






A biogeographical perspective on species abundance distributions: recent advances and opportunities for future research

Thomas J. Matthews^{1,2*} , Paulo A.V. Borges² ,
Eduardo Brito de Azevedo³ and Robert J. Whittaker^{4,5} 

¹GEES (School of Geography, Earth and Environmental Sciences), The University of Birmingham, Birmingham B15 2TT, UK, ²CE3C – Centre for Ecology Evolution and Environmental Changes/Azorean Biodiversity Group and Universidade dos Açores – Departamento de Ciências e Engenharia do Ambiente, 9700-042 Angra do Heroísmo, Açores, Portugal, ³Dep. de Ciências Agrárias, CMMG (Centro de Estudos do Clima Meteorologia e Mudanças Globais) CITA-A, Universidade dos Açores, 9700-042 Angra do Heroísmo, Portugal, ⁴Conservation Biogeography and Macroecology Programme, School of Geography and the Environment, University of Oxford, South Parks Road, Oxford OX1 3QY, UK, ⁵Center for Macroecology, Evolution and Climate, Department of Biology, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark

*Correspondence: Thomas J. Matthews, School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham B15 2TT, UK.
E-mail: t.j.matthews@bham.ac.uk

ABSTRACT

It has become increasingly recognized that multiple processes can generate similar shapes of species abundance distributions (SADs), with the result that the fit of a given SAD model cannot unambiguously provide evidence in support of a given theory or model. An alternative approach to comparing the fit of different SAD models to data from a single site is to collect abundance data from a variety of sites, and then build models to analyse how different SAD properties (e.g. form, skewness) vary with different predictor variables. Such a biogeographical approach to SAD research is potentially very revealing, yet there has been a general lack of interest in SADs in the biogeographical literature. In this *Perspective*, we address this issue by highlighting findings of recent analyses of SADs that we consider to be of intrinsic biogeographical interest. We use arthropod data drawn from the Azorean archipelago to further highlight how analyses of SAD form and function may be biogeographically informative. We hope that, by reviewing a number of novel approaches, our article may prove to be a catalyst for a greater interest in analysing SADs in biogeography.

Keywords

applied biogeography, arthropods, Azores, function regression, gambin model, lognormal model, logseries model, species abundance distribution

INTRODUCTION

A species abundance distribution (herein ‘SAD’) describes how the total number of individuals sampled within a given community is distributed amongst the sampled species. Patterns of SADs have been studied for over 70 years, with seminal papers published in the 1940s by Fisher *et al.* (1943) and Preston (1948), but have received renewed interest in the last decade (McGill *et al.*, 2007; McGill, 2011; Matthews & Whittaker, 2015). One common recent use of the SAD has been as a means to test different ecological theories. For example, following the introduction of Hubbell’s (2001) spatially implicit neutral model (SINM) for application in biogeography and biodiversity studies, many authors attempted to test the theory by evaluating the fit of the SAD predicted by Hubbell’s SINM to empirical data – with varying levels of success (e.g. Hubbell, 2001; McGill, 2003; Etienne, 2005; for a review see Matthews & Whittaker, 2014).

Unfortunately, it has become increasingly apparent that multiple processes can generate similar shapes of SAD curves, thus causing a problem of equifinality: the fit of a given SAD model cannot unambiguously provide evidence in support of a given theory (McGill *et al.*, 2007). An alternative and less frequently applied approach to that of simply comparing the fit of different SAD models to data from a single site, is to collect species abundance data from a variety of sites to build models that can then be used to analyse how different SAD properties (e.g. form, skewness) vary with different predictor (environmental) variables. In other words, the question of which mechanisms drive SADs can be approached instead by assessing which biogeographical variables control the relative abundances of species. Such an approach to SAD research may be more revealing and encourage interest in using SADs in biogeographical studies.

Here, we set out to highlight the use of SADs as a fruitful and instructive approach in biogeographical research and we

hope that this article may encourage greater interest in analysing SADs amongst biogeographers. We begin by describing how two different regression-based methods (standard linear regression and function regression) can be used to determine the mechanisms underlying SADs and provide examples from the recent literature of studies that have used these approaches. This is followed by an analysis of arthropod data from the Azorean archipelago to illustrate how the study of SADs can generate interesting information from a biogeographical point of view.

USE OF REGRESSION METHODS FOR EXPLAINING VARIATION IN SADS

Assessing SAD model parameters within a standard linear regression framework

One useful way of assessing which variables underpin various SAD properties is to use a parameter from a given SAD model as the response variable in, for example, a regression model. Traditionally, the lognormal and logseries models have been used in such an approach (see Matthews & Whittaker, 2015). For instance, Sæther *et al.* (2013) used a Scottish lake macro-benthos community dataset to show that the σ^2 parameter (the variance) of the lognormal model varied significantly according to the severity of pollution.

While the lognormal and logseries models are sound statistical representations of many empirical communities and still represent useful tools, more recently introduced models offer additional possibilities for such analyses. For example, the immigration parameter (m) of Hubbell's (2001) SINM can be used to make inferences regarding the importance of dispersal from the metacommunity into the local community: if m is close to 1, the local community is a random sample of the regional metacommunity, whereas if m is close to zero, the local community receives very few immigrants from the metacommunity (Hubbell, 2001; Matthews & Whittaker, 2014). The m parameter can easily be estimated by maximum likelihood fitting of the zero-sum multinomial SAD model (Etienne, 2005). If multiple sites are surveyed it is then possible to regress m against various predictor variables to determine what factors influence the importance of dispersal in ecological communities. For example, in one recent analysis Chust *et al.* (2013) found that m was consistently lower in tropical relative to temperate marine phytoplankton communities. The frequency dependence parameter (δ) of Jabot & Chave's (2011) non-neutral generalization of Hubbell's model provides an alternative parameter to use in such exercises (e.g. see Jabot & Chave's 2011, Figure 2 for a correlation between δ and precipitation for a variety of tropical tree plots).

The gambin SAD model (Matthews *et al.*, 2014a) can also be used in biogeographical analyses (e.g. Dornelas *et al.*, 2011). The gambin model combines the flexible gamma distribution with a binomial sampling process. It is a single free parameter model and its parameter (α) characterizes the shape of the SAD. Low values characterize logseries SAD

shapes, while higher values indicate lognormal curve shapes (Ugland *et al.*, 2007; Matthews *et al.*, 2014a). Extreme values can indicate more complex situations in which common species are the most prevalent type of species in the community. Thus, α condenses the shape of the SAD into a single value that can then be used in regression models. The gambin model is beneficial in this regard as it is flexible and has been found to fit a wide range of SAD shapes.

A recent paper published in *Journal of Vegetation Science* by Ulrich *et al.* (2015) provides a useful example of what can be gained from a biogeographical approach to SAD research. They examined the SADs of 605 tree assemblages across six continents and fitted two SAD models (logseries and lognormal) in rank-abundance form and related various properties of the observed SADs (e.g. shape and evenness) to geographical and climatic variables (e.g. latitude, elevation and evapotranspiration) using linear regression. Relationships between latitude and SAD evenness and shape were found. Logseries distributions were more prevalent at lower latitudes, while there was an increase in the prevalence of lognormal distributions towards northern latitudes. Again, while these results are interesting in themselves, what is perhaps more intriguing is that this approach allows the user to make inferences regarding classic SAD and community assembly theory. For example, the lognormal distribution has traditionally been used to model undisturbed ecological communities (Ugland *et al.*, 2007; Matthews & Whittaker, 2015). In contrast, logseries SADs have been linked to disturbed communities and communities in severe environments with low productivity (Gray *et al.*, 1979; Hill & Hamer, 1998; Ugland *et al.*, 2007). As such, a number of studies have used deviation from a lognormal distribution as a means of assessing the impact of disturbance (e.g. pollution) on natural communities (e.g. Hill & Hamer, 1998; reviewed in Matthews & Whittaker, 2015). The results of Ulrich *et al.* (2015) are at variance with this theory, as it was found that logseries distributions were associated with species rich, productive and low-latitude tree communities. Instead, their results point to the importance of dispersal and stochastic processes in shaping the SADs of tree communities. It is worth noting that the R^2 values of their models were generally quite low (< 0.3), indicating that there must be other important 'hidden variables' that were not included in the study, thus pointing towards information needs for future biogeographical SAD studies and meta-analyses. For example, isolation is known to be an important variable in many ecological systems and may therefore explain some of the additional variation in Ulrich *et al.*'s data. A distance-decay approach may be useful in such circumstances to evaluate the impact of between-patch distances on SAD metrics.

Function regression and the species abundance distribution

A recent paper by Yen *et al.* (2015) introduces the method of function regression in the context of ecological applications. Function based regression models have been used in

the statistical literature for some time (see Müller & Stadtmüller, 2005), but as Yen *et al.* note, applications in biogeography and ecology are uncommon. Function regression models enable functions to be used as the response variable (and as predictor variables, although this possibility is not discussed here), which allows the user to regress a function on different predictor variables (Yen *et al.*, 2015). Questions such as ‘how does the form of the SAD change in response to fragment area and/or isolation?’ are perfectly suited to this methodological approach. However, instead of condensing the SAD into a single value, for example the gambin α parameter, function regression models allow us to use function-valued data and therefore the whole SAD, as the response variable. Multiple methods are available to fit function regression models (e.g. Bayesian estimation using reversible-jump Markov chain Monte Carlo computation). The original ‘FREE’ R package (see Yen *et al.*, 2015) provides functionality for six of these different methods. An updated version of the package (‘FREELITE’) that includes fewer fitting methods is also available. The use of function regression in SAD research is technically more complex than the simple linear models discussed above, but the former are arguably more revealing models as they allow the user to determine how different parts of the SAD are affected by a given predictor, e.g. the relative abundances of very common species. As such, ‘a function-valued method is likely to provide much deeper ecological and evolutionary insight’ (Yen *et al.*, 2015, p. 18). In their paper, Yen *et al.* provided examples of function regression using individual size distributions of various taxa.

EXEMPLIFICATION WITH AZOREAN ARTHROPOD SADS

Using standard linear regression to examine variation in gambin’s alpha parameter

To highlight the utility of using SAD model parameters within a linear regression framework, we focused on the α parameter of the gambin model (described above). We used well-specified arthropod SAD data from the long-running Biodiversity of Arthropods from the Laurisilva of Azores (BALA) project in the Azores (see Borges *et al.*, 2005; Ribeiro *et al.*, 2005) in combination with a linear regression modelling framework to determine which environmental variables explained variance in the shape of the SAD. Arthropods were sampled using a combination of pitfall traps and a canopy beating method, following a rigorously standardized protocol. Eighteen fragments of native Laurisilva forest were sampled across seven islands in the Azorean archipelago over 13 years. The full sampling methodology can be found in Gaspar *et al.* (2008). The gambin model was fitted to the SADs from these 18 fragments using the ‘GAMBIN’ R package (Matthews *et al.*, 2014a), recording the α parameter in each case. As differences in sample size have been found to influence α (Matthews *et al.*, 2014a), we used a re-sampling

approach in which we first determined the fragment with the smallest number of individuals, denoting the number of individuals in this fragment as n . For each of the remaining 17 fragments, we then randomly sampled n individuals and fit the gambin model to the sampled data. This process was repeated 100 times in each case and the average α value of 100 iterations taken. The response variable therefore constituted the standardised α values for 18 forest fragments. We collected data on five predictor variables, representing different fragment characteristics: fragment area, fragment isolation, precipitation, temperature and relative humidity (RH). Climatic data were obtained from the CIELO model (de Azevedo *et al.*, 1999). Fragment area and isolation were obtained from Gaspar *et al.* (2008) and were calculated using a geographic matrix of centroids using the DIVA-GIS software (Hijmans *et al.*, 2005). Isolation was measured in metres as the distance between fragments within an island, except in the case of the single fragment on the island of Santa Maria, for which we used the distance to the closest fragment on the nearest island of São Miguel. All predictors were log-transformed (base e), which induced normality. Variance inflation factors were used to assess multi-collinearity between predictors, while Cook’s distance was used to identify any outliers. Two data points were removed as outliers, so that our results are for 16 fragments only. The dredge function in the ‘MuMIN’ R package (Bartón, 2012) was used to fit a complete set of models that were compared using Akaike’s information criterion corrected for small sample size (AIC_c ; Burnham & Anderson, 2002). Weight of evidence values were calculated for each predictor as the sum of the AIC_c weights from all models in which a predictor was included (Burnham & Anderson, 2002).

The best model contained isolation and temperature, both of which had relatively high WoE values (Table 1). The best model had an adjusted R^2 of 0.74, which is noteworthy, as it indicates that a substantial part of the variation in α between sites can be attributed to isolation and temperature alone, for these 16 fragments. Isolation was included in all models within 10 ΔAIC_c of the best model, and had a near maximum WoE value (0.99). The effect of isolation was negative (see Fig. 1), meaning that increasing isolation results in a decrease in α and therefore a shift from a more lognormal-like SAD towards a more logseries-shaped distribution. This is an interesting finding because it could be expected that more isolated fragments/islands will have fewer really rare species of forest-dependent arthropods due to reduced rescue effects and re-colonisations following patch level extinctions. Thus, based on this line of reasoning, more isolated fragments would be expected to have a lower proportion of rarer species and therefore have SADs closer to lognormal in form. One possible ecological interpretation of this result is that, as these fragments are isolated in a matrix of human-modified habitats, our samples contain a relatively high proportion of tourist species represented by few individuals (Borges *et al.*, 2008), and that the proportion of tourist species is higher in the more isolated fragments. For example, in our previous

Table 1 The results of the linear regression analysis. The response variable was the standardized gambin α value from 16 arthropod SADs in native Laurisilva forest fragments, in the Azores. The five predictor variables were fragment area, isolation, precipitation (Precip.), relative humidity (RH) and temperature (Temp.). The best model and all models within 4 ΔAIC_c of the best model are shown. The weight of evidence (WoE) of each predictor was calculated by summing the AIC_c weights ($wAIC_c$) of each model in which a predictor was included. NI indicates a variable was not included in a model.

Model	Area	Isolation	Precip.	RH	Temp.	ΔAIC_c	$wAIC_c$
1	NI	-0.19	NI	NI	2.01	0	0.67
2	NI	-0.19	NI	3.00	2.59	3.60	0.11
WoE	0.11	0.99	0.10	0.15	0.97		

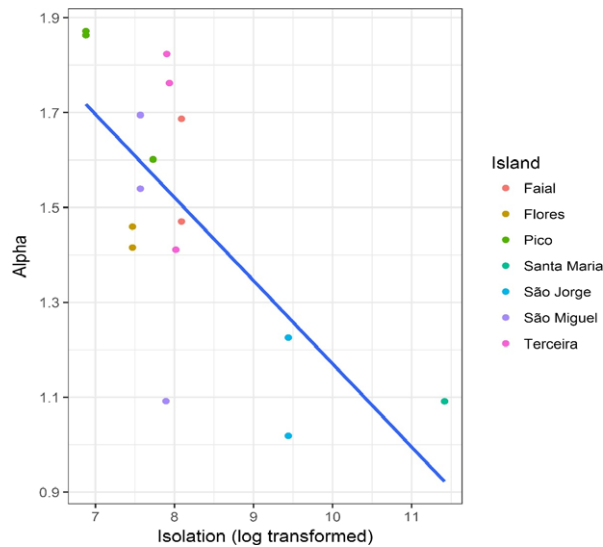


Figure 1 The relationship between fragment isolation and the standardized alpha parameter value of the gambin species abundance distribution model. The data are 16 arthropod SADs from native Laurisilva fragments in the Azores. Fragment isolation ranged from 970 m to 90,780 m. The blue line represents the best fit linear regression model. The islands in which the fragments are located are provided in the key. [Colour figure can be viewed at wileyonlinelibrary.com]

work we found that the presence of non-forest specialist species within fragments affects the shape of the SAD (Matthews *et al.*, 2014b). Nonetheless, the more important point in the context of the current paper is how this methodological approach can be used to determine the variables driving variation in SADs between sites, and in turn provides useful information on the key processes underpinning community structure.

Examining variation in the shape of Azorean arthropod SADs using function regression

To our knowledge function regression has not been used in conjunction with SADs and so we return to the Azorean arthropod data described above to illustrate the approach.

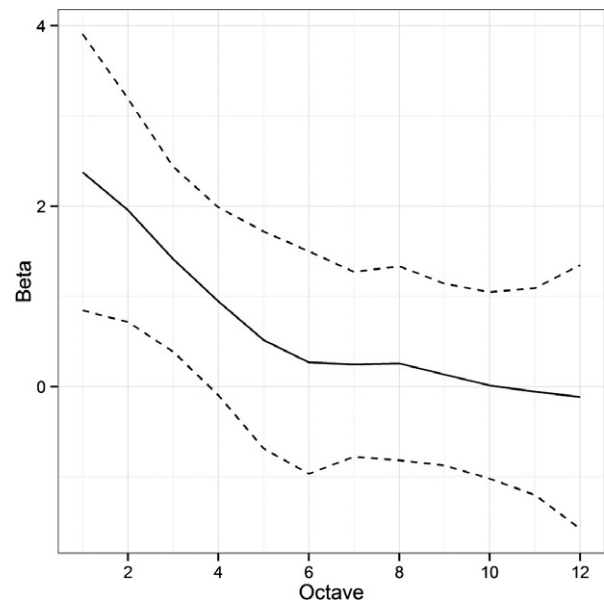


Figure 2 Fitted parameter estimates from a function regression model. The fitted curve indicates the effect (Beta) of the predictor variable (fragment isolation) on the species abundance distributions of arthropods in 18 native Laurisilva forest fragments, in the Azores. A higher value of Beta indicates a greater effect of isolation on that particular area of the SAD (see the main text for further information). The solid line represents the mean value, and the dashed lines represent the approximate 95% pointwise credible intervals. The SAD data were binned into octaves using a simple \log_2 transformation: octave 1 contains the number of species with 1 individual, octave 2 the number of species with two or three individuals, octave 3 the number of species with 4–7 individuals, and so forth.

For this analysis, we used the SAD data from the 18 forest fragments as the response variable, but this time in their raw form. We used two predictor variables: fragment area and isolation. Following Yen *et al.* (2015), both predictors were standardized to have a mean of zero and a standard deviation of one. The function regression model was fitted using the `FREE` R package, and we used the 'INLA' (Integrated nested Laplace approximation) method as it was found to perform well in Yen *et al.*'s various tests. A Gaussian error structure was assumed. To convert the SAD data into a matrix, we binned the data from each fragment into octaves using functionality available in the gambin R package (Matthews *et al.*, 2014a). This procedure uses a simple \log_2 transformation that doubles the number of abundance classes within each octave (see method 3 in Gray *et al.*, 2006). A matrix was then created in which the columns represented abundance octaves, and the rows represented fragments.

The R^2 value of the resultant model was high (0.87), but the credible intervals around the model coefficient estimates were quite large (e.g. Fig. 2). In the function regression analysis, isolation had a relatively large positive effect on the number of species in the rarer octaves (Table 2; Fig. 2), which in turn will make the SAD more logseries-like. The

Table 2 The mean fitted model coefficients for a function regression model with two predictor variables: fragment area and isolation. The response variable was 18 arthropod species abundance distributions (binned into octaves) from 18 fragments of native Laurisilva forest, in the Azores. The SAD data were binned into octaves using a simple log₂ transformation.

	Octave											
	1	2	3	4	5	6	7	8	9	10	11	12
Intercept	26.5	18.3	13.0	10.2	9.5	8.4	6.8	4.8	2.9	1.2	0.5	0.0
Area	-1.5	-1.2	-0.9	-0.7	-0.5	-0.5	-0.4	-0.4	-0.3	-0.2	-0.1	0.0
Isolation	2.4	2.0	1.4	0.9	0.5	0.3	0.2	0.3	0.1	0.0	-0.1	-0.1

results of this analysis match up well with our analysis above in which we used the gambin α parameter as the response variable in a standard linear regression model. However, the use of the function regression model allows us to make additional observations, which we were unable to make using the simple linear model. For example, from Fig. 2 it is clear that the effect of isolation is most pronounced for the two rarest octaves (1:2), and the effect decreases towards the more common octaves. Thus, it does appear that isolation is resulting in a relatively higher proportion of rare, possibly tourist (see e.g. Borges *et al.*, 2008), species in isolated fragments. A single parameter value (e.g. α) will never be able to convey this detailed level of information; which is one reason why function regression represents a potentially useful tool for SAD studies.

CONCLUSIONS

Many authors have commented on how simply evaluating the fit of a predicted SAD is a poor test of any ecological theory (e.g. McGill, 2003; McGill *et al.*, 2007; Matthews & Whittaker, 2014; May *et al.*, 2015). Perhaps then a more fruitful avenue for SAD research is to adopt a more biogeographic perspective, and examine the factors that underpin the observed variation in SAD form between sites in both space and time. A large number of biogeographical studies have attempted to interpret the parameters of the power species–area relationship model (c and z) ecologically by assessing which predictor variables explain variation in the parameters across datasets (Connor & McCoy, 1979; Triantis *et al.*, 2012; Matthews *et al.*, 2015). Similar biogeographic analyses involving parameters of SAD models are much less prevalent in the biogeographical literature, probably due in part to the additional data requirements involved in constructing SADs. In fact, obtaining standardized abundance values for many species across large scales is not trivial, particularly for invertebrates. However, many SAD datasets have now been published (Borges *et al.*, 2005; Ribeiro *et al.*, 2005; Ulrich *et al.*, 2010, 2015), including a number of large-scale total counts (i.e. not samples; e.g. the Barro Colorado Island 50 ha tree plot). Coupled with this increasing availability of data, it is hoped that, by reviewing a number of novel approaches, the present paper may act as a catalyst for a greater uptake and application of SADs in biogeography.

ACKNOWLEDGEMENTS

T.J.M. acknowledges funding from the Royal Geographical Society, the Sir Richard Stapley Trust and the Sidney Perry Foundation. Jian Yen kindly answered questions relating to the application of function regression models. The Azorean data are from project BALA, funded by Direcção Regional dos Recursos Florestais (Project: 17.01-080203, 1999–2004) and PO Açores 2016–2018 PROAACXXIs. P.A.V.B. was partly supported by project FCT-PTDC/BIA-BIC/5558/2014–MOMENTOS.

REFERENCES

- de Azevedo, E.B., Pereira, L.S. & Itier, B. (1999) Modelling the local climate in island environments: water balance applications. *Agricultural Water Management*, **40**, 393–403.
- Bartoń, K. (2012) *MuMIn: multi-model inference*. R package version 1.2.
- Borges, P.A.V., Aguiar, C., Amaral, J. *et al.* (2005) Ranking protected areas in the Azores using standardized sampling of soil epigeal arthropods. *Biodiversity and Conservation*, **14**, 2029–2060.
- Borges, P.A.V., Ugland, K.I., Dinis, F.O. & Gaspar, C. (2008). Insect and spider rarity in an oceanic island (Terceira, Azores): true rare and pseudo-rare species. *Insect Ecology and Conservation* (ed. by S. Fattorini), pp. 47–70. Research Signpost, Kerala, India.
- Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multi-model inference: a practical information-theoretic approach*, 2nd edn. Springer, New-York.
- Chust, G., Irigoien, X., Chave, J. & Harris, R.P. (2013) Latitudinal phytoplankton distribution and the neutral theory of biodiversity. *Global Ecology and Biogeography*, **22**, 531–543.
- Connor, E.F. & McCoy, E.D. (1979) Statistics and biology of the species–area relationship. *The American Naturalist*, **113**, 791–833.
- Dornelas, M., Soykan, C.U. & Ugland, K.I. (2011) Biodiversity and disturbance. *Biological diversity: frontiers in measurement and assessment* (ed. by A.E. Magurran and B.J. McGill), pp. 237–251. Oxford University Press, Oxford.
- Etienne, R.S. (2005) A new sampling formula for neutral biodiversity. *Ecology Letters*, **8**, 253–260.

- Fisher, R.A., Corbet, A.S. & Williams, C.B. (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, **12**, 42–58.
- Gaspar, C., Borges, P.A.V. & Gaston, K.J. (2008) Diversity and distribution of arthropods in native forests of the Azores archipelago. *Arquipélago Life and Marine Sciences*, **25**, 1–30.
- Gray, J.S., Waldichuk, M., Newton, A.J., Berry, R.J., Holden, A.V. & Pearson, T.H. (1979) Pollution-induced changes in populations. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, **286**, 545–561.
- Gray, J.S., Bjørgesæter, A. & Uglund, K.I. (2006) On plotting species abundance distributions. *Journal of Animal Ecology*, **75**, 752–756.
- Hijmans, R.J., Guarino, L., Jarvis, A., O'Brien, R. & Mathur, P. (2005) DIVA-GIS software. Version 5.2.0.2 (cited 31 November 2008). Available at: <http://www.diva-gis.org>.
- Hill, J.K. & Hamer, K.C. (1998) Using species abundance models as indicators of habitat disturbance in tropical forests. *Journal of Applied Ecology*, **35**, 458–460.
- Hubbell, S.P. (2001) *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton.
- Jabot, F. & Chave, J. (2011) Analyzing tropical forest tree species abundance distributions using a nonneutral model and through approximate Bayesian inference. *The American Naturalist*, **178**, E37–E47.
- Matthews, T.J. & Whittaker, R.J. (2014) Neutral theory and the species abundance distribution: recent developments and prospects for unifying niche and neutral perspectives. *Ecology and Evolution*, **4**, 2263–2277.
- Matthews, T.J. & Whittaker, R.J. (2015) On the species abundance distribution in applied ecology and biodiversity management. *Journal of Applied Ecology*, **52**, 443–454.
- Matthews, T.J., Borregaard, M.K., Uglund, K.I., Borges, P.A.V., Rigal, F., Cardoso, P. & Whittaker, R.J. (2014a) The gambin model provides a superior fit to species abundance distributions with a single free parameter: evidence, implementation and interpretation. *Ecography*, **37**, 1002–1011.
- Matthews, T.J., Borges, P.A.V. & Whittaker, R.J. (2014b) Multimodal species abundance distributions: a deconstruction approach reveals the processes behind the pattern. *Oikos*, **123**, 533–544.
- Matthews, T.J., Guilhaumon, F., Triantis, K.A., Borregaard, M.K. & Whittaker, R.J. (2015) On the form of species–area relationships in habitat islands and true islands. *Global Ecology and Biogeography*, **25**, 847–858.
- May, F., Huth, A. & Wiegand, T. (2015) Moving beyond abundance distributions: neutral theory and spatial patterns in a tropical forest. *Proceedings of the Royal Society B: Biological Sciences*, **282**, 20141657.
- McGill, B.J. (2003) A test of the unified neutral theory of biodiversity. *Nature*, **422**, 881–885.
- McGill, B.J. (2011) Species abundance distributions. *Biological diversity: frontiers in measurement and assessment* (ed. by A.E. Magurran and B.J. McGill), pp. 105–122. Oxford University Press, Oxford.
- McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas, M., Enquist, B.J., Green, J.L., He, F., Hurlbert, A.H., Magurran, A.E., Marquet, P.A., Maurer, B.A., Ostling, A., Soykan, C.U., Uglund, K.I. & White, E.P. (2007) Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, **10**, 995–1015.
- Müller, H.-G. & Stadtmüller, U. (2005) Generalized functional linear models. *The Annals of Statistics*, **33**, 774–805.
- Preston, F.W. (1948) The commonness, and rarity, of species. *Ecology*, **29**, 254–283.
- Ribeiro, S.P., Borges, P.A.V., Gaspar, C., Melo, C., Serrano, A.R.M., Amaral, J., Aguiar, C., André, G. & Quartau, J.A. (2005) Canopy insect herbivores in the Azorean Laurisilva forests: key host plant species in a highly generalist insect community. *Ecography*, **28**, 315–330.
- Sæther, B.-E., Engen, S. & Grøtan, V. (2013) Species diversity and community similarity in fluctuating environments: parametric approaches using species abundance distributions. *Journal of Animal Ecology*, **82**, 721–738.
- Triantis, K.A., Guilhaumon, F. & Whittaker, R.J. (2012) The island species–area relationship: biology and statistics. *Journal of Biogeography*, **39**, 215–231.
- Uglund, K.I., Lambshead, P.J.D., McGill, B., Gray, J.S., O'Dea, N., Ladle, R.J. & Whittaker, R.J. (2007) Modelling dimensionality in species abundance distributions: description and evaluation of the Gambin model. *Evolutionary Ecology Research*, **9**, 313–324.
- Ulrich, W., Ollik, M. & Uglund, K.I. (2010) A meta-analysis of species–abundance distributions. *Oikos*, **119**, 1149–1155.
- Ulrich, W., Kusumoto, B., Shiono, T. & Kubota, Y. (2015) Climatic and geographical correlates of global forest tree species abundance distributions and community evenness. *Journal of Vegetation Science*, **27**, 295–305.
- Yen, J.D.L., Thomson, J.R., Paganin, D.M., Keith, J.M. & MacNally, R. (2015) Function regression in ecology and evolution: FREE. *Methods in Ecology and Evolution*, **6**, 17–26.

BIOSKETCH

Tom Matthews is a research fellow at the University of Birmingham. He is interested in the study of macroecological patterns in habitat islands, and the application of island theory to fragmented landscapes.

Author contributions: T.J.M. and R.J.W. conceived the ideas; P.A.V.B. and E.B.deA. collected the data; T.J.M. ran the analyses; and T.J.M. and R.J.W. wrote the paper with a significant contribution from P.A.V.B. and E.B.deA. All authors discussed the results and commented on the manuscript.

Editor: Daniel Chapman