

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

Matthews, Thomas; de Azvedo, Eduardo; Borges, Paulo; Whittaker, Robert

DOI:  
[10.1111/jbi.13008](https://doi.org/10.1111/jbi.13008)

License:  
None: All rights reserved

Document Version  
Peer reviewed version

Citation for published version (Harvard):  
Matthews, T, de Azvedo, E, Borges, P & Whittaker, R 2017, 'A biogeographical perspective on species abundance distributions: recent advances and opportunities for future research', *Journal of Biogeography*, vol. 44, no. 8, pp. 1705–1710. <https://doi.org/10.1111/jbi.13008>

[Link to publication on Research at Birmingham portal](#)

## Publisher Rights Statement:

This is the peer reviewed version of the following article: Matthews, T. J., Borges, P. A.V., de Azevedo, E. B. and Whittaker, R. J. (2017), A biogeographical perspective on species abundance distributions: recent advances and opportunities for future research. *Journal of Biogeography*, 44: 1705–1710., which has been published in final form at <http://dx.doi.org/10.1111/jbi.13008>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

Eligibility for repository: Checked on 5/5/2017

## General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

## Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

1 **Submission to: Journal of Biogeography**

2 **Article Type: Perspective**

3

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

6

7 Thomas J. Matthews<sup>1, 2</sup>, Paulo A.V. Borges<sup>2</sup>, Eduardo Brito de Azevedo<sup>3</sup>, and Robert J.  
8 Whittaker<sup>4, 5</sup>

9 <sup>1</sup>GEES (School of Geography, Earth and Environmental Sciences), The University of  
10 Birmingham, Birmingham, B15 2TT

11 <sup>2</sup>CE3C – Centre for Ecology, Evolution and Environmental Changes / Azorean Biodiversity  
12 Group and Universidade dos Açores –Departamento de Ciências e Engenharia do Ambiente,  
13 9700-042, Angra do Heroísmo, Açores, Portugal

14 <sup>3</sup>CMMG (Centro de Estudos do Clima, Meteorologia e Mudanças Globais) CITA-A,  
15 Universidade dos Açores, Dep. de Ciências Agrárias, 9700-042 Angra do Heroísmo, Portugal

16 <sup>4</sup>Conservation Biogeography and Macroecology Programme, School of Geography and the  
17 Environment, University of Oxford, South Parks Road, Oxford, OX1 3QY, UK

18 <sup>5</sup>Center for Macroecology, Evolution and Climate, Department of Biology, University of  
19 Copenhagen, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark

20

21 \*Correspondence: Thomas J. Matthews, School of Geography, Earth and Environmental  
22 Sciences, University of Birmingham, Birmingham, B15 2TT, UK

23 Email: t.j.matthews@bham.ac.uk

24

25 **Running header: Biogeography and species abundance distributions**

26

27 **Word count:** abstract (185 words); main text (2956); whole paper (4148); 2 Tables; 2  
28 Figures

29

30

31

## 32 **ABSTRACT**

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

## 47 **Keywords**

48 applied biogeography, function regression, gambin model, lognormal model, logseries model,  
49 species abundance distribution

50

## 51 **INTRODUCTION**

52 A species abundance distribution (herein ‘SAD’) describes how the total number of  
53 individuals sampled within a given community is distributed amongst the sampled species.  
54 [Patterns of SADs have](#) been studied for over 70 years, with seminal papers published in the  
55 1940s by Fisher *et al.* (1943) and Preston (1948), [but have](#) received renewed interest in the  
56 last decade (McGill *et al.*, 2007; McGill, 2011; Matthews & Whittaker, 2015). One common  
57 recent use of the SAD has been as a means to test different ecological theories. For example,  
58 following the introduction of Hubbell’s (2001) spatially implicit neutral model (SINM) for  
59 application in biogeography and biodiversity studies, many authors attempted to test the  
60 theory by evaluating the fit of the SAD predicted by [Hubbell’s SINM to empirical data](#) –

61 with varying levels of success (e.g. Hubbell, 2001; McGill, 2003; Etienne, 2005; for a review  
62 see Matthews & Whittaker, 2014).

63 Unfortunately, it has become increasingly apparent that multiple processes can generate  
64 similar shapes of SAD curves, thus causing a problem of equifinality: the fit of a given SAD  
65 model cannot unambiguously provide evidence in support of a given theory (McGill *et al.*,  
66 2007). An alternative and less frequently applied approach to that of simply comparing the fit  
67 of different SAD models to data from a single site, is to collect species abundance data from a  
68 variety of sites to build models that can then be used to analyse how different SAD properties  
69 (e.g. form, skewness) vary with different predictor (environmental) variables. In other words,  
70 the question of which mechanisms drive SADs can be approached instead by assessing which  
71 biogeographical variables control the relative abundances of species. Such an approach to  
72 SAD research may be more revealing and encourage interest in using SADs in  
73 biogeographical studies.

74 Here we set out to highlight the use of SADs as a fruitful and instructive approach in  
75 biogeographical research and we hope that this article may encourage greater interest in  
76 analysing SADs amongst biogeographers. We begin by describing how two different  
77 regression-based methods (standard linear regression and function regression) can be used to  
78 determine the mechanisms underlying SADs and provide examples from the recent literature  
79 of studies that have used these approaches. This is followed by an analysis of arthropod data  
80 from the Azorean archipelago to illustrate how the study of SADs can generate interesting  
81 information from a biogeographical point of view.

## 82 **USE OF REGRESSION METHODS FOR EXPLAINING VARIATION IN SADS**

### 83 **Assessing SAD model parameters within a standard linear regression framework**

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

90 Whilst the lognormal and logseries models are sound statistical representations of many  
91 empirical communities and still represent useful tools, more recently introduced models offer

92 additional possibilities for such analyses. For example, the immigration parameter ( $m$ ) of  
93 Hubbell's (2001) SINM can be used to make inferences regarding the importance of dispersal  
94 from the metacommunity into the local community: if  $m$  is close to 1, the local community is  
95 a random sample of the regional metacommunity, whereas if  $m$  is close to zero, the local  
96 community receives very few immigrants from the metacommunity (Hubbell, 2001;  
97 Matthews & Whittaker, 2014). The  $m$  parameter can easily be estimated by maximum  
98 likelihood fitting of the zero-sum multinomial SAD model (Etienne, 2005). If multiple sites  
99 are surveyed it is then possible to regress  $m$  against various predictor variables to determine  
100 what factors influence the importance of dispersal in ecological communities. For example, in  
101 one recent analysis Chust *et al.* (2013) found that  $m$  was consistently lower in tropical relative  
102 to temperate marine phytoplankton communities. [The frequency dependence parameter](#)  
103 [\(delta\) of Jabot & Chave's \(2011\) non-neutral generalisation of Hubbell's model provides an](#)  
104 [alternative parameter to use in such exercises \(e.g. see Jabot & Chave's 2011, Figure 2 for a](#)  
105 [correlation between delta and precipitation for a variety of tropical tree plots\).](#)

106 The gambin SAD model ([Matthews \*et al.\*, 2014a](#)) can also be used in biogeographical  
107 analyses (e.g. Dornelas *et al.*, 2011). The gambin model combines the flexible gamma  
108 distribution with a binomial sampling process. It is a single free parameter model [and its](#)  
109 [parameter \( \$\alpha\$ \) characterizes the shape of the SAD.](#) Low values characterize logseries SAD  
110 shapes, while higher values indicate lognormal curve shapes (Ugland *et al.*, 2007; Matthews  
111 *et al.*, 2014a). Extreme values can indicate more complex situations in which common  
112 species are the most prevalent type of species in the community. Thus,  $\alpha$  condenses the shape  
113 of the SAD into a single value that can then be used in regression models. The gambin model  
114 is beneficial in this regard as it is flexible and has been found to fit a wide range of SAD  
115 shapes.

116 A recent paper published in *Journal of Vegetation Science* (Ulrich *et al.*, 2015) [provides a](#)  
117 [useful example of what can be gained from a biogeographical approach to SAD research.](#)  
118 Ulrich *et al.* (2015) examined the SADs of 605 tree assemblages across six continents. Ulrich  
119 *et al.* fitted two SAD models (logseries and lognormal) in rank-abundance form and related  
120 various properties of the observed SADs (e.g. shape and evenness) to geographical and  
121 climatic variables (e.g. latitude, elevation and evapotranspiration) using linear regression. A  
122 relationship between latitude and SAD evenness and shape was found. Logseries distributions  
123 were more prevalent at lower latitudes, whilst there was an increase in the prevalence of  
124 lognormal distributions towards northern latitudes. Again, whilst these results are interesting

125 in themselves, what is perhaps more intriguing is that this approach allows the user to make  
126 inferences regarding classic SAD and community assembly theory. For example, the  
127 lognormal distribution has traditionally been used to model undisturbed ecological  
128 communities (Ugland *et al.*, 2007; Matthews & Whittaker, 2015). In contrast, logseries SADs  
129 have been linked to disturbed communities and communities in severe environments with low  
130 productivity (Gray *et al.*, 1979; Hill & Hamer, 1998; Ugland *et al.*, 2007). As such, a number  
131 of studies have used deviation from a lognormal distribution as a means of assessing the  
132 impact of disturbance (e.g. pollution) on natural communities (e.g. Hill & Hamer, 1998;  
133 reviewed in Matthews & Whittaker, 2015). The results of Ulrich *et al.* (2015) are at variance  
134 with this theory, as it was found that logseries distributions were associated with species rich,  
135 productive and low-latitude tree communities. Instead, their results point to the importance of  
136 dispersal and stochastic processes in shaping the SADs of tree communities. It is worth  
137 noting that the  $R^2$  values of the models in Ulrich *et al.* were generally quite low ( $< 0.3$ ),  
138 indicating that there must be other important "hidden variables" that were not included in the  
139 study, thus pointing towards information needs for future biogeographical SAD studies and  
140 meta-analyses. For example, isolation is known to be an important variable in many  
141 ecological systems and may therefore explain some of the additional variation in Ulrich *et*  
142 *al.*'s data. A distance-decay approach may be useful in such circumstances to evaluate the  
143 impact of between patch distances on SAD metrics.

#### 144 **Function regression and the species abundance distribution**

145 A recent paper by Yen *et al.* (2015) introduces the method of function regression in the  
146 context of ecological applications. Function based regression models have been used in the  
147 statistical literature for some time (see Müller & Stadtmüller, 2005), but as Yen *et al.* note,  
148 applications in biogeography and ecology are uncommon. Function regression models enable  
149 functions to be used as the response variable (and as predictor variables, although this  
150 possibility is not discussed here), which allows the user to regress a function on different  
151 predictor variables (Yen *et al.*, 2015). Questions such as 'how does the form of the SAD  
152 change in response to fragment area and/or isolation?' are perfectly suited to this  
153 methodological approach. However, instead of condensing the SAD into a single value, for  
154 example the gambin  $\alpha$  parameter, function regression models allow us to use function-valued  
155 data and therefore the whole SAD, as the response variable. Multiple methods are available to  
156 fit function regression models (e.g. Bayesian estimation using reversible-jump Markov chain  
157 Monte Carlo computation). The original 'FREE' R package (see Yen *et al.*, 2015) provides

158 functionality for six of these different methods. An updated version of the package  
159 ('FREElite') that includes fewer fitting methods is also available. The use of function  
160 regression in SAD research is technically more complex than the simple linear models  
161 discussed above, but the former are arguably more revealing models as they allow the user to  
162 determine how different parts of the SAD are affected by a given predictor, e.g. the relative  
163 abundances of very common species. As such, "a function- valued method is likely to provide  
164 much deeper ecological and evolutionary insight" (Yen *et al.*, 2015, p. 18). In their paper,  
165 Yen *et al.* provided examples of function regression using individual size distributions of  
166 various taxa.

## 167 **EXEMPLIFICATION WITH AZOREAN ARTHOPOD SADS**

### 168 **Using standard linear regression to examine variation in gambin's alpha parameter**

169 [To highlight the utility of using SAD model parameters within a linear regression framework](#)  
170 [we focused on the  \$\alpha\$  parameter of the gambin model \(described above\).](#) We used well  
171 specified arthropod SAD data from the long-running Biodiversity of Arthropods from the  
172 Laurisilva of Azores (BALA) project in the Azores (see Borges *et al.*, 2005; Ribeiro *et al.*,  
173 2005) in combination with a linear regression modelling framework to determine which  
174 environmental variables explained variance in the shape of the SAD. [Arthropods were](#)  
175 [sampled using a combination of pitfall traps and a canopy beating method following a](#)  
176 [rigorously standardized protocol.](#) [Eighteen fragments of native Laurisilva forest were](#)  
177 [sampled across seven islands in the Azorean archipelago over 13 years.](#) The full sampling  
178 methodology can be found in Gaspar *et al.* (2008). The gambin model was fitted to the SADs  
179 from these 18 fragments using the 'gambin' R package ([Matthews \*et al.\*, 2014a](#)), [recording](#)  
180 [the  \$\alpha\$  parameter in each case.](#) As differences in sample size have been found to influence  $\alpha$   
181 ([Matthews \*et al.\*, 2014a](#)), we used a re-sampling approach in which we first determined the  
182 fragment with the smallest number of individuals, denoting the number of individuals in this  
183 fragment as  $n$ . For each of the remaining 17 fragments, we then randomly sampled  $n$   
184 individuals and fit the gambin model to the sampled data. This process was repeated 100  
185 times in each case and the average  $\alpha$  value of 100 iterations taken. The response variable  
186 therefore constituted the standardised  $\alpha$  values for 18 forest fragments. We collected data on  
187 five predictor variables, representing different fragment characteristics: fragment area,  
188 fragment isolation, precipitation, temperature and relative humidity (RH). Climatic data were  
189 obtained from the CIELO model ([Azevedo \*et al.\*, 1999](#)). Fragment area and isolation were

190 obtained from Gaspar *et al.* (2008) and were calculated using a geographic matrix of  
191 centroids using the DIVA-GIS software (Hijmans *et al.*, 2005). Isolation was measured in  
192 meters as the distance between fragments within an island, except in the case of the single  
193 fragment on the island of Santa Maria, for which we used the distance to closest fragment on  
194 the nearest island of São Miguel. All predictors were log-transformed (base e), which induced  
195 normality. Variance inflation factors were used to assess multi-collinearity between  
196 predictors, while Cook's distance was used to identify any outliers. Two data points were  
197 removed as outliers, so that our results are for 16 fragments only. The dredge function in the  
198 'MuMIn' R package (Bartoń, 2012) was used to fit a complete set of models that were  
199 compared using Akaike's information criterion corrected for small sample size (AIC<sub>c</sub>;  
200 Burnham & Anderson, 2002). Weight of evidence values were calculated for each predictor  
201 as the sum of the AIC<sub>c</sub> weights from all models in which a predictor was included (Burnham  
202 & Anderson, 2002).

203         The best model contained isolation and temperature, and both of these variables had  
204 relatively high WoE values (Table 1). The best model had an adjusted R<sup>2</sup> of 0.74, which is  
205 noteworthy, as it indicates that a substantial part of the variation in  $\alpha$  between sites can be  
206 attributed to isolation and temperature alone, for these 16 fragments. Isolation was included  
207 in all models within 10  $\Delta$ AIC<sub>c</sub> of the best model, and had a near maximum WoE value (0.99).  
208 The effect of isolation was negative (see Fig. 1), meaning that increasing isolation results in a  
209 decrease in  $\alpha$  and therefore a shift from a more lognormal-like SAD towards a more  
210 logseries-shaped distribution. This is an interesting finding because it is could be expected  
211 that more isolated fragments/islands will have fewer really rare species of forest-dependent  
212 arthropods due to reduced rescue effects and re-colonisations following patch level  
213 extinctions. Thus, based on this line of reasoning, more isolated fragments would be expected  
214 to have a lower proportion of rarer species and therefore have SADs closer to lognormal in  
215 form. One possible ecological interpretation of this result is that, as these fragments are  
216 isolated in a matrix of human-modified habitats, our samples contain a relatively high  
217 proportion of tourist species represented by few individuals (Borges *et al.*, 2008), and that the  
218 proportion of tourist species is higher in the more isolated fragments. For example, in our  
219 previous work we found that the presence of non-forest specialist species within fragments  
220 affects the shape of the SAD (Matthews *et al.*, 2014b). Nonetheless, the most remarkable  
221 point in the context of the current paper is how this methodological approach can be used to



222 determine the variables driving variation in SADs between sites, and in turn provides useful  
223 information on the key processes underpinning community structure.

#### 224 **Examining variation in the shape Azorean arthropod SADs using function regression**

225 To our knowledge function regression has not been used in conjunction with SADs and so we  
226 return to the Azorean arthropod data described above to illustrate the approach. For this  
227 analysis, we used the SAD data from the 18 forest fragments as the response variable, but this  
228 time in their raw form. We used two predictor variables: fragment area and isolation.

229 Following Yen *et al.* (2015), both predictors were standardized to have a mean of zero and a  
230 standard deviation of one. The function regression model was fitted using the FREE R  
231 package, and we used the ‘INLA’ (Integrated nested Laplace approximation) method as it  
232 was found to perform well in Yen *et al.*’s various tests. A Gaussian error structure was  
233 assumed. To convert the SAD data into a matrix, we binned the data from each fragment into  
234 octaves using functionality available in the gambin R package (Matthews *et al.*, 2014a). This  
235 procedure uses a simple log<sub>2</sub> transformation that doubles the number of abundance classes  
236 within each octave (see method 3 in Gray *et al.*, 2006). A matrix was then created in which  
237 the columns represented abundance octaves, and the rows represented fragments.

238 The R<sup>2</sup> value of the resultant model was high (0.87), but the credible intervals around  
239 the model coefficient estimates were quite large (e.g. Fig. 2). In the function regression  
240 analysis, isolation had a relatively large positive effect on the number of species in the rarer  
241 octaves (Table 2 and Fig. 2), which in turn will make the SAD more logseries-like. The  
242 results of this analysis match up well with our analysis above in which we used the gambin  $\alpha$   
243 parameter as the response variable in a standard linear regression model. However, the use of  
244 the function regression model allows us to make additional observations, which we were  
245 unable to make using the simple linear model. For example, from Fig. 2 it is clear that the  
246 effect of isolation is most pronounced for the two rarest octaves (1:2), and the effect  
247 decreases towards the more common octaves. Thus, it does appear that isolation is resulting  
248 in a relatively higher proportion of rare, possibly tourist (see e.g. Borges *et al.*, 2008), species  
249 in isolated fragments. A single parameter value (e.g.  $\alpha$ ) will never be able to convey this  
250 detailed level of information; which is one reason why function regression represents a useful  
251 tool for SAD studies.

252

## 253 CONCLUSIONS

254 Many authors have commented on how simply evaluating the fit of a predicted SAD is a poor  
255 test of any ecological theory (e.g. McGill, 2003; McGill *et al.*, 2007; Matthews & Whittaker,  
256 2014; May *et al.*, 2015). Perhaps then a more fruitful avenue for SAD research is to adopt a  
257 more biogeographic perspective, and examine the factors that underpin the observed variation  
258 in SAD form between sites in both space and time. A large number of biogeographical  
259 studies have attempted to interpret the parameters of the power species–area relationship  
260 model ( $c$  and  $z$ ) ecologically by assessing which predictor variables explain variation in the  
261 parameters across datasets (Connor & McCoy, 1979; Triantis *et al.*, 2012; Matthews *et al.*,  
262 2015). Similar biogeographic analyses involving parameters of SAD models are much less  
263 prevalent in the biogeography literature, probably due in part to the additional data  
264 requirements involved in constructing SADs. In fact, obtaining standardized abundance  
265 values for many species across large scales is not trivial, particularly for invertebrates.  
266 However, many SAD datasets have now been published (Borges *et al.*, 2005; Ribeiro *et al.*,  
267 2005; Ulrich *et al.*, 2010, 2015), including a number of large-scale total counts (i.e. not  
268 samples; e.g. the Barro Colorado Island 50ha tree plot). Coupled with this increasing  
269 availability of data, it is hoped that, by reviewing a number of novel approaches, the present  
270 paper may act as a catalyst for a greater uptake and application of SADs in biogeography.

## 271 ACKNOWLEDGEMENTS

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

278

## 279 REFERENCES

- 280 de Azevedo, E.B., Pereira, L.S. & Itier, B. (1999) Modelling the local climate in island  
281 environments: water balance applications. *Agricultural Water Management*, **40**, 393-  
282 403.
- 283 Bartoń, K. (2012) *MuMIn: multi-model inference*. R package version 1.2.

- 284 Borges, P.A.V., Aguiar, C., Amaral, J., Amorim, I.R., André, G., Arraiol, A., A., B., Dinis,  
285 F., Enghoff, H., Gaspar, C., Ilharco, F., Mahnert, V., Melo, C., Pereira, F., Quartau,  
286 J.A., Ribeiro, S., Ribes, J., Serrano, A.R.M., Sousa, A.B., Strassen, R.Z., Vieira, L.,  
287 Vieira, V., Vitorino, A. & Wunderlich, J. (2005) Ranking protected areas in the  
288 Azores using standardized sampling of soil epigeal arthropods. *Biodiversity and*  
289 *Conservation*, **14**, 2029-2060.
- 290 Borges, P.A.V., Ugland, K.I, Dinis, F.O. & Gaspar, C. (2008). Insect and spider rarity in an  
291 oceanic island (Terceira, Azores): true rare and pseudo-rare species. *Insect Ecology*  
292 *and Conservation* (ed. by S. Fattorini), pp. 47-70. Research Signpost, Kerala, India.
- 293 Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multi-model inference: a*  
294 *practical information-theoretic approach*, 2nd edn. Springer, New-York.
- 295 Connor, E.F. & McCoy, E.D. (1979) Statistics and biology of the species-area relationship.  
296 *American Naturalist*, **113**, 791-833.
- 297 Dornelas, M., Soykan, C.U. & Ugland, K.I. (2011) Biodiversity and disturbance. *Biological*  
298 *diversity: frontiers in measurement and assessment* (ed. by A.E. Magurran and B.J.  
299 McGill), pp. 237-251. Oxford University Press, Oxford.
- 300 Etienne, R.S. (2005) A new sampling formula for neutral biodiversity. *Ecology Letters*, **8**,  
301 253-260.
- 302 Fisher, R.A., Corbet, A.S. & Williams, C.B. (1943) The relation between the number of  
303 species and the number of individuals in a random sample of an animal population.  
304 *Journal of Animal Ecology*, **12**, 42-58.
- 305 Gaspar, C., Borges, P.A.V. & Gaston, K.J. (2008) Diversity and distribution of arthropods in  
306 native forests of the Azores archipelago. *Arquipélago Life and Marine Sciences*, **25**,  
307 1-30.
- 308 Gray, J.S., Waldichuk, M., Newton, A.J., Berry, R.J., Holden, A.V. & Pearson, T.H. (1979)  
309 Pollution-induced changes in populations. *Philosophical Transactions of the Royal*  
310 *Society of London. B, Biological Sciences*, **286**, 545-561.
- 311 Gray, J.S., Bjørgesæter, A. & Ugland, K.I. (2006) On plotting species abundance  
312 distributions. *Journal of Animal Ecology*, **75**, 752-756.
- 313 Hill, J.K. & Hamer, K.C. (1998) Using species abundance models as indicators of habitat  
314 disturbance in tropical forests. *Journal of Applied Ecology*, **35**, 458-460.
- 315 Hijmans, R.J., Guarino, L., Jarvis, A., O'Brien, R. & Mathur, P. (2005) DIVA-GIS software.  
316 Version 5.2.0.2. (cited 31 November 2008). Available from: <http://www.diva-gis.org>.
- 317 Hubbell, S.P. (2001) *The unified neutral theory of biodiversity and biogeography*. Princeton  
318 University Press, Princeton.

- 319 MacArthur, R.H. & Wilson, E.O. (1967) *The theory of island biogeography*. Princeton  
320 University Press, Princeton.
- 321 Matthews, T.J. & Whittaker, R.J. (2014) Neutral theory and the species abundance  
322 distribution: recent developments and prospects for unifying niche and neutral  
323 perspectives. *Ecology and Evolution*, **4**, 2263–2277.
- 324 Matthews, T.J. & Whittaker, R.J. (2015) On the species abundance distribution in applied  
325 ecology and biodiversity management. *Journal of Applied Ecology*, **52**, 443–454.
- 326 Matthews, T.J., Borregaard, M.K., Ugland, K.I., Borges, P.A.V., Rigal, F., Cardoso, P. &  
327 Whittaker, R.J. (2014a) The gambin model provides a superior fit to species  
328 abundance distributions with a single free parameter: evidence, implementation and  
329 interpretation. *Ecography*, **37**, 1002–1011.
- 330 Matthews, T.J., Borges, P.A.V. & Whittaker, R.J. (2014b) Multimodal species abundance  
331 distributions: a deconstruction approach reveals the processes behind the pattern.  
332 *Oikos*, **123**, 533–544.
- 333 Matthews, T.J., Guilhaumon, F., Triantis, K.A., Borregaard, M.K. & Whittaker, R.J. (2015)  
334 On the form of species–area relationships in habitat islands and true islands. *Global  
335 Ecology and Biogeography*, **25**, 847–858.
- 336 May, F., Huth, A. & Wiegand, T. (2015) Moving beyond abundance distributions: neutral  
337 theory and spatial patterns in a tropical forest. *Proceedings of the Royal Society B:  
338 Biological Sciences*, **282**, 20141657.
- 339 McGill, B.J. (2003) A test of the unified neutral theory of biodiversity. *Nature*, **422**, 881–885.
- 340 McGill, B.J. (2011) Species abundance distributions. *Biological diversity: frontiers in  
341 measurement and assessment* (ed. by A.E. Magurran and B.J. McGill), pp. 105–122.  
342 Oxford University Press, Oxford.
- 343 McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas,  
344 M., Enquist, B.J., Green, J.L., He, F., Hurlbert, A.H., Magurran, A.E., Marquet, P.A.,  
345 Maurer, B.A., Ostling, A., Soykan, C.U., Ugland, K.I. & White, E.P. (2007) Species  
346 abundance distributions: moving beyond single prediction theories to integration  
347 within an ecological framework. *Ecology Letters*, **10**, 995–1015.
- 348 Müller, H.-G. & Stadtmüller, U. (2005) Generalized functional linear models. *The Annals of  
349 Statistics*, **33**, 774–805.
- 350 Preston, F.W. (1948) The commonness, and rarity, of species. *Ecology*, **29**, 254–283.
- 351 Preston, F.W. (1962) The canonical distribution of commonness and rarity: part II. *Ecology*,  
352 **43**, pp. 410–432.

- 353 Ribeiro, S.P., Borges, P.A.V., Gaspar, C., Melo, C., Serrano, A.R.M., Amaral, J., Aguiar, C.,  
354 André, G. & Quartau, J.A. (2005) Canopy insect herbivores in the Azorean Laurisilva  
355 forests: key host plant species in a highly generalist insect community. *Ecography*,  
356 **28**, 315-330.
- 357 Sæther, B.-E., Engen, S. & Grøtan, V. (2013) Species diversity and community similarity in  
358 fluctuating environments: parametric approaches using species abundance  
359 distributions. *Journal of Animal Ecology*, **82**, 721-738.
- 360 Triantis, K.A., Guilhaumon, F. & Whittaker, R.J. (2012) The island species–area relationship:  
361 biology and statistics. *Journal of Biogeography*, **39**, 215-231.
- 362 Ugland, K.I., Lamshead, P.J.D., McGill, B., Gray, J.S., O’Dea, N., Ladle, R.J. & Whittaker,  
363 R.J. (2007) Modelling dimensionality in species abundance distributions: description  
364 and evaluation of the Gambin model. *Evolutionary Ecology Research*, **9**, 313-324.
- 365 Ulrich, W., Ollik, M. & Ugland, K.I. (2010) A meta-analysis of species–abundance  
366 distributions. *Oikos*, **119**, 1149-1155.
- 367 Ulrich, W., Kusumoto, B., Shiono, T. & Kubota, Y. (2015) Climatic and geographical  
368 correlates of global forest tree species abundance distributions and community  
369 evenness. *Journal of Vegetation Science*. **27**, 295-305.
- 370 Yen, J.D.L., Thomson, J.R., Paganin, D.M., Keith, J.M. & Mac Nally, R. (2015) Function  
371 regression in ecology and evolution: FREE. *Methods in Ecology and Evolution*, **6**, 17-  
372 26.

373

374

## 375 **BIOSKETCHES**

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

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

383 Editor: Daniel Chapman

384 **TABLES**

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

Model	Area	Isolation	Precip.	RH	Temp.	$\Delta AIC_c$	$wAIC_c$
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		

392

393

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

	Octave											
	1	2	3	4	5	6	7	8	9	10	11	12
Intercept	26.5	18.3	13	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

399

400

401

402

403

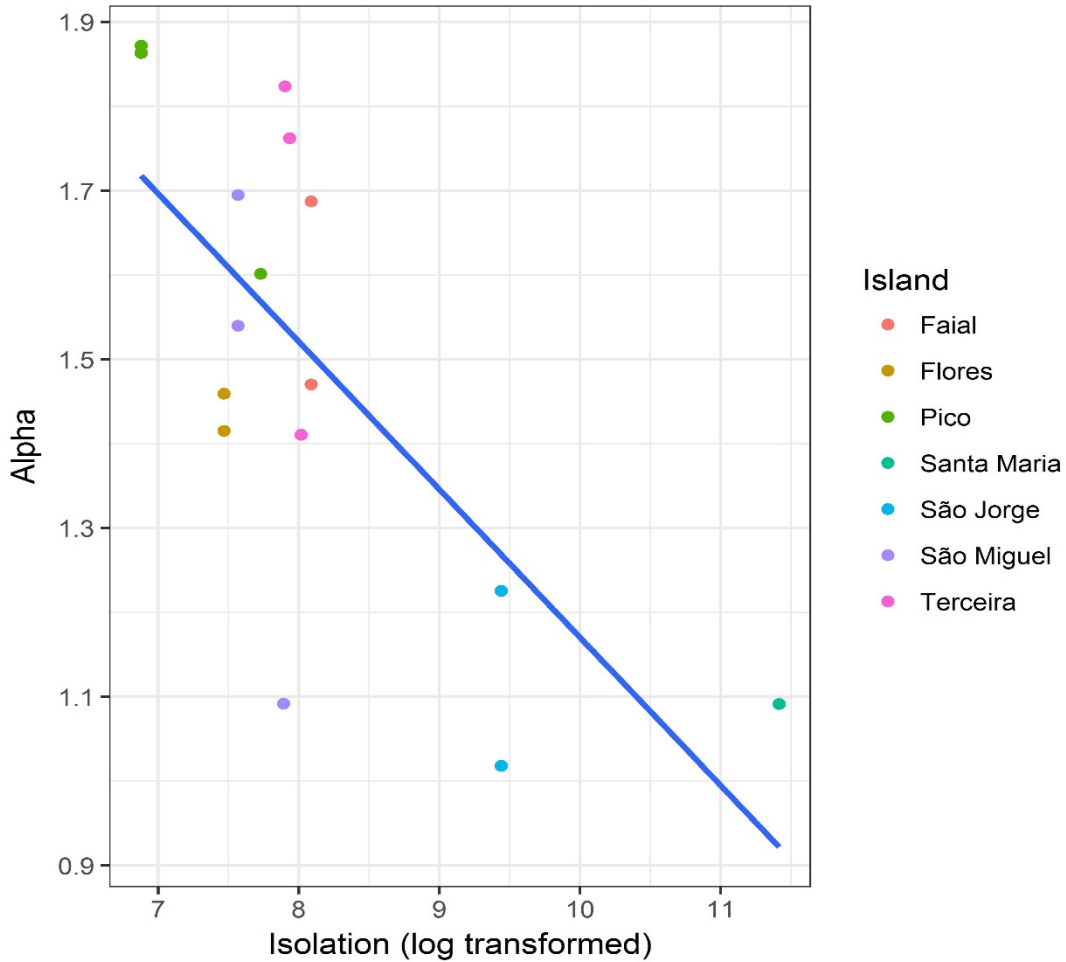
404

405

406

407

408



425 **Figure 1** The relationship between fragment isolation and the standardized alpha parameter  
 426 value of the gambin species abundance distribution model. The data are 16 arthropod SADs  
 427 from native Laurisilva fragments in the Azores. Fragment isolation ranged from 970 m to  
 428 90780 m. The blue line represents the best fit linear regression model. The islands in which  
 429 the fragments are located are provided in the key.

430

431

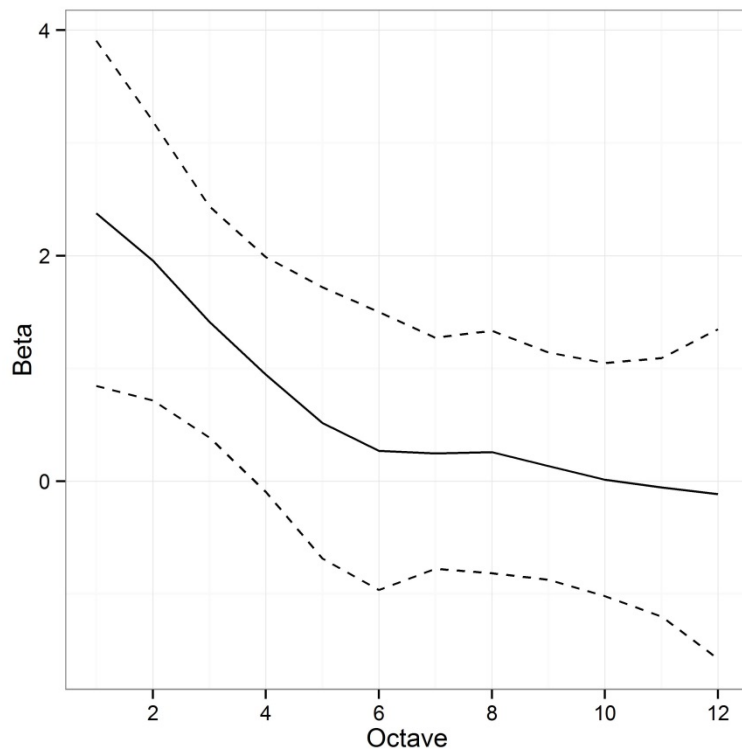
432

433

434

435

436



450 **Figure 2** Fitted parameter estimates from a function regression model. The fitted curve  
 451 indicates the effect (Beta) of the predictor variable (fragment isolation) on the species  
 452 abundance distributions of arthropods in 18 native Laurisilva forest fragments, in the Azores.  
 453 A higher value of Beta indicates a greater effect of isolation on that particular area of the  
 454 SAD (see the main text for further information). The solid line represents the mean value, and  
 455 the dashed lines represent the approximate 95% pointwise credible intervals. The SAD data  
 456 were binned into octaves procedure using a simple log2 transformation: octave 1 contains the  
 457 number of species with 1 individual, octave 2 the number of species with 2 or 3 individuals,  
 458 octave 3 the number of species with 4 to 7 individuals, and so forth.

459

460

461

462