdel–Monem, A.A. et al. 1975. K/Ar ages from the eastern Azores group (Santa Maria, São Miguel and the Formigas Islands). Lithos 4: 247–254.

Beck, J. et. al. 2012. What´s on the horizon for macroecology? Ecography 35: 673–683.

Borda–de–Água, L. et al. 2007. Scaling biodiversity under neutrality. In Storch, D. et al. (eds), Scaling Biodiversity. Cambridge Univ Press, pp. 347–375.

Borges, P.A.V. et al. 2005. Ranking protected areas in the Azores using standardized sampling of soil epigean arthropods. Biodiver Conserv 14: 2029–2060.

Borges, P.A.V. et al. 2006. Invasibility and species richness of island endemic arthropods: a general model of endemic vs. exotic species. J. Biogeogr 33: 169–187.

Borges, P.A.V. et al. 2007. Biodiversity patterns of cavernicolous ground–beetles and their conservation status in the Azores, with the description of a new species: *Trechus isabelae* n. sp. (Coleoptera, Carabidae, Trechinae). Zootaxa 1478: 21–31

Borges P.A.V. et al. 2000. Ranking the Azorean Natural Forest Reserves for conservation using their endemic arthropods. J Insect Conserv 4: 129–147

Borges, P.A.V. et al. 2008. Insect and spider rarity in an oceanic island (Terceira, Azores): true rare and pseudo–rare species. In Fattorini, S. (ed), Insect Ecology and Conservation, Research Signpost, pp. 47–70.

Borges, P.A.V. and Wunderlich, J. 2008. Spider biodiversity patterns and their conservation in the Azorean archipelago, with description of new taxa. – Syst Biodivers 6: 249–282.

Büssenschütt, M. and Pahl–Wostl, C. 1999. Diversity patterns in climax communities. Oikos 87: 531–540.

Cardoso, P. et al. 2009. A spatial scale assessment of habitat effects on arthropod communities of an oceanic island. Acta Oecol 35: 590–597.

Cardoso, P. et al. 2010. Drivers of diversity in Macaronesian spiders and the role of species extinctions. J Biogeogr 37: 1034–1046.

Cardoso, P. et al. 2014. Partitioning taxon, phylogenetic and functional beta diversity into replacement and richness difference components. J Biogeogr 41: 749–761.

Cox, C. B. and Moore, P. D. 1999. Biogeography: an ecological and evolutionary approach, 6th edn. Blackwell Science Ltd, Oxford.

Demand, J. et al. 1982. Prospection géothermique, íles de Faial et de Pico (Açores). Rapport géologique, geochimique et gravimétrique. Technical report, BRGM 82 SGN 003 GTH

Dornelas, M. 2010. Disturbance and change in biodiversity. Phil Trans R Soc B 365: 3719–3727.

Fattorini, S. 2005. A simple method to fit geometric series and broken stick models in community ecology and island biogeography. Acta Oecol 28: 199–205.

Florencio, M. et al. 2013. Arthropod assemblage homogenisation in oceanic islands: the role of exotic and indigenous species under landscape disturbance. Divers Distrib 19: 1450–1460.

Fraterrigo, J.M. and Rusak, J. A. 2008. Disturbance–driven changes in the variability of ecological patterns and processes. Ecol Lett 11: 756–770.

Freckleton, R. P. et al. 2005. Large□scale population dynamics, abundance–occupancy relationships and the scaling from local to regional population size. J Anim Ecol 74: 353–364.

Gaspar, C. et al. 2008. Diversity and distribution of arthropods in native forests of the Azores archipelago. Arquipélago Life and Marine Sciences 25: 1–30.

Gaston, K.J. and Blackburn, T.M. 1996. Range size–body size relationships, evidence of scale dependence. Oikos 75: 479–485.

Gaston, K.J. and Lawton, J.H. 1990. Effects of scale and habitat on the relationship between regional distribution and local abundance. Oikos 58: 329–335.

Gaston, K.J. et al. 2000. Abundance–occupancy relationships. J Appl Ecol 37 (Suppl. 1): 39–59.

Gaston, K.J. et al. 2006. Abundance, spatial variance and occupancy: arthropod species distribution in the Azores. J Anim Ecol 75: 646–656

Giller, P. S. 1984. Community structure and the niche. Chapman and Hall.

Gravel, D. et al. 2006. Reconciling niche and neutrality: The continuum hypothesis. Ecol Lett 9: 399–409.

Hammer, Ø. et al. 2001. PAST – PAlaeontological STatistics, ver. 1.89. Palaeontol Electron 4(1): 1–9.

Hanski, I. 1994. A practical model of metapopulation dynamics. J Anim Ecol 63: 151–162.

Hayek, L. C. and Buzas, M. A. 2010. Surveying natural populations. Quantitative tools for assessing biodiversity. Columbia Univ Press.

He, F. and Tang, D. 2008. Estimating the niche preemption parameter of the geometric series. Acta Oecol 33: 105–107.

Higgins, C. L. and Strauss, R. E. 2008. Modeling stream fish assemblages with niche apportionment models: patterns. processes, and scale dependence. T Am Fish Soc 137: 696–706.

Hobbs, R. J. 2000. Land–use changes and invasions. In Mooney, H.A. and Hobbs, R. J. (eds), Invasive species in a changing world. Island Press, pp. 55–64.

Holloway, J.D. 1996. The Lepidoptera of Norfolk Island, actual and potential, their origins and dynamics. In Keast, J.A. and Miller, SE. (eds) The origin and evolution of Pacific Island biotas, New Guinea to Eastern Polynesia. Patterns and processes, SPB Academic Publishing, pp. 123–151.

Hothorn, T. et al. 2008. Simultaneous inference in general parametric models. Biom J 50: 346–363.

Hubbell, S.P. 2001. The unified neutral theory of biodiversity and biogeography. Princeton Univ Press.

Laffan, S.W. and Crisp, M.D. 2003. Assessing endemism at multiple spatial scales, with an example from the Australian vascular flora. J Biogeog 30: 511–520.

Lande, R. et al. 2003. Stochastic population dynamics in ecology and conservation. Oxford Univ Press.

Lawton, J.H. 1996. Patterns in ecology. Oikos 75: 145–147.

Lomolino et al. 2010. Biogeography, 4th ed. Sinauer Associates.

Magurran, A.E. 1988. Ecological diversity and its measurement. Princeton Univ.

Magurran, A.E. 2003. Measuring biological diversity. Blackwell Publishing.

Magurran, A. E. and Henderson, P. A. 2003. Explaining the excess of rare species in natural species abundance distributions. Nature 422: 714–716.

Marquet, P.A. et al. 2004. Diversity emerging: toward a deconstruction of biodiversity patterns. In Lomolino, M. and Heaney L.R. (eds), Frontiers of biogeography: new directions in the geography of nature. Sinauer Associates, pp. 191–209.

Martins, A.M.F. 1993. The Azores – Westernmost Europe: Where evolution can be caught red–handed. Bol Mus Mun Funchal, Sup. 2: 181–198.

Matthews, T.J and Whittaker, R. J. 2014a. Fitting and comparing competing models of the species abundance distribution: assessment and prospect. Front Biogeogr 6(2): fb_20607. Retrieved from: http://escholarship.org/uc/item/3gz504j3

Matthews, T. J. and Whittaker, R.J. 2014b. Neutral theory and the species abundance distribution: recent developments and prospects for unifying niche and neutral perspectives. Ecol Evol 4(11): 2263–2277.

Matthews, T. J. et al. 2014a. Multimodal species abundance distributions: a deconstruction approach reveals the processes behind the pattern. Oikos, 123: 533–544.

Matthews, T. J. et al. 2014b. The gambin model provides a superior fit to species abundance distributions with a single free parameter: evidence, implementation and interpretation. Ecography 37: 1002–1011.

May. R. M. 1975. Patterns of species abundances and diversity. In Cody M.L. and Diamond J.M. (eds), Ecology and evolution of communities, Harward Univ Press, pp. 81–118.

May, R.M. 1994. The effects of spatial scale on ecological questions and answers. In Edwards, P.J. et al. (es), Large–scale ecology and conservation biology. Blackwell Scientific Publications, pp. 1–17.

McGill, B. J. et al. 2007. Species Abundance Distributions: Moving beyond single prediction theories to integration within an ecological framework. Ecol Lett,10: 995–1015.

McNally, R. 2007. Use of the abundance spectrum and relative–abundance distributions to analyze assemblage change in massively altered landscapes. Am Nat 170: 319–330.

Meijer, S.S. et al. 2011. The effects of land–use change on arthropod richness and abundance on Santa Maria Island (Azores): unmanaged plantations favour endemic beetles. J Insect Conserv 15: 505–522.

Passy, S.I. 2012. A hierarchical theory of macroecology. Ecol Lett 15: 923–934.

Prado, P.I. and Miranda, M.D. 2014. sads: Maximum Likelihood Models for Species Abundance Distributions. R package version 0.1.10. http://CRAN.R–project.org/package=sads

Preston, F.W. 1948. The commonness, and rarity, of species. Ecology 29: 254–283.

Preston, F.W. 1962. The canonical distribution of commonness and rarity: Part I. Ecology 43: 185–215.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R–project.org/.

Ribeiro, S.P. et al. 2005. Canopy insect herbivores in the Azorean Laurisilva forests: key host plant species in a highly generalist insect community. Ecography 28: 315–330.

Sutherland, C. et al. 2012. Multi–scale processes in metapopulations: contributions of stage structure, rescue effect, and correlated extinctions. Ecology 93: 2465–2473.

Solé, R.V. et al. 2004. Habitat fragmentation and biodiversity collapse under recruitment limitation. Ecol Complex 1: 65–75.

Starzomski, B.M. et al. 2008. On the relationship between regional and local species richness a test of saturation theory. Ecology 89: 1921–1930.

Storch, D. et al. 2008. The quest for macroecological patterns: geometry of species distributions at multiple spatial scales. Ecol Lett 11: 771–784.

Taylor, L.R. 1961. Aggregation, variance and the mean. Nature 189: 732–735.

Terzopoulou, S. et al. 2015. Drivers of extinction: the case of Azorean beetles. Biol Lett 11: 1–4.

Tokeshi, M. 1993. Species abundance patterns and community structure. Adv Ecol Res 24: 112–186.

Triantis, K.A. et al. 2010. Extinction debt on oceanic islands. Ecography 33: 285–294.

Turquin, M.J. 1973 Une biocénose cavernicole originale pour le Bugey: le puits de Rappe. C. R. 96^{e} Congr. Nat. Soc. Savantes, Toulouse 1971, Sciences 3: 235–256.

Ugland, K.I. et al. 2007. Modelling dimensionality in species abundance distributions: description and evaluation of the Gambin model. Evol Ecol Res 9: 1–12.

Ulrich, W. et al. 2010. A meta–analysis of species–abundance distributions. Oikos 119: 1149–1155.

Vergnon, R. et al. 2009. Niches versus neutrality: uncovering the drivers of diversity in a species–rich community. Ecol Lett 12: 1079–1090.

Wilson, J.B. 1993. Would we recognise a broken–stick community if we found one? Oikos 67: 181–183.

Supplementary material (Appendix oik.XXXXX at <www.oikosjournal.org/readers/appendix>).

Appendix 1

Tables

Table 1. Geometric series models for the species abundance rank distribution of the arthropods of the Azorean Islands calculated for single island endemics (SIE), Azorean endemics (ARE) and native (but non endemic) species (NAT). Calculations were done using samples from both only native forests (epigean soil and canopy arthropods) and all habitats (epigean soil arthropods). All regressions are significant (P<0.001)

Sample	Intercept (±Standard Error)	Slope (±Standard Error)	R^2
Native forests - Canopy			
SIE	3.771 (0.159)	-0.266 (0.019)	0.938
AZE	3.740 (0.080)	-0.106 (0.003)	0.957
NAT	3.663 (0.032)	-0.113 (0.001)	0.992
Native forests - Soil			
SIE	2.543 (0.055)	-0.108 (0.004)	0.967
AZE	2.746 (0.042)	-0.119 (0.003)	0.984
NAT	3.218 (0.045)	-0.111 (0.002)	0.982
All habitats - Soil			
SIE	2.620 (0.045)	-0.106 (0.003)	0.979
AZE	3.149 (0.068)	-0.093 (0.003)	0.955
NAT	3.467 (0.045)	-0.061 (0.001)	0.962

Table 2. Results of the Gambin models for the species abundance rank distribution of the arthropods of the Azorean Islands calculated for single island endemics (SIE), Azorean endemics (ARE) and native (but non endemic) species (NAT). Calculations were done using samples from both only native forests (epigean soil and canopy arthropods) and all habitats (epigean soil arthropods). The variable α determining the shape of the gamma distribution and the *P* of the χ^2 test are given. Additionally, the mean α and its associated 95% confidence interval from the standardization procedure are given.

Sample	α	χ^2 *P*	Standardization (1000 runs) set up to 1000 individuals	
Native forests - Canopy			Mean α	95%
SIE	5.466	0.542	8.004	[5.191; 11.291]
AZE	1.731	0.17	3.346	[2.196; 5.128]
NAT	1.48	0.693	1.165	[0.836; 1.736]
Native forests - Soil				
SIE	3.662	0.52	3.992	[2.911; 5.368]
AZE	1.925	0.941	1.603	[1.014; 2.208]
NAT	1.133	0.677	1.521	[0.931; 2.471]
All habitats -Soil				
SIE	3.663	0.507	3.857	[2.967; 4.570]
AZE	0.924	0.865	0.929	[0.552; 1.493]
NAT	1.622	0.976	1.449	[0.957; 2.154]

Table 3. Diversity and evenness of arthropod species of the Azorean islands calculated for single island endemics (SIE), Azorean endemics (AZE) and native (but non endemic) species (NAT). Calculations were done using standardized samples from both only native forests (epigean soil and canopy arthropods) and all habitats (epigean soil arthropods). 95%CI are given in parentheses.

Sample	Diversity	95%	Evenness	95%
Native forests - Canopy				
SIE	1.926	[1.904; 1.944]	0.490	[0.482; 0.568]
AZE	2.064	[2.044; 2.082]	0.171	[0.171; 0.205]
NAT	2.432	[2.412; 2.448]	0.178	[0.178; 0.204]
Native forests - Soil				
SIE	2.413	[2.343; 2.460]	0.372	[0.358; 0.427]
AZE	2.162	[2.103; 2.205]	0.256	[0.249; 0.296]
NAT	2.160	[2.129; 2.184]	0.158	[0.160; 0.192]
All habitats -Soil				
SIE	1.948	[1.884; 1.995]	0.351	[0.332; 0.384]
AZE	1.136	[1.092; 1.172]	0.107	[0.107; 0.140]
NAT	2.830	[2.795; 2.853]	0.199	[0.201; 0.228]

Figure 1. Top panels (first row). Distribution of the 1000 Gambin alpha values generated by the standardization procedures for Canopy samples in native of forest, soil samples in native forests and soil samples in all habitats. Greens indicated SIE, blue AZE and orange NAT. Bottom panels: Abundance rank curves for the arthropod species of the Azorean islands calculated for single island endemics (SIE), Azorean endemics (AZE) and native (but non endemic, NAT) soil epigean and canopy species collected in native forests and soil epigean species in all habitats. AZE and NAT datasets have been randomized 1000 times to the same number of individuals as SIE. Averages and 95% confidence limits envelopes are presented.