

# Multimodal species abundance distributions: a deconstruction approach reveals the processes behind the pattern

Thomas J. Matthews, Paulo A. V. Borges and Robert J. Whittaker

*T. J. Matthews (thomas.matthews@ouce.ox.ac.uk), and R. J. Whittaker, Conservation Biogeography and Macroecology Programme, School of Geography and the Environment, Univ. of Oxford, South Parks Road, Oxford, OX1 3QY, UK. – TJM and P. A. V. Borges, Azorean Biodiversity Group (ABG, CITA-A) and Portuguese Platform for Enhancing Ecological Research and Sustainability (PEERS), Depto de Ciências Agrárias, Univ. of the Azores, Rua Capitão João d'Ávila, Pico da Urze, PT-9700-042, Angra do Heroísmo, Portugal. RJW also at: Center for Macroecology, Evolution and Climate, Dept of Biology, Univ. of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark.*

It is becoming increasingly apparent that the species abundance distribution of many ecological communities contains multiple modes, a phenomenon that has been largely overlooked. Here, we test for multiple modes in the species abundance distribution using a combination of one, two and three mode Poisson lognormal distributions and an extensive arthropod dataset from the Azores. We consider the abundance distribution of twelve native laurisilva forest fragments and the combination of fragments within five islands, allowing us to detect whether patterns are consistent across scales. An information theoretic approach is employed to determine the best model in each case. To explore the processes driving multimodal abundance distributions we tested various potential mechanisms. We classified species as core if they are present in over half the fragments in our study, and as satellite if they are sampled in fewer than half the fragments. Furthermore, species are classified based on body size, whether they are indigenous (i.e. endemic or native non-endemic) or introduced to the Azores, abundance in land uses other than native forest, and dispersal ability. We find that models incorporating multiple modes perform best for most fragments and islands. A large number of communities are bimodal, comprising a mode of very rare species and a mode of relatively common species. Deconstructing the full assemblages into their constituent subsets reveals that the combination of ecologically different groups of species into a single sample underpins the multimodal pattern. Specifically, the rarer mode prevalently contains a higher proportion of satellite taxa, introduced species and species that are more adapted to anthropogenic land uses that surround the native forest.

The observation that ecological systems contain a small number of very abundant species and numerous relatively rare species is often described as one of ecology's few universal laws, with very few known deviations (Fisher et al. 1943, Preston 1948, Pielou 1969, McGill et al. 2007, Ulrich et al. 2010). A species abundance distribution (SAD) describes the abundance of all species recorded within an ecological community of interest (Ulrich et al. 2010). Over 25 different functions have been proposed to model SADs (McGill et al. 2007, McGill 2011); the two most widely used being the logseries (Fisher et al. 1943) and the lognormal (Preston 1948). As the continuous lognormal distribution allows fractional abundances, the Poisson lognormal (Bulmer 1974) is often used instead of the standard continuous lognormal. Several authors (McGill et al. 2007) have called for a shift away from simply determining how well particular distributions fit different datasets, to a more process and mechanistic oriented approach. One particular emerging area of interest in this regard has been the increased recognition of the possibility of multiple modes within the SAD.

While the possibility of multimodal SADs has long been discussed (Pielou 1969, Ugland and Gray 1982, Gray et al. 2005, Borges et al. 2008, Borda-de-Agua et al. 2012) it has largely been overlooked and few studies explicitly test for multiple modes when fitting SAD models (but see Dornelas and Connolly 2008, Vergnon et al. 2012). Even when multiple modes have been observed, the pattern has often been dismissed as an uncommon anomaly (Gaston 1994). Pielou (1969) was one of the earliest to acknowledge the possibility of multimodal SADs, postulating that increasing a sample by broadening the type of species included, for example from a focus on warblers to all bird species, could result in the combination of multiple logseries distributions, each with its own set of parameters. Ugland and Gray (1982) also debated multimodal distributions, this time in the context of communities disturbed away from equilibrium. They argued that a realistic community model should take account of the asymmetry in abundances typical of ecological communities, illustrating this idea by subdividing data for a community into three distinct abundance groups, roughly translated as rare, intermediately abundant

and common (Ugland and Gray 1982, see an application in Borges et al. 2008). Within each group the SAD is symmetric, which leads to a mixture of three lognormal distributions; the combination of which results in an asymmetric SAD for the whole assemblage. More recently, Gray et al. (2005, 2006a) found strong evidence of bimodal lognormal distributions for a mixture of marine and terrestrial datasets.

Dornelas and Connolly (2008) advanced this approach to focus on multimodal Poisson lognormal (herein 'PLN') distributions. They fitted a combination of one to four mode PLN distributions in addition to the logseries for a large sample of Australian coral communities and compared the various distributions using maximum likelihood methods. At the scale of the entire sample the three-mode PLN was found to have the greatest support (Dornelas and Connolly 2008). Thus, it appears likely that multimodal SADs are common in nature; it is simply that they have been overlooked. Vergnon et al. (2012) provided further evidence supporting this conjecture. Sourcing a number of previously published datasets from the literature relating to an array of taxa, Vergnon et al. (2012) confirmed that multimodality is not simply confined to a few anomalous systems.

Numerous potential mechanisms driving the aforementioned multimodality have been proposed. The combination of multiple samples through space and/or time has been shown to result in the amalgamation (the 'amalgamation hypothesis') of two different groups of species (Hanski 1982, Magurran and Henderson 2003, Ulrich and Zalewski 2006, Unterseher et al. 2011). Firstly, there is a group of core species: those which are widely distributed and present/abundant within numerous samples (Ulrich and Zalewski 2006). The second group, often termed satellite or occasional species (herein satellite), are only recorded in a few samples and are generally viewed as representing migrants from the surrounding species pool (Unterseher et al. 2011). The two sets of species leave different signals in the SAD and when grouped together can result in the multimodal SADs which have been observed (Gray et al. 2005). In addition to the core/satellite dichotomy, it is possible that the grouping of other distinct species subsets into one sample may lead to multiple modes. For instance, dispersal ability has been postulated as an important determinant of a species' abundance within a community (Borda-de-Água et al. 2007, Alonso et al. 2008). It has also been suggested that introduced species may influence the SAD of a given community (Labra et al. 2005): a process which is likely to be prevalent on many islands, as such systems have received a disproportionate amount of species introductions (Whittaker and Fernández-Palacios 2007). Moreover, it is likely that these various subset divisions are not entirely independent; for example, it is possible that in general satellite species include many better dispersers than the core group and are present in the native forest due to mass effects (Shmida and Wilson 1985). Deconstructing entire assemblages into these different subsets and analysing each subset for particular patterns individually has been advocated as an effective method for elucidating new perspectives on macroecological and biogeographical patterns (Marquet et al. 2004).

Emergent neutrality theory (Scheffer and van Nes 2006, Segura et al. 2011, Vergnon et al. 2012; herein 'ENT') has

also been put forward as a potential explanation for multimodal SADs (Vergnon et al. 2012). ENT contends that species are organised along a niche axis such that multiple modes of coexisting species are observable. The species within the 'core' of these niche axis modes are relatively abundant while the species in the 'valleys' are relatively rare, which results in a bimodal SAD (Vergnon et al. 2012). Multimodal SADs are also observed if the species within the different niche axis modes differ significantly in terms of their abundance. Finally, neutral theory proper has also been shown to result in multimodal SADs (Borda-de-Água et al. 2007), but testing this proposition is beyond the scope of the present study.

In this paper, we test for multimodal PLN SADs using an extensive arthropod dataset from the Azores comprising 12 fragments of native laurisilva forest, across five islands. Analyses are performed at both the fragment and island scale to determine whether the presence of multiple modes is dependent on the scale of amalgamation of samples. We will show that, while not a universal pattern, multiple modes are present in the SADs of numerous samples. In order to infer the processes driving the observed multimodal SADs, we test various potential explanatory variables. The full assemblage is deconstructed into various subsets, including core and satellite groups, and the proportions of these different subsets within the different modes are examined. We also test the possibility that self-limiting similarity along a body size niche axis appropriately characterises Azorean laurisilva forest arthropod assemblages.

## Methods

### Study site and sampling methods

The data used in the current manuscript are part of a long-term ecological study conducted in the Azores archipelago, the BALA project (1999–2004; Biodiversity of Arthropods from the Laurisilva of the Azores; Borges et al. 2005, Ribeiro et al. 2005). As part of this study a total of 100 transects (150 × 5 m) were randomly placed within 18 fragments of protected native forest in seven Azorean islands. In each transect, arthropods from the soil (mainly epigeal) and herbaceous vegetation were surveyed with a set of 30 pitfall traps (Borges et al. 2005), while arthropods associated with dominant woody plant species were sampled using a beating tray, focused on three tree species (10 samples per tree; for details see Ribeiro et al. 2005). In total, 6770 samples (3420 pitfall traps and 3350 beating samples) were collected. The taxonomic groups were selected to cover the majority of arthropod diversity, excluding only those groups for which taxonomic expertise was not available. All Araneae, Opilionida, Pseudoscorpionida, Myriapoda and Insecta (excluding Diptera and Hymenoptera) were assigned to species or to morphospecies through comparison with a reference collection. Most taxa were assigned additional classifications based on colonization status (termed here 'indigenous' or 'introduced') and ecological information (predator, herbivores, saprophagous). We classified as introduced those species known from regions outside the Macaronesian archipelagos and that are believed to have

been introduced after human settlement in the Azores in the 15th century. This was generally inferred either from historical records of detected species introductions, or from the current distribution of a species being closely associated with human activity. Otherwise, we assumed the species to be indigenous (Borges et al. 2010). For roughly 16% of the morphospecies only the family or genus is known. We classified such morphospecies as indigenous or introduced based on the dominant category (i.e. introduced or indigenous) of species in the family or genus (Borges et al. 2010). The majority of species (84% of the overall species richness) belong to four taxonomic orders (Araneae, Hemiptera, Coleoptera and Lepidoptera; more details in Gaspar et al. 2008).

As fitting the various multimodal SAD models was very time intensive a subset of 12 fragments and five islands were selected from the BALA dataset for analysis (Fig. 1). For each fragment the individual transects were grouped together to form the dataset for that particular fragment. The minimum number of transects sampled within the fragments is four, and thus in order to maintain consistency across the samples four transects were pooled in all fragments. For fragments in which more than four transects were sampled, four were selected at random. To create the island datasets the same transects that were combined to form the individual fragment samples were pooled for all fragments on a particular island. A detailed description of the sampling methods and study sites can be found in Gaspar et al. (2008). Hereafter ‘sample’ refers to the unit of analysis for which a SAD was constructed, i.e. fragment or island. Thus, a reference to ‘all samples’ relates to all 12 fragment and five island datasets.

### Species classification

In order to facilitate greater understanding of the mechanisms behind multimodal SADs, the entire assemblage was deconstructed into distinct subsets. Species were classified as possessing either high or low dispersal ability by a trained taxonomist (PAVB). Following Magurran and Henderson (2003; see also Hanski 1982), species were further classified into core and satellite subgroups. However, this was achieved through focus on the pooling of species through space as opposed to time (Gray et al. 2005, Ulrich and Zalewski 2006). A species was classified as a core species if it was present in more than 50% of the fragments in our study (cf. Ulrich and Zalewski 2006), or a satellite species if it was present in 50% or less. However, as Ulrich and Olrik (2004) note, it is likely that there is a ‘continuum’ between core and satellite species and any dichotomy is likely to be a simplification of the ecological reality. Therefore, by way of a sensitivity analysis, we also used a different core/satellite classification method to determine whether our choice of 50% affected results. For this purpose we split species into three groups: those present in 25% or fewer fragments, 26–75%, and 76–100% of fragments.

As part of a separate project three other land-uses in addition to native forest were sampled (Cardoso et al. 2009, 2013): exotic plantation forest, semi-natural pasture, and intensive pasture. This allowed a further classification of species based on the abundance of species in these other land uses. As such, species were classified as tourist species (Borges et al. 2008) if they were recorded as having a greater abundance in a land-use other than native forest. Thus, tourist species are either those species that are more adapted

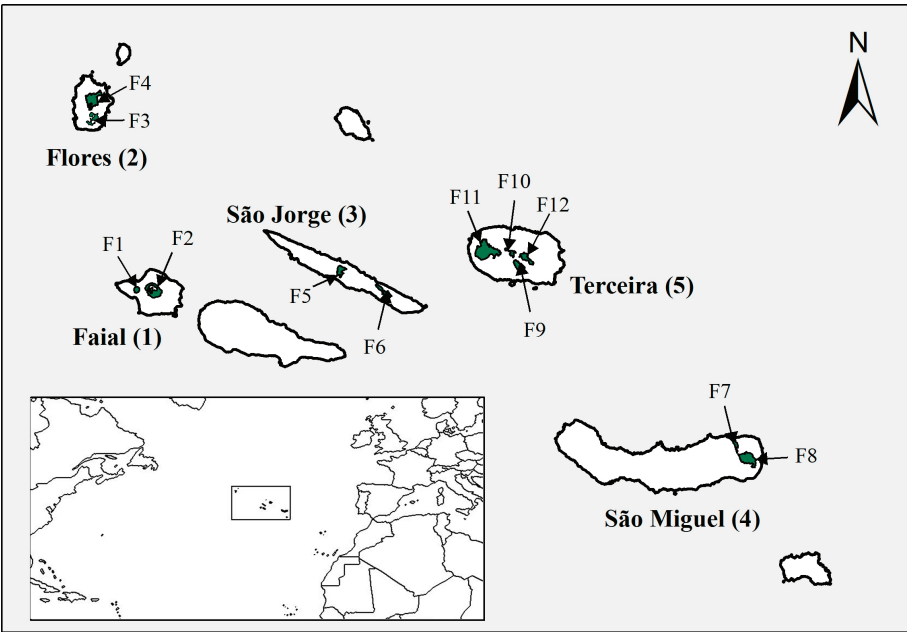


Figure 1. Map of study sites in the Azores. The names of the five islands are given and the numbers in parentheses refer to the corresponding island information in Table 2. Within these five islands the locations of the 12 native forest fragments are presented alongside the fragment codes which correspond to the individual fragment information in Table 1. Due to issues with clarity the figure is not to scale; in particular, the native forest fragments have been enlarged to allow them to appear visible.

to anthropogenic land uses and are present in the native forest due to 'mass effect' like processes, or species which are habitat generalists and are found in large numbers in all four land uses. In order to keep sample size constant across land uses, transects were selected at random, and pooled, for each of the three additional land-uses based on the number of transects used to construct individual fragments. The body size of each species was also measured: 1) for spiders, body length was collated from the literature, separately for males and females to address the possible effects of sexual dimorphism. As female and male body lengths were highly correlated (Pearson correlation = 0.96,  $p < 0.001$ ), the average value between females and males was used; 2) for the other arthropods, we measured specimens available in our collection. Finally, species were classified as indigenous or introduced.

## Statistical analyses

### Model selection

To determine whether the SADs of the various samples exhibited multiple modes, a combination of one, two and three mode Poisson lognormal distributions (PLN, PLN2 and PLN3 respectively) were compared using a multi-model information theoretic framework (Burnham and Anderson 2002). The models were fit to each of the 12 fragments and five islands and model performance was calculated with the Bayesian information criterion (BIC) and Akaike's information criterion corrected for small sample size ( $AIC_c$ ). The BIC is a useful measure in studies of this nature as it penalises to a greater extent for additional parameters than does  $AIC_c$  (Burnham and Anderson 2004). As the PLN2 and PLN3 models have five and eight parameters respectively, the number of parameters is of particular importance in comparing model fits in our study. The  $AIC_c$  was preferred to the standard Akaike's information criterion (AIC) as it is more applicable when datasets are small, and because  $AIC_c$  converges to AIC when the sample size is large (Burnham and Anderson 2004). The best model(s) within each sample was selected by setting a minimum difference criterion in BIC and  $AIC_c$  of two (Burnham and Anderson 2002).

Due to the complexity involved in fitting the PLN2 and PLN3 distributions, and the increased chance of numerous local maxima, parameter searches were launched from multiple starting values, and the parameter values corresponding to the lowest negative log likelihood selected for use in the multi-model comparison. Initially, starting values were estimated from analysing the frequency distributions of the raw data. A more objective parameter search method was also conducted. This involved setting upper and lower limit boundaries for each parameter and then creating a loop that randomly selected values within these bounds for each parameter and then running the optimisation algorithm using these values, reporting the likelihood, parameter values and whether convergence was successful, after each iteration. Furthermore, a variety of maximum likelihood optimisation algorithms in R (i.e. 'optim', 'nlminb') were employed in the parameter searches.

The functions used to fit the PLN models were derived from Dornelas and Connolly (2008), particularly their supplementary material, and readers are directed there for

the probability distribution formulae and full R code. In order to determine whether the PLN was in fact a good choice of model to begin with, a logseries distribution was also fitted to each sample. A likelihood ratio test (LRT) based on a  $\chi^2$  distribution was also used following the multi-model comparison to determine if the increase in fit of more complex models was significant. Considering only nested models, a more complex model (i.e. more parameters) will always provide a better fit than a less complex model, and while  $AIC_c$  and BIC take the number of parameters into account, it can be useful to check the LRT statistic. The degrees of freedom for each model were the number of free parameters pertaining to each model. As LRT can only be used for nested models, the logseries was not included in this particular analysis. All analyses were carried out in the R statistical environment. In addition to the statistical approach, all samples were plotted on histograms using a  $\log_2$  binning method whereby bin 1 equals the number of singleton species, bin 2 equals the number of species with 2–3 individuals, bin 3 equals 4–7, and so on (method 3 in Gray et al. 2006b). Fits of the best distributions were analysed by eye in each case to check for any issues or peculiar fits.

### Post model selection

We tested various mechanisms that may result in multimodal SADs. To test the effect of combining species subsets on the SAD, for each of the samples in which the PLN2 or PLN3 distributions provided the best fit, the octaves corresponding to the modes of the distribution were recorded and the species within these octaves labelled accordingly. For all samples in which the PLN2 distribution provided the best fit, a Welch's two sample t-test (0.05 significance level) was used to determine if the percentage of core, tourist, indigenous and high dispersal ability species significantly varied between the first and second mode. Furthermore, to determine whether the proportion of particular subsets within each mode were significantly different from a random sample of the same amount of species from the regional pool, a null distribution was created. For each of the modes within the samples in which the PLN2 provided the best fit the number of species representing the modal octave class was recorded. This number of species was then randomly sampled without replacement from the regional pool and the proportion of core, tourist, indigenous, and high dispersal ability species calculated. The same method was used for mean body size. The number of iterations of this procedure was set at 10 000 for each mode. In order to make the regional pool sample more ecologically realistic, for any particular fragment the regional pool was restricted to data from the island in which the fragment was located. A 0.05 significance level was employed and thus an observed proportion falling in the top and bottom 2.5% of the sampled distribution was deemed to be significant. In addition to comparing the percentage of core and satellite species between modes within each sample, the PLN, PLN2 and logseries were calculated separately for the core subsets, and the PLN and logseries for the satellite subsets. It was then possible to determine for each sample if the core and satellite subsets were best modelled by separate distributions (sensu Magurran and Henderson 2003, Gray et al. 2005, Ulrich and Zalewski 2006). The PLN2 distribution was not calculated



for the satellite subsets as the reduction in sample size coupled with the large number of singletons made optimisation convergence problematic.

To determine if ENT provided an explanation for multimodality, we followed Vergnon et al. (2012) and used body size as the basis of the niche axis along which assemblages were theorised to be organised. Body size frequency distributions were created for each of the samples using both untransformed and log(base ten) transformed body size values and any modes identified through visual inspection. Additionally, for each sample, body size was plotted against species abundance (again, using untransformed and log transformed data) and all plots examined. Lastly, within samples that exhibited a significantly bimodal SAD, the mean body size of all species within the modal octaves was calculated and a Welch's two sample t-test used to assess whether the differences in body size between the two modes was significant across samples. Obviously, measures other than body size can be used to construct a niche axis, and equally factors other than ENT can lead to multimodal body size distributions (Holling 1992), but we use body size as it has been recommended as an apt way of testing ENT (Segura et al. 2011, Vergnon et al. 2012).

## Results

In total, 56 338 individuals representing 399 unique species were analysed. Sample size varied from 3107 to 7366 individuals (mean = 4879) at the fragment scale (Table 1), and from 8942 to 17 973 individuals (mean = 11 702) at the island scale (Table 2). Ecological information for all species used in the analyses is presented in Supplementary material Appendix 1 Table A1.

### Model comparison

The data showed significant evidence of multimodality in a number of samples. Six of the twelve fragments were best represented by the PLN2 distribution according to both BIC and AIC<sub>c</sub> (Table 3). Three were best represented by the

Table 1. Site information for the different native forest fragments from the Azores used in the analyses, including the area and the number of arthropod individuals (n) and species (S) sampled within the four transects used to create the sample in each instance. Arthropods were sampled using a standardised pitfall trap and canopy beating methodology between 1999 and 2004. 'Island' relates to island within the Azores in which the fragment is located.

Fragment	Area (ha)	Island	n	S
1. Cabeço do Fogo	36	Faial	5586	92
2. Caldeira do Faial	191	Faial	3356	80
3. Lagoas Funda e Rasa	240	Flores	5989	117
4. Morro Alto	1331	Flores	3107	89
5. Pico Pinheiro	74	São Jorge	7366	107
6. Topo	220	São Jorge	4939	102
7. Atalhada	10	São Miguel	3990	130
8. Pico da Vara	306	São Miguel	6204	118
9. Caldeira Guilherme Moniz	223	Terceira	6018	86
10. Pico Galhardo	38	Terceira	4608	111
11. Serra de Santa Bárbara	1347	Terceira	3131	89
12. Terra Brava	180	Terceira	4216	103

Table 2. Site information for the different Azorean islands used in the analyses. For each island the number of fragments that were combined to create the island dataset is provided along with the combined area, which is the summed area of the combined fragments. + This is not indicative of the actual number of native forest fragments that have been studied in the BALA project on a particular island, it is simply the number selected for this study. The number of arthropod species (S) and number of individuals (n) sampled in each island is also given.

Island	Number of fragments +	Combined area (ha)	n	S
1. Faial	2	227	8942	125
2. Flores	2	1571	9096	141
3. São Jorge	2	294	12 305	145
4. São Miguel	2	316	10 194	178
5. Terceira	4	1788	17 973	175

PLN3 distribution, again according to both model selection criteria. The remaining three fragments did not have a clear best model. For fragment 1 and 2 (both from Faial) the PLN2 distribution provided the best fit according to AIC<sub>c</sub>. However, according to BIC, the PLN2 and PLN, and PLN2 and logseries were both in the best model set (i.e. either the lowest BIC value or within a value of two of the lowest value) for fragment 1 and 2, respectively (Table 3). For fragment 3 (Flores island) the PLN provided the best fit according to BIC, but PLN, PLN2 and PLN3 were all within the best model set according to AIC<sub>c</sub> (Table 3). At the island scale, two island samples were best fit by PLN2 (São Miguel, Terceira), and one by PLN3 (Flores), according to both criteria. The remaining two islands (Faial and São Jorge) had different best models if one used BIC or AIC<sub>c</sub> (Table 4). The LRT results are largely concurrent with model selection based on BIC and AIC<sub>c</sub> (All LRT results are presented in Supplementary material Appendix 2 Table A2), indicating that the better fit of PLN2 and PLN3 compared to PLN is significant despite the increase in number of parameters. The difference in goodness of fit between PLN2 and PLN3 was often not significant (Supplementary material Appendix 2 Table A2), suggesting that there is often no significant improvement in fit going from a bimodal to a trimodal model, although in all samples in which the PLN3 distribution provided the best fit according to the information criterion, the LRT results confirmed this.

For the samples in which PLN2 provided the best fit, one mode was apparent at the rare end of the distribution with the second modal octave generally being 16–32 or 33–64 individuals (see Fig. 2a for an example). Datasets best fit by PLN3 followed this pattern but with a third modal class at the more abundant end of the distribution (Fig. 2b). Nonetheless, visual inspection of the SAD histograms for certain samples indicated that the multimodal pattern was not always obvious despite the best model being either PLN2 or PLN3 (see Fig. 2c for an example). In sum, multimodality within the SAD, while not a universal pattern, does appear to be a common occurrence within natural Laurisilva forest arthropod communities in the Azores.

### Post model selection

The percentage of core, tourist, and indigenous species was found to differ significantly between mode one (rarer species)

Table 3. Model selection results for arthropod SADs of twelve fragments of native Laurisilva forest in the Azores. Arthropods were sampled using a standardised pitfall trap and canopy beating transect methodology between 1999 and 2004. Each fragment consists of four transects. For each fragment the maximum log likelihood (MLL) is provided for each of the four SAD models: the standard Poisson lognormal (PLN), the two (PLN2) and three (PLN3) mode Poisson lognormal, and the logseries. The Bayesian information criterion (BIC) and Akaike's information criterion corrected for small sample size (AIC<sub>c</sub>) calculated using the MLL are also presented. The PLN distribution has two parameters, PLN2 has five, PLN3 has eight, and the logseries one parameter. Fragment codes and abbreviated island name/initials relate to those in Table 1. Values in bold represent the best model for the corresponding fragment, for both BIC and AIC<sub>c</sub>. Where several models are within an AIC<sub>c</sub> or BIC value of two of one another, each is highlighted in bold.

Fragment	SAD Model	MLL	BIC	AIC <sub>c</sub>
1 FAI	PLN	366.90	<b>747.36</b>	740.06
	PLN2	359.40	<b>745.92</b>	<b>731.78</b>
	PLN3	356.91	754.51	734.01
	Logseries	370.75	750.54	745.63
2 FAI	PLN	321.23	656.02	648.73
	PLN2	312.77	<b>651.84</b>	<b>638.70</b>
	PLN3	311.53	662.49	643.63
	Logseries	322.13	<b>653.01</b>	648.41
3 FLO	PLN	447.17	<b>908.65</b>	<b>900.54</b>
	PLN2	444.09	916.75	<b>900.94</b>
	PLN3	440.45	923.75	<b>900.57</b>
	Logseries	451.83	913.18	907.76
4 FLO	PLN	331.02	676.32	668.25
	PLN2	328.17	683.27	669.37
	PLN3	315.93	<b>671.19</b>	<b>652.48</b>
	Logseries	332.94	674.86	670.02
5 SJG	PLN	422.58	858.31	851.48
	PLN2	413.82	<b>855.67</b>	<b>840.48</b>
	PLN3	412.89	867.84	845.64
	Logseries	428.67	866.68	861.45
6 SJG	PLN	385.12	784.11	776.48
	PLN2	374.64	<b>777.03</b>	<b>762.16</b>
	PLN3	373.16	787.94	766.27
	Logseries	385.70	780.64	775.51
7 SMG	PLN	473.41	960.69	953.06
	PLN2	461.99	<b>953.18</b>	<b>936.66</b>
	PLN3	461.52	966.84	942.54
	Logseries	473.50	956.74	951.10
8 SMG	PLN	474.65	963.91	955.49
	PLN2	464.08	956.78	940.91
	PLN3	442.90	<b>928.74</b>	<b>905.47</b>
	Logseries	474.51	958.56	953.12
9 TER	PLN	364.18	741.82	734.63
	PLN2	356.03	738.78	725.12
	PLN3	307.86	<b>655.81</b>	<b>636.09</b>
	Logseries	363.55	736.02	731.25
10 TER	PLN	419.17	851.70	844.63
	PLN2	409.65	<b>847.56</b>	<b>832.11</b>
	PLN3	407.48	857.34	834.74
	Logseries	420.79	851.00	845.69
11 TER	PLN	211.20	436.52	428.62
	PLN2	197.20	<b>420.61</b>	<b>407.56</b>
	PLN3	194.54	426.23	410.55
	Logseries	208.35	424.96	420.91
12 TER	PLN	400.22	812.82	806.85
	PLN2	388.77	<b>803.76</b>	<b>790.71</b>
	PLN3	386.57	812.47	793.75
	Logseries	398.68	806.62	801.47

Table 4. Model selection results for arthropod SADs of five Azorean islands. Arthropods were sampled in the native Laurisilva forest using a standardised pitfall trap and canopy beating transect methodology between 1999 and 2004. Column and model abbreviations and the number of parameters of each model are as for Table 3. Values in bold represent the best model for the corresponding island, for both BIC and AIC<sub>c</sub>. Where several models are within an AIC<sub>c</sub> or BIC value of two of one another, each is highlighted in bold.

Island	SAD Model	MLL	BIC	AIC <sub>c</sub>
1 Faial	PLN	400.22	814.34	806.68
	PLN2	388.77	<b>805.35</b>	<b>790.42</b>
	PLN3	387.57	818.67	794.69
	Logseries	398.68	<b>806.62</b>	801.47
	PLN	568.17	1151.21	1142.52
2 Flores	PLN2	560.22	1150.18	1133.07
	PLN3	550.77	<b>1146.15</b>	<b>1120.91</b>
	Logseries	572.87	1155.65	1149.83
3 São Jorge	PLN	582.15	1179.25	1170.47
	PLN2	573.19	<b>1176.28</b>	1158.99
	PLN3	567.85	1180.56	<b>1155.03</b>
	Logseries	593.13	1196.23	1190.35
4 São Miguel	PLN	766.36	1548.52	1538.84
	PLN2	756.69	<b>1544.99</b>	<b>1525.83</b>
	PLN3	754.65	1556.71	1528.28
	Logseries	775.09	1560.72	1554.25
5 Terceira	PLN	704.63	1424.66	1415.40
	PLN2	695.57	<b>1421.95</b>	<b>1403.65</b>
	PLN3	693.92	1434.05	1406.96
	Logseries	710.62	1431.51	1425.31

and mode two (more common species) for samples best modelled by PLN2 or PLN3 according to a Welch's two sample t-test (Table 5); mode one at the rarer end of the distribution always had a higher percentage of satellite and tourist species and a lower proportion of indigenous species. These results were consistent at both the fragment and island scale (Table 5), although the difference in the proportion of indigenous species across modes was not significant at the island scale. The percentage of species with high dispersal ability was also found to differ significantly between mode one and two, but the direction of the change was inconsistent among samples, that is, in some samples mode one had a higher percentage of species with high dispersal ability than mode 2, and vice versa. Thus, the effect of differences in dispersal ability as a mechanism for multimodal SADs in the Azores is unclear. No significant difference was found between any of the subsets between the second and third mode for samples best modelled by PLN3 (results not presented); although the degrees of freedom was reduced as PLN3 provided the best fit for fewer samples.

We found minimal evidence for the structuring of communities within each sample along a body size niche axis as outlined by ENT. Body size distributions were all distinctly unimodal, and the body size-abundance plots did not produce any apparent patterns. Mean body size was not found to significantly differ between mode one and mode two for datasets best modelled by PLN2 or PLN3 (Table 5).

While the randomly sampled null distribution varied between samples because they were drawn from different subsets of the archipelagic species pool and because of the number of species in each modal octave class, a rough rule of thumb in relation to the proportion of core species was

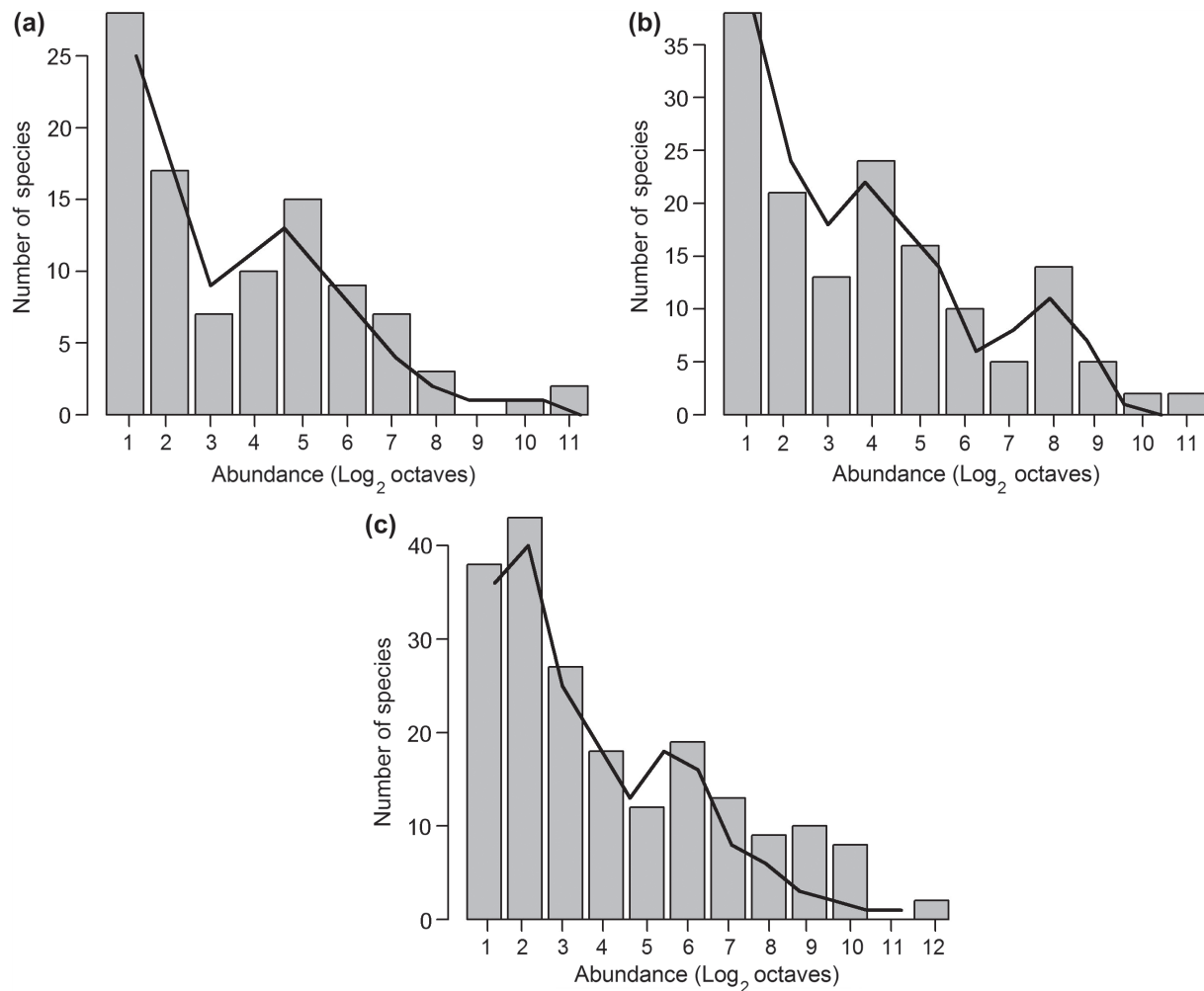


Figure 2. Exemplar multimodal species abundance distributions with fit lines for Azorean arthropods sampled using a standardised pitfall trap and canopy beating sampling methodology. Data were gathered from 1999–2004. (a) Fit of the PLN2 distribution (black line) to the data from fragment 1 (located on the island of Faial). (b) Fit of the PLN3 distribution (black line) to the data from island 2 (Flores). (c) Example of a SAD for which the PLN2 provides the best fit according to both information criterion (BIC and AIC<sub>c</sub>) but for which the bimodal pattern is not immediately striking. Data are from island 4 (São Miguel). Island samples were created by pooling the fragment data on each island. In each plot abundances are grouped into Log<sub>2</sub> octaves: octave 1 contains the number of species with 1 individual, octave 2 the number of species with 2–3 individuals, octave 3 the number of species with 4–7 individuals, and so forth.

Table 5. Welch's two sample t-test results for the composition of the different modes within the SAD, with the degrees of freedom (DF) and the corresponding p-value, for both the fragment and island samples, and the direction of change. The values only relate to samples in which the PLN2 distribution provided the best fit, that is, there were two distinctive modes. The direction of change relates to whether the proportion of a subset was consistently higher or lower in the more common mode across samples. For instance, considering dispersal ability, in some samples the rarer mode had a higher proportion of high dispersal ability species and in other samples the opposite pattern was observed. Values in bold represent results significant at the 0.05 level.

Subset	Direction of change	t	DF	p
Fragments				
Percentage of tourist species	lower in more common mode	7.76	20	<b>0.000</b>
Percentage of core species	higher in more common mode	−13.43	19	<b>0.000</b>
Percentage of high dispersal ability species	mixed	2.82	20	<b>0.011</b>
Body size	not applicable	−0.38	15	0.708
Indigenous species	higher in more common mode	−1.07	17	<b>0.002</b>
Islands				
Percentage of tourist species	lower in more common mode	4.59	8	<b>0.002</b>
Percentage of core species	higher in more common mode	−5.26	4	<b>0.005</b>
Percentage of high dispersal ability species	mixed	2.32	8	<b>0.049</b>
Body size	not applicable	−1.38	8	0.204
Indigenous species	higher in more common mode	−2.36	6	0.059

that the 2.5th and 97.5th percentiles were located at 5% and 35% core species. Thus, at both the fragment and island scale it was universally the case that there were more core species in the more abundant mode than expected by chance. For the rarer mode, there was never a lower proportion of core species than expected by chance at the fragment scale, but at the island scale three of the four samples in which PLN2 provided the best fit had a significantly lower proportion of core species than expected by chance. At the fragment scale, the rarer mode always had a higher proportion of tourist species than expected by chance, although this difference was significant ( $p < 0.05$ ) in only five of the eight cases. No significant differences in proportion of tourists were found at the island scale. In relation to the more abundant mode, two fragments and three islands had a lower proportion of tourist species than expected by chance. Considering the proportion of indigenous species, for two fragment samples and one island sample there were more indigenous species in the more common mode than expected by chance, and in no case was there significantly fewer indigenous species in the rarer mode. No consistently significant results in terms of the proportion of species with high dispersal ability or the mean body size were found.

Re-running the model selection using the core and satellite subsets separately revealed mixed results (BIC and  $AIC_c$  values for all models and for each sample are presented in Supplementary material Appendix 3 Table A3). For the core subset at the fragment scale it was often unclear which of the logseries or PLN provided the best fit as in only seven samples did a particular distribution provide the best fit according to both BIC and  $AIC_c$  using a minimum difference in each criterion of two. Nevertheless, apart from the single fragment where the PLN2 distribution was the best model, and a further two fragments where the PLN2 distribution was within the best model set according to  $AIC_c$ , the good fit of the PLN2 more generally, and thus the presence of multiple modes, was not apparent when focusing purely on the core subset of species. For the satellite subset the PLN performed best according to both criteria for all but one fragment. Results at the island scale were similar. Considering the core subset, the PLN provided the best for all island samples according to BIC and in four out of five instances according to  $AIC_c$ . The PLN performed best in all five instances for the satellite subset, according to both criteria.

That the PLN provided a better fit than the logseries for the satellite species in most samples is intriguing as analysis of the SAD frequency distributions reveals a logseries type pattern in the majority of instances. Figure 3 illustrates the pattern exhibited in most samples, although the exact form of the relationship varies between samples. The complete assemblage (Fig. 3a) exhibits two distinct modes, one of very rare species represented by only one individual in the sample, and one of relatively more common species. Deconstructing into core and satellite subsets indicates that the core subset exhibits a more lognormal type pattern, with the mode corresponding to the common species mode in the complete assemblage SAD. Accordingly, the satellite subset SAD exhibits a more logseries pattern (but is best fit by the PLN; Fig. 3c). It is clear then, for this sample at least,

that combining the core and satellite subsets within the complete assemblage results in two distinct modes in the SAD. Results were largely consistent between the two sets of core/satellite classification methods and we are thus confident that our choice of 50% as the grouping threshold did not affect the results. For example, Fig. 3 (d–f) shows the plots for the same data as Fig. 3 (a–c) but using the three group classification. It is evident that it is the ‘25% or fewer group’ which is generating the first mode in the SAD of the whole assemblage (Fig. 3a), and that the 76–100% group is generating the second mode.

## Discussion

The performance of Poisson lognormal SAD models that incorporate multiple modes was compared with the traditional one mode PLN and logseries distributions for twelve forest fragments, and the combination of fragments within five islands. Multiple modes were found in a large number of samples, with the PLN2 distribution most prevalent, followed by the PLN3 distribution. Consistent with what we have termed the amalgamation hypothesis (Introduction) it appears that the amalgamation of core and satellite, and indigenous and introduced subsets leads to the observed multiple modes. The presence of tourist species, which in many (but not all) instances are also satellite species, and which are present in larger abundances in more anthropogenic land uses, inflates the number of rare species in the samples and thus leads to the emergence of a rare mode in the SAD.

### A deconstruction approach

Combining assemblage deconstruction within a multimodal SAD framework has been put forward as a promising way of advancing SAD research (McGill et al. 2007). Nonetheless, our finding that grouping together core and satellite species into one assemblage results in two distinct components of the SAD is not unique. For instance, some studies have reported different SAD patterns for core and satellite subsets using time series data (Magurran and Henderson, Ulrich and Ollik 2004). Ulrich and Zalewski (2006) have also found different SAD patterns for core and satellite groups, in their case using a very similar methodology to the present study, determining group membership based on the proportion of samples across space in which each species was recorded (and see Unterseher et al. 2011 for an example using molecular taxonomic units).

Our findings differ from the aforementioned studies in that the satellite subsets were best modelled by a PLN distribution and the core species by a mixture of PLN, PLN2 and logseries distributions. Thus, in certain instances both the core and satellite subsets were best modelled by PLN distributions but with different modal octaves, a pattern similar to that observed by Ugland and Gray (1982) in marine benthic communities. Regardless of which distribution best fits the core subsets, for the bimodal samples it is clear that the proportion of core species is significantly higher in the more common mode. Examining the concept of core species reveals that this finding is to be expected.



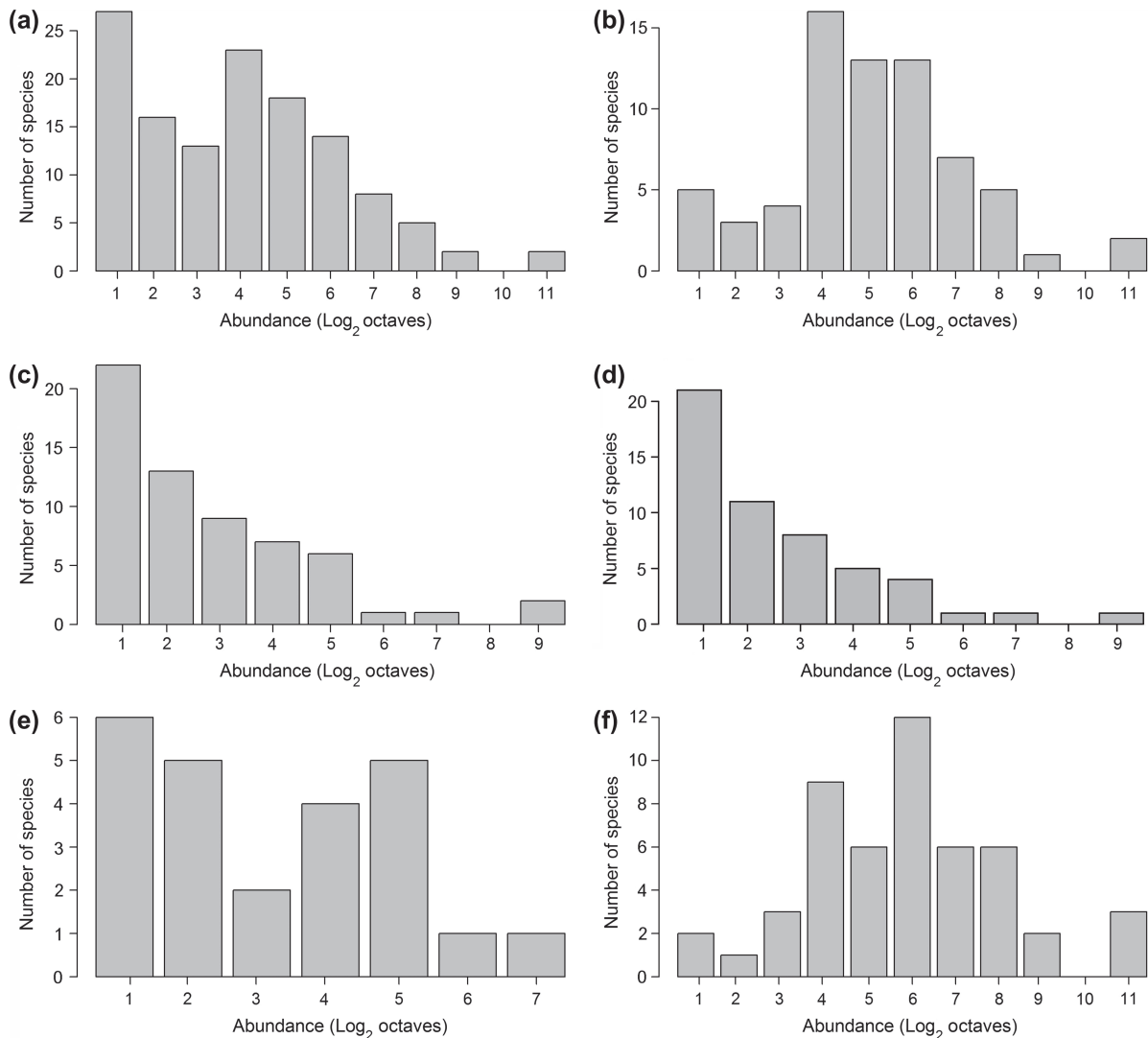


Figure 3. The SAD of arthropods from the island of Faial in the Azores (Table 2, Fig. 1). Species were sampled using a standardised pitfall trap and canopy beating sampling methodology from 1999–2004. Plots correspond to the SAD of: (a) the whole assemblage, deconstructed into, (b) core species and, (c) satellite species. Two modes are clearly present in (a) which correspond well with modal octaves in (b) and (c). (d–f) correspond to the three group core and satellite approach, whereby d) represents species found in 25% or fewer fragments, (e) 26–75% of fragments and (f) 76–100% of fragments. In each plot abundances are grouped into  $\log_2$  octaves: octave 1 contains the number of species with 1 individual, octave 2 the number of species with 2–3 individuals, and so forth.

Core species, as classified herein, are those which are theorised to be the constituent members of the community, organised according to traditionally invoked deterministic niche-based processes, such as competition and predation (Begon et al. 1990, Ulrich and Zalewski 2006). As such, it intuitively makes sense that they will, by and large, obtain higher abundances than the satellite species, which are predominantly immigrants from outside the community. It is also possible that our method of classifying species as core and satellite has influenced our findings. While our method of classification is independent of species abundance at the fragment scale in that a species classified as core or satellite can theoretically have any abundance in a particular fragment, and we found that our 50% threshold was robust, a positive occupancy–abundance relationship is known to exist for a variety of taxa and systems, including for Azorean

arthropods (Rigal et al. 2013). As we have classified species based on incidence, it is possible that the core species have a higher abundance simply as a result of this relationship. However, the large overlap in identity between satellite and tourist species indicates that many satellite species obtain much larger abundances in other land uses and we are confident that our results are not simply a result of a positive occupancy–abundance relationship. Equally, as we have pooled data from individual fragments to create our island datasets, there is a chance that a degree of circularity exists in our island datasets, i.e. whereby the classification of a species as core is not independent of its abundance. For example, if an island contained two thirds of the fragments in our dataset, the pooled abundance of any core species would have to be greater than one, and thus the species could not be represented in the singleton octave of the SAD. In reality,

this is only an issue if any islands contain a large proportion of the fragments used in the core-satellite classification. As the island in our dataset which contains the most sampled fragments (Terceira; Table 2), only contains four fragments, it is unlikely that this circularity is driving the observed patterns at the island scale.

While we found no evidence of communities being structured along a body size niche axis, this does not disprove ENT in respect to our study system. We only focused on body size as a measure of the niche, which while recommended by Vergnon et al. (2012), is not the sole niche axis along which ecological communities could be structured (Scheffer and van Nes 2006, Vergnon et al. 2012). As no additional measures of a niche axis were tested, all we can say with surety is that ENT based on body size as an indicator of the niche is not the cause of the observed multimodal SADs in our data. It is also possible that focusing on the body size of arthropods in general obscures any general pattern as this grouping comprises a mix of various taxa, e.g. spiders, beetles, and Lepidoptera, each with a different body size plan or 'bau plan'.

#### **Satellites, tourists and introduced species**

It has been postulated that satellite species dynamics are mainly controlled by stochastic processes, specifically random dispersal, as opposed to the niche-based mechanisms hypothesised to be structuring core species (Magurran and Henderson 2003, Ulrich and Ollik 2004). Our results seemingly back up this line of argument. In the Azorean invertebrate arthropod community it appears that species classified as satellite species are predominantly, but not universally, also tourist species. That is, they are not simply species which are only occasionally found in a set of samples (i.e. satellite species), but are also species which are more abundant in other non-native habitats (i.e. tourist species; Borges et al. 2008). Thus, it is the immigration of small numbers of individuals from these different habitats into the native forest that results in the mode of rare species. This is a key finding. While other studies have focused on the core/satellite dichotomy and assumed that satellite species have different habitat requirements, we have gone a step further and shown that the presence of tourist species in an assemblage significantly affects the signal in the SAD.

Our data derive from native forest fragments on the Azores, a system where there has been large scale conversion of native forest to semi-natural and intensive pasture alongside a long and continuing process of anthropogenic introduction of non-native species (Borges et al. 2008, Cardoso et al. 2010). Exotic forestry plantations are also a common feature of the landscape and, currently, native forest represents only a small amount of the total Azorean land use (Cardoso et al. 2009). These changes, in combination, mean that many arthropod species now found in the Azores are better adapted to the more anthropogenic habitats. It is thus evident that the concepts of satellite species and tourist species are interlinked. The species sampled in only a few fragments, i.e. the satellite species, are largely immigrants from land uses outside the fragment and will thus naturally be represented in our samples by only a few individuals (see also Gray et al. 2005, 2006b). However, linked to the

issues described above, it is possible that our classification of tourist species influenced our results. We classified a species as tourist if it possessed a greater abundance in any of the three anthropogenic habitats relative to the native forest. Thus, implicit in this classification is our assumption that species should be abundant in one habitat (e.g. native forest) and rare in the others (e.g. the more anthropogenic habitats). However, this is not necessarily the case and it is possible that a positive relationship between abundance in the different habitat types exists, such that certain species are abundant in both sets of habitats and vice versa. Obviously, this positive relationship will characterise certain species. For instance, in our data roughly 10–20% of tourist species are core species, depending on the sample in question, and these are likely to be predominantly habitat generalists; species found in large abundances in multiple habitats (Shmida and Wilson 1985, Borges et al. 2008, Cardoso et al. 2009, 2013). Nonetheless, in general we do not think this assumption is problematic in our analysis, as overall we found a negative correlation between abundance in the native forest and maximum abundance in the additional habitats (both variables log transformed; Pearson's product-moment correlation:  $r = -0.23$ ,  $p = 0.008$ ).

It is also the case that there is a higher proportion of introduced species in the rarer mode of the bimodal SADs. The Azorean invertebrate arthropod fauna includes a vast number of introduced species (Borges et al. 2010), evidenced by the observation that even in the common mode of the SAD up to 40% of the species in the modal octave were classified as introduced. For instance, one of the most widespread introduced species in man-made habitats, the millipede *Ommatoiulus moreletii*, was also successful in invading native forest, where it is now abundant. It has been shown that introduced species are integrated alongside indigenous species in the contemporary Azorean arthropod communities such that they are mostly undetectable through the study of abundance–occupancy plots (Rigal et al. 2013). This implies that current island faunas with a high proportion of introduced species may have an inflated number of rare species in native habitats due to source–sink dynamics (see also Borges et al. 2008).

#### **Distinguishing between two and three modes**

While the grouping of core and satellite, and indigenous and introduced species in a sample appears to explain bimodal SADs, it does not provide an answer as to why certain samples were best modelled by the PLN3 distribution. For instance, taking fragment 9 (Table 1) as an example it can be seen that PLN3 provides the best fit for the whole assemblage (Table 3), and PLN2 the best for the core species (Supplementary material Appendix 3 Table A3a). Inspection of the frequency distributions reveals the modal octaves for the core species are 8–15 and 64–127 species, which is largely congruent with the pattern for the whole assemblage (although without the mode of very rare species), while the satellite species have a large number of singletons. Hence, the two more common modes of the whole assemblage are both mainly comprised of core species, with few to zero tourist species, meaning that there must be a further

causal mechanism driving this pattern. The grouping of species into ecological guilds may add some explanatory power (Borges et al. 2008) and it is also possible that in such communities there are three distinct groups of species, as postulated by Ugland and Gray (1982, see also Borges et al. 2008). Further research is needed to elicit the structuring mechanism(s) governing these multimodal groupings.

We found that for certain samples distinct multiple modes were not particularly evident visually (e.g. cf. Fig. 1c), although they were detected to be multimodal in our analyses. Thus, it may be the case that many well studied communities are in fact multimodal: it is simply that appropriate statistical methods have not been applied to detect this (see also McGill et al. 2007, Vergnon et al. 2012). As the R code is now available (Dornelas and Connolly 2008) to test multimodal PLN distributions we call on researchers who are exploring SAD patterns in different datasets to statistically test for these patterns, even if they are not immediately apparent visually.

In sum, the amalgamation of samples through time and space can lead to a bimodal – or even trimodal – SAD within which there is a discernible group of occasional species each represented by only a few individuals, leading to a distinct mode at the rare end of the SAD. In the Azorean laurisilva forest arthropod community many of these satellite taxa are introduced species best adapted to non-native land uses but occasionally present and sampled in the native forest. Thus, deconstructing the full assemblage into various subsets (Marquet et al. 2004, Borges et al. 2008) can aid in elucidating the underlying patterns and we therefore recommend this approach for future analyses of SADs. Nonetheless, it is imperative that care is taken in determining ecologically rational deconstructions as simply classifying all relatively rare species from a sample as satellite species, for example, will obviously result in a change in the form of the core and satellite SADs. Multimodal SADs may be more common than originally perceived and we advocate testing for them using appropriate rigorous statistical methods.

**Acknowledgements** – We are grateful to Sean Connolly and Zhe Sha for support with the various methods employed. Luís Borda-de-Água provided constructive comments which improved the manuscript. Alison Pool is thanked for her help with large amounts of data entry and organising the figures. We are grateful to all the researchers who collaborated in the field and laboratory work, and to the Azorean Forest Services and Environment Services for providing local support on each island. Acknowledgements are also due to the taxonomists who assisted in species identification and to Enésima Mendonça and Sofia Terzopoulou for the body size measurements. Data used in this manuscript were obtained from the projects funded by Direção Regional dos Recursos Florestais (project: 17.01-080203, 1999–2004) and Direção Regional da Ciência e Tecnologia (project: ‘Consequences of land-use change on Azorean fauna and flora - the 2010 Target’, M.2.1.2/I/003/2008. TJM’s work in the Azores was funded through a Santander Academic Travel Grant, and he also acknowledges funding from the EPA Cephalosporin Fund and the Sidney Perry Foundation. PAVB was partly funded by the project FCT-PTDC/BIA-BIC/119255/2010 – ‘Biodiversity on oceanic islands: towards a unified theory’.

## References

- Alonso, D. et al. 2008. The implicit assumption of symmetry and the species abundance distribution. – *Ecol. Lett.* 11: 93–105.
- Begon, M. et al. 1990. *Ecology: individuals, populations and communities*. – Blackwell.
- 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.
- Borda-de-Água, L. et al. 2012. Spatial scaling of species abundance distributions. – *Ecography* 35: 549–556.
- Borges, P. A. V. et al. 2005. Ranking protected areas in the Azores using standardized sampling of soil epigeic arthropods. – *Biodivers. Conserv.* 14: 2029–2060.
- 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. et al. 2010. List of arthropods (Arthropoda). – In: Borges, P. A. V. et al. (eds), *A list of the terrestrial and marine biota from the Azores*. Princípiá, pp. 179–246.
- Bulmer, M. G. 1974. On fitting the Poisson lognormal distribution to species abundance data. – *Biometrics* 30: 101–110.
- Burnham, K. P. and Anderson, D. R. 2002. *Model selection and multi-model inference: a practical information-theoretic approach*, 2nd edn. – Springer.
- Burnham, K. P. and Anderson, D. R. 2004. Multimodel inference: understanding AIC and BIC in model selection. – *Sociological Meth. Res.* 33: 261–304.
- Cardoso, P. et al. 2009. A spatial scale assessment of habitat effects on arthropod communities of an oceanic island. – *Acta Oecol.* 30: 590–597.
- Cardoso, P. et al. 2010. Patterns of alpha and beta diversity of epigeic arthropods at contrasting land-uses of an oceanic island (Terceira, Azores). – In: Serrano, A. R. M. et al. (eds), *Terrestrial arthropods of Macronesia – biodiversity, ecology and evolution*. Soc. Portuguesa Entomol., Lisboa, pp. 73–88.
- Cardoso, P. et al. 2013. Integrating landscape disturbance and indicator species in conservation studies. – *PLoS ONE* 8: e63294.
- Dornelas, M. and Connolly, S. R. 2008. Multiple modes in a coral species abundance distribution. – *Ecol. Lett.* 11: 1008–1016.
- Fisher, R. A. et al. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. – *J. Anim. Ecol.* 12: 42–58.
- Gaspar, C. et al. 2008. Diversity and distribution of arthropods in native forests of the Azores archipelago. – *Arquipélago Life Mar. Sci.* 25: 1–30.
- Gaston, K. J. 1994. *Rarity*. – Chapman and Hall.
- Gray, J. S. et al. 2005. The impact of rare species on natural assemblages. – *J. Anim. Ecol.* 74: 1131–1139.
- Gray, J. S. et al. 2006a. Are there differences in structure between marine and terrestrial assemblages? – *J. Exp. Mar. Biol. Ecol.* 330: 19–26.
- Gray, J. S. et al. 2006b. On plotting species abundance distributions. – *J. Anim. Ecol.* 75: 752–756.
- Hanski, I. 1982. Dynamics of regional distribution: the core and satellite species hypothesis. – *Oikos* 38: 210–221.
- Holling, C. S. 1992. Cross-scale morphology, geometry and dynamics of ecosystems. – *Ecol. Monogr.* 62: 447–502.
- Labra, F. A. et al. 2005. Distribution and abundance: scaling patterns in exotic and native bird species. – In: Sax, D. F. et al. (eds), *Species invasions: insights into ecology, evolution, and biogeography*. Sinauer, pp. 421–446.
- 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. V. and Heaney, L. R. (eds), *Frontiers of biogeography: new directions in the geography of nature*. Sinauer, pp. 191–210.
- McGill, B. J. 2011. Species abundance distributions. – In: Magurran, A. E. and McGill, B. J. (eds), *Biological diversity: frontiers in measurement and assessment*. Oxford Univ. Press, pp. 105–122.
- 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.
- Pielou, E. C. 1969. *An introduction to mathematical ecology*. – Wiley.
- Preston, F. W. 1948. The commonness, and rarity of species. – *Ecology* 29: 254–283.
- 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.
- Rigal, F. et al. 2013. Integration of non-indigenous species within the interspecific abundance–occupancy relationship. – *Acta Oecol.* 48: 69–75.
- Scheffer, M. and van Nes, E. H. 2006. Self-organized similarity, the evolutionary emergence of groups of similar species. – *Proc. Natl Acad. Sci. USA* 103: 6230–6235.
- Segura, A. M. et al. 2011. Emergent neutrality drives phytoplankton species coexistence. – *Proc. R. Soc. B.* 278: 2355–2361.
- Shmida, A. and Wilson, M. V. 1985. Biological determinants of species diversity. – *J. Biogeogr.* 12: 1–20.
- Ugland, K. I. and Gray, J. S. 1982. Lognormal distributions and the concept of community equilibrium. – *Oikos* 39: 171–178.
- Ulrich, W. and Ollik, M. 2004. Frequent and occasional species and the shape of relative-abundance distributions. – *Divers. Distrib.* 10: 263–269.
- Ulrich, W. and Zalewski, M. 2006. Abundance and co-occurrence patterns of core and satellite species of ground beetles on small lake islands. – *Oikos* 114: 338–348.
- Ulrich, W. et al. 2010. A meta-analysis of species–abundance distributions. – *Oikos* 119: 1149–1155.
- Unterseher, M. et al. 2011. Species abundance distributions and richness estimations in fungal metagenomics – lessons learned from community ecology. – *Mol. Ecol.* 20: 275–285.
- Vergnon, R. et al. 2012. Emergent neutrality leads to multimodal species abundance distributions. – *Nat. Commun.* 3: 663.
- Whittaker, R. J. and Fernández-Palacios, J. M. 2007. *Island biogeography: ecology, evolution and conservation*. – Oxford Univ. Press.

Supplementary material (available online as Appendix oik-00829 at <[www.oikosoffice.lu.se/appendix](http://www.oikosoffice.lu.se/appendix)>). Appendix 1–3.