

1 Title:

2 **Prevalence of multimodal species abundance distributions is linked to spatial and**
3 **taxonomic breadth**

4

5 Laura Henriques Antão ^{1,2*}, Sean Connolly ³, Anne Magurran ¹, Amadeu Soares ² & Maria Dornelas ¹

6

7 ¹ Centre for Biological Diversity, University of St Andrews. Sir Harold Mitchell Building, Greenside
8 Place, KY16 9TF, St Andrews, United Kingdom.

9 ² Department of Biology & CESAM, Universidade de Aveiro. Campus Universitário de Santiago,
10 3810-193 Aveiro, Portugal.

11 ³ School of Marine and Tropical Biology, James Cook University. Townsville, QLD 4811, Australia.

12

13 Emails:

14 Laura Henriques Antão – liha@st-andrews.ac.uk

15 Sean Connolly – sean.connolly@jcu.edu.au

16 Anne Magurran – aem1@st-andrews.ac.uk

17 Amadeu Soares – asoares@ua.pt

18 Maria Dornelas – maadd@st-andrews.ac.uk

19

20 The authors declare no conflict of interest.

21

22 Keywords:

23 Community structure; diversity patterns; Poisson Lognormal mixture; logseries; lognormal; spatial
24 scale; taxonomic breadth

25

26 Short running-title:

27 Multimodal SADs linked to spatial and taxonomic breadth

28

29 Corresponding author:

30 * Laura Henriques Antão – liha@st-andrews.ac.uk

31

32 Number of words in the abstract: 240

33 Number of words in the main body of text: 4.829

34

35 Number of general references: 42

36 Number of data-source references: 123

37

38 4 figures; 1 table

39

40 Supporting Information: 3 appendices

41 **Abstract**

42

43 **Aim:** Species abundance distributions (SADs) are a synthetic measure of biodiversity and community
44 structure. Although typically described by unimodal logseries or lognormal distributions, empirical
45 SADs can also exhibit multiple modes. However, we do not know how prevalent multimodality is,
46 nor do we have an understanding of the factors leading to this pattern. Here we quantify the
47 prevalence of multimodality in SADs across a wide range of taxa, habitats and spatial extents.

48

49 **Location:** Global.

50

51 **Methods:** We used AIC_c and Likelihood Ratio tests (LRT) to test whether distributions with more
52 than one mode accurately describe the abundance distributions of the underlying communities. We
53 analysed 117 empirical datasets from intensely sampled communities, including taxa ranging from
54 birds, plants, fish and invertebrates, from terrestrial, marine and freshwater habitats.

55

56 **Results:** We find evidence for multimodality in 14.5% of the SADs when using both AIC_c and LRT.
57 This is a conservative estimate, as AIC_c alone estimates a prevalence of multimodality of 22%. We
58 additionally show that the pattern is more common in data encompassing broader spatial scales and
59 greater taxonomic breadth, suggesting that multimodality increases with ecological heterogeneity.

60

61 **Main conclusions:** We suggest that higher levels of ecological heterogeneity, underpinned by larger
62 spatial extent and higher taxonomic breadth, can yield multimodal SADs. Our analysis shows that
63 multimodality occurs with a prevalence that warrants its systematic consideration when assessing
64 SAD shape and emphasizes the need for macroecological theories to include multimodality in the
65 range of SADs they predict.

66 **Introduction**

67

68 Species Abundance Distributions (SADs) depict the relative abundance of the species present in a
69 community and describe one of the most fundamental patterns of species diversity – most
70 communities contain many rare and only a few common species (McGill *et al.*, 2007). Empirical
71 datasets consistently produce species abundance distributions that are quasi-hyperbolic on an
72 arithmetic scale – the ubiquitous ‘hollow curve’. However, on a logarithmic scale of abundance,
73 SADs exhibit more variability, with species abundance distributions alternately exhibiting no internal
74 mode - most species occur at the lowest abundance class (i.e. as singletons), one internal mode, or
75 multiple internal modes. Despite seven decades of study and dozens of different models proposed
76 (McGill *et al.*, 2007), there is still no consensus about what drives variation in SADs shape, nor how it
77 might be connected to factors structuring ecological communities (Fisher *et al.*, 1943; Preston, 1948;
78 Magurran & Henderson, 2003; McGill, 2003b; Green & Plotkin, 2007; Dornelas *et al.*, 2009). The
79 extent to which current biodiversity theories are able to accommodate and explain such variation is a
80 critical criterion to their evaluation and application (McGill *et al.*, 2007).

81

82 The two distributions recurrently proposed to describe SADs are the logseries (Fisher *et al.*, 1943) and
83 the lognormal (Preston, 1948) (Fig. 1). While many intensely sampled communities seem to follow a
84 lognormal distribution (Magurran, 2004), it has become increasingly clear that empirical SADs often
85 deviate from a lognormal by having more than one internal mode (Ugland & Gray, 1982; Gray *et al.*,
86 2005; Dornelas & Connolly, 2008). Multimodality is seldom reported and its implications little
87 explored (McGill *et al.*, 2007), with some notable, but dispersed, exceptions. Ugland & Gray (1982)
88 proposed three lognormal distributions, corresponding to rare, intermediate abundant and common
89 species, to describe non-equilibrium marine benthic communities. Magurran & Henderson (2003)
90 ‘deconstructed’ an estuarine fish community into two groups - ‘core’ and ‘occasional’, based on
91 species persistence and habitat preferences, where the first group was better fit by a lognormal, while
92 the ‘occasional’ group of rare species followed a logseries distribution. Gray *et al.* (2005) showed that

93 a mixture of two lognormal distributions provided a good fit to a marine benthos and a tropical tree
94 data, again separating the species into ‘abundant’ and ‘rare’.

95

96 In the first statistical analysis comparing the fit of distributions with varying numbers of modes,
97 Dornelas & Connolly (2008) showed that the SAD of an intensely sampled coral community was
98 multimodal. However, the different modes could not be explained by mixture of species associated
99 with different habitats, and were only partially explained by different spatial aggregation. Recently,
100 Matthews *et al.* (2014), using the same methodology for an arthropod community, showed that
101 multimodal distributions performed better for many of the samples analysed, and that grouping
102 ecologically different species leads to multimodality, with the rarest species mode containing a higher
103 proportion of satellite, introduced and species better adapted to other habitats. However, the effect of
104 dispersal ability was unclear, and a body size niche axis was unrelated to the multimodal patterns. The
105 commonality among these studies is that they indicate that multimodality is linked to ecological
106 heterogeneity, broadly defined as groups of species with different ecological or functional
107 characteristics. This suggests that multimodality should have higher prevalence among communities
108 with higher ecological heterogeneity. Our concept of ecological heterogeneity is intended to
109 encompass the spatial, environmental, taxonomic and functional aspects of ecological systems, rather
110 than simply the number of species or functional groups.

111

112 The prevalence of multimodality in empirical SADs is as yet unknown. In a recent theoretical study,
113 Barabás *et al.* (2013) reported that stochastic versions of both resource partitioning and neutral models
114 can produce multimodal SADs with a 50% prevalence. The authors argue that in nature, individual
115 realizations are likely to differ from the mean predicted pattern due to stochastic processes, while
116 disputing that the Emergent Neutrality model proposed by Vergnon *et al.* (2012) is the only one able
117 to produce multimodal SADs. Thus, assessing the prevalence of multimodality in empirical datasets is
118 warranted to establish the generality of the pattern, as well as help elucidate how it can be related to
119 different ecological explanations.

120

121 Here, we undertake a comprehensive global assessment of the prevalence of multimodality for a wide
122 range of communities. This is, to our knowledge, the first assessment of the prevalence of multiple
123 modes in SADs. We improve the method of multimodality detection and show with high confidence
124 that multimodality occurs in 17 out of 117 assemblages. Additionally, we test the hypothesis that
125 more heterogeneous communities are more likely to exhibit multimodality. We show that
126 multimodality has higher prevalence for large scale or taxonomically heterogeneous communities.
127 Broader spatial extent and higher taxonomic breadth (as measured by family diversity) underpin
128 higher ecological heterogeneity, and hence we suggest these as potential explanations for
129 multimodality in SADs.

130

131

132 **Materials and Methods**

133

134 Model Selection

135 To test whether distributions with more than one mode accurately reflect the abundance distributions
136 of the underlying communities we used maximum likelihood methods to explicitly compare the fit of
137 mixtures of 1, 2 and 3 Poisson Lognormal distributions (1PLN, 2PLN and 3PLN, respectively)
138 (Pielou, 1969; Bulmer, 1974); a logseries distribution was also included (Fig. 1). All the calculations
139 were performed in the software R (R Core Team, 2013). Functions to fit the PLN mixtures and to
140 calculate maximum likelihood estimates (MLE) were adapted from Dornelas & Connolly (2008) but
141 using the `dpoilog()` function from `poilog` package (Grøtan & Engen, 2008); the log-likelihood
142 functions are otherwise similar and best fit parameters were found by minimizing the negative log-
143 likelihood (functions available in Appendix S1). Parameter estimation was performed using the R
144 optimization routine `nlm` and parameter searches were initialized from multiple starting points due
145 to the possibility of several local maxima for more complex distributions (Dornelas & Connolly,
146 2008; Connolly & Dornelas, 2011).

147

148 Model comparison was performed under a multi-model information-theoretic framework (Burnham &
149 Anderson, 2002), using the second order Akaike's information criterion for small sample sizes (AIC_c ,
150 Burnham & Anderson, 2002) and Bayesian information criterion (BIC, Schwarz 1978). AIC_c was
151 used throughout as it converges to AIC when sample size is large (Burnham & Anderson, 2002,
152 2004). AIC and BIC are model selection tools that provide quantitative relative support for alternative
153 hypotheses, while finding a compromise between goodness of fit and model complexity. AIC tends to
154 overestimate the number of distributions in mixture models, while BIC tends to underestimate them
155 (McLachlan & Peel, 2000; Henson *et al.*, 2007). Hence, we evaluated the performance of these two
156 model selection criteria with a simulation study.

157

158 We evaluated model performance in slightly different ways in the empirical and simulation studies.
159 For the analysis of the empirical data, relative support for the models was calculated as ΔAIC , which
160 is the difference between the AIC_c of each model, and the lowest AIC_c in the model set. Differences
161 larger than 2 indicate substantial evidence against the model with the higher AIC_c (or BIC) (Burnham
162 & Anderson, 2002). However, for the simulation study, the "true model" (the model used to generate
163 the simulated data) is known. Therefore, we calculated AIC differences relative to this *true* model, a
164 quantity we term AICdiff. Specifically, AICdiff is the AIC_c of the *true* model, minus the smallest
165 AIC_c of the remaining models. This quantity is negative whenever the *true* model is the best fitting
166 model (the one with the lowest AIC score). Conversely, if one or more of the alternative models
167 actually fits better than the *true* model does, then AICdiff will be positive (note that $AICdiff=0$ does
168 not indicate the best fitting model). An analogous quantity was calculated for BIC.

169

170

171 Simulation Study

172 Because the *PLN-mixture* method has only been applied to specific datasets (Dornelas & Connolly,
173 2008; Vergnon *et al.*, 2012; Matthews *et al.*, 2014), we conducted a simulation study to assess how it
174 performed under a broad range of parameter combinations. We were specifically interested in
175 determining which conditions lead to selection of a model with the wrong number of internal modes.

176 We defined a *false positive* as simulated samples where a multimodal distribution was selected with
177 high confidence when the *true* distribution generating the sample was not multimodal; and a *false*
178 *negative* as simulated samples where the *true* distribution was multimodal but for which a ‘non-
179 multimodal’ distribution was selected. A range of species richness and parameter values for the four
180 alternative abundance distributions models was used to generate simulated count data. The spectrum
181 of parameters used was designed to cover a realistic range for species abundance data (Connolly &
182 Thibaut, 2012), and to provide a quantitative picture of whether and when the method fails to select
183 the *true* number of underlying modes. A total of 162 parameter combinations were examined; for each
184 parameter combination, 100 simulated SAD samples were generated and the alternative log-likelihood
185 functions were fit (see Appendix S2 for more details and code).

186

187

188 Parametric Bootstrap

189 Following the simulation study results, we identified some 1PLN parameter combinations where AIC_c
190 strongly selected a more complex model than the one generating the data with a frequency of up to
191 ~25% of the simulated samples (Fig. S2.2 in Appendix S2). We additionally calculated likelihood
192 ratio tests (LRT) to minimise the chance of a multimodal distribution being selected due to
193 overfitting. Likelihood ratio tests assess if the improvement in goodness of fit of a more complex
194 models is greater than would be expected by chance, if the simpler model were true. LRT are only
195 applicable to nested models, so the logseries was not included in this analysis. Because the null
196 distribution of LRT is known to occasionally deviate from a chi squared distribution (McLachlan,
197 1987; McLachlan & Peel, 2000), we generated null LRT frequency distributions from 1PLN
198 simulated communities. This allows calculating the equivalent of a p-value for the null hypothesis that
199 the sampled data are consistent with a 1PLN distribution, thus providing an alternative assessment of
200 whether a multimodal model provided the best-fit for that parameter combination. For the simulation
201 study, we illustrate this by comparing LRT distributions for two parameter combinations, one from
202 the parameter space where AIC_c successfully selected 1PLN, and the other from the space where AIC_c
203 has a higher probability of selecting a more complex model.

204

205 For the empirical data, we conducted a parametric bootstrap likelihood ratio test (PBLRT; see Knappe
206 & de Valpine 2012 for an example) for all the SADs selected as multimodal by AIC_c . The parametric
207 bootstrap procedure consisted of randomly generating species abundance values from a 1PLN density
208 function parameterized using the model's maximum likelihood estimates for that empirical dataset
209 (Connolly *et al.*, 2009). As these analyses are very computationally intensive (Dornelas & Connolly,
210 2008; Connolly & Dornelas, 2011), 100 parametric bootstrap samples were generated for each
211 dataset, using $\hat{\mu}$ and $\hat{\sigma}$ (the estimated mean and standard deviation of log-abundances) and sample size
212 as the observed number of species, and the log-likelihood functions were fit (details and code in
213 Appendix S2). This procedure allowed comparing the *empirical* likelihood ratio, calculated from the
214 empirical SAD fitting, with the frequency distribution expected under the null hypothesis that the data
215 are actually a single PLN.

216

217

218 Empirical Data

219 117 datasets from intensely sampled communities were collected from 3 online repositories: OBIS
220 (Ocean Biogeographic Information System, <http://www.iobis.org/>), Ecological Data Wiki
221 (<http://ecologicaldata.org/>) and GBIF (Global Biodiversity Information Facility, <http://www.gbif.org/>)
222 (Fig. 2). These data are freely available and a complete list of the data sources can be found in
223 Appendix S3. For each dataset a simplified vector of abundances was obtained, corresponding to one
224 year of sampling only (the most recent year with at least 10,000 individuals where multiple years were
225 sampled). This was intended to prevent interannual variability from inducing multimodality *sensu*
226 Magurran & Henderson (2003), as we were interested in assessing the prevalence of multimodal
227 SADs independent of a temporal effect of species abundances fluctuations among years.

228

229 Datasets were classified according to spatial extent and taxonomic breadth. These two variables were
230 intended to represent different axes of ecological heterogeneity. Regarding spatial extent, as explicit
231 estimates of extent were not available for all datasets, datasets were classified as *Local* when data

232 originated from plots or sampling stations within less than 1° latitude/longitude, as *Regional* when
233 data comprised larger areas (e.g. countrywide or larger biome patches), and as *Continental* when data
234 spanned broader areas such as the whole eastern North American coast or Antarctica. Regarding
235 taxonomic breadth, we used the number of families to quantify this variable. We analysed whether the
236 prevalence of multimodality was influenced by spatial extent and taxonomic breadth (and their
237 interaction) using two models: first, we used a binomial generalised linear model (GLM), aggregating
238 1PLN and logseries as ‘non-multimodal’. Additionally, we used a multinomial Bayesian generalised
239 linear model to assess the prevalence of multimodality, 1PLN and logseries separately. We performed
240 the Markov chain Monte Carlo (MCMC) estimation using the R package MCMCglmm (Hadfield,
241 2010). We fitted a model with a random intercept to obtain improved parameter estimates for each
242 level of the fixed effects (see MCMCglmm vignette (Hadfield, 2010) and Gelman & Hill, 2007),
243 running 5,000,000 iterations with a burn-in of 100,000 and a thinning interval of 25.

244

245

246 **Results**

247

248 Simulation Study

249 Overall, the *PLN-mixture* method was robust to large variation in the parameters used to perform the
250 simulations. The false positive frequency was very low, particularly for BIC where in only 1% of the
251 cases was a multimodal distribution selected with high confidence as the best fit model when the *true*
252 distribution was not multimodal, and for AIC_c it was 6%. Species richness (S) had strong and
253 disparate effects for the 1PLN simulations. For AIC_c, the percentage of false positives increased with
254 S, while for BIC the percentage of failures decreased. When inspecting the 1PLN simulation results in
255 more detail, some particular parameter combinations led AIC_c to consistently and strongly favour
256 more complex models than the one generating the data (Fig. S2.2 in Appendix S2).

257

258 The overall false negative frequency, i.e. simulations where the distribution generating the sampled
259 communities was multimodal but for which a ‘non-multimodal’ distribution was selected as best fit,

260 was 25% for AIC_c and 39% for BIC. For 2PLN and 3PLN simulations, the *true* model was selected
261 when the modes were clearly separated, for smaller σ values and for higher species richness. BIC
262 started to select a simpler model as the distance between the modes decreased ‘earlier’ than AIC_c,
263 which was still able to select the *true* model for closer modes. Overall, the position of the modes,
264 species richness and particularly σ values showed strong effects in the best-fit model selection, for
265 both AIC_c and BIC, sometimes with different directions (see Appendix S2 for more detailed results
266 and figures for each set of simulations).

267

268 When likelihood ratio tests were used in addition to AIC_c, the chance of selecting a more complex
269 model decreased compared to when using AIC_c alone (Fig. S2.5 in Appendix S2). For the parameter
270 space where AIC_c very rarely selected a multimodal distribution, the LRT distribution overlapped
271 with the AIC_c selection pattern (Fig. S2.5 a and b). When AIC_c had a higher false positive frequency,
272 using the LRT reduced the chance of erroneously selecting a multimodal model. Furthermore, the
273 parametric bootstrap p-value is more conservative than the critical value from a chi squared
274 distribution for the latter case (Fig. S2.5 c and d). Hence, both AIC_c and PBLRT were used to analyse
275 the empirical SADs, as the high false negative frequency for BIC suggests that it might not effectively
276 detect multimodality.

277

278

279 Empirical Data

280 Of the 117 SADs analysed, AIC_c selected a multimodal distribution for 47 SADs, 26 of which with
281 high confidence. For many SADs, estimated 1PLN parameters fell within the parameter space for
282 which AIC_c often selects a multimodal model with high confidence when the true distribution is
283 unimodal (specifically with an estimated standard deviation of log abundance, $\hat{\sigma}$, of about 2). On the
284 other hand, all the SADs selected as logseries also had estimated $\sigma \geq 2$ for the 1PLN model. This
285 suggests that the method is not overfitting generally, but can occasionally select a more complex
286 model. On visual inspection, none of the fitted curves seemed to be odd-looking or out of phase with

287 the empirical SAD (Figs. 3 and S3.1 in Appendix S3), although it is possible that SADs that appear
288 unimodal are better fit by multimodal distributions, and vice-versa (Matthews *et al.*, 2014).

289

290 For the SADs selected as multimodal, PBLRT supported AIC_c model selection for 17 SADs
291 (*empirical* likelihood ratio values were higher than the bootstrap p-value from the PBLRT
292 distribution; Fig. 3 and Table S2.2). For the cases where the PBLRT results did not support
293 multimodality, the second best model was assumed to be the best model (either logseries or 1PLN).
294 Overall, 17 SADs are multimodal with high confidence, 1PLN was the best model for 54 and for 46 it
295 was logseries. None of the datasets selected as logseries had continental spatial scale (Table S3.1).

296

297 Both spatial extent and taxonomic breadth have a positive effect on the prevalence of multimodality
298 (Table 1). For the binomial GLM, SADs with Local spatial extent were significantly less likely to be
299 multimodal ($p=0.0073$) vs Continental and Regional scales, and there is a positive effect of the
300 interaction between number of families and the Local scale ($p=0.00407$). When using the
301 multinomial GLM, SADs with Local spatial extent were again significantly less likely to be
302 multimodal vs 1PLN (Fig. 4; $p\text{MCMC}=0.01943$), but not at Continental and Regional scales. There is
303 a positive effect of the interaction between number of families and the Local scale, with the
304 proportion of multimodality vs 1PLN increasing as the number of families increases ($p\text{MCMC}=
305 0.00106$). In other words, relative to 1PLN, multimodality is significantly less prevalent at Local
306 scales and low family richness, compared to when family richness is higher or spatial extent is
307 Regional or Continental. Conversely, logseries is less prevalent vs 1PLN at Continental scales
308 ($p\text{MCMC}=0.01636$), and more prevalent at Regional and Local scales ($p\text{MCMC}=0.00923$ and
309 $p\text{MCMC}=0.01578$, respectively; Table 1). These effects are independent of number of families,
310 which does not influence significantly the proportion of logseries vs 1PLN.

311

312

313 **Discussion**

314

315 Here we show that 17 out of 117 SADs are multimodal with high confidence (~15%). Further, there is
316 a higher prevalence of multimodality for communities with broader spatial scale or higher taxonomic
317 breadth, suggesting that multimodality increases with ecological heterogeneity. This warrants
318 systematic consideration of multimodality in the quantification of SAD shape.

319

320 Our analysis across different taxa, biomes and species richness indicates that multimodality is not an
321 artefact of particular SADs. The only particularity of the SADs analysed is that they were intensely
322 sampled, and there is no reason to suspect that this holds any influence as to whether the underlying
323 ecological community is multimodal. Furthermore, because each empirical SAD analysed
324 corresponds to only one year of sampling, multimodality reflects the structure of the community at a
325 particular point in time. Additionally, we infer multimodality only when it is supported by both AIC_c
326 and PBLRT. Given that false negatives were more prevalent than false positives in our simulation
327 study, this renders our conclusions highly conservative. A caveat of our study is that the SADs
328 analysed here do not fully represent the spectrum of community variability in terms of spatial and
329 taxonomic coverage; furthermore, our sample of SADs was not intended to be representative of taxa,
330 habitat, climatic regions or even realm. Nevertheless, our results show a positive effect of both spatial
331 scale and taxonomic breadth on the prevalence of multimodality, regardless of taxa and realm.

332

333 The prevalence of multimodality we found differs from that suggested by Barabás *et al.* (2013). Our
334 simulation study showed that depending on the parameter combination, sampled communities from a
335 single PLN can indeed produce apparently multimodal SADs, as the authors suggested. However, we
336 believe that the method developed here improves our ability to test for multimodality. Despite there
337 being no direct correspondence between Barabás *et al.*'s parameterization and ours, their Fig. 4
338 suggests that the mode of the average unimodal distributions is located around octave 6 of the SAD,
339 with the distributions spanning 11 octaves. This could be compared to our 1PLN simulations for
340 larger μ , σ and species richness values, which fall in the parameter space for which AIC_c has a higher
341 chance of erroneously selecting multimodality. Thus it would be interesting to investigate whether

342 performing the additional LRT to the SADs generated using Barabás *et al.*'s parameterization would
343 still yield similar multimodality frequencies.

344

345

346 General explanations for multimodality

347 Scale is fundamental to understanding biodiversity patterns (Levin, 1992; McGill, 2010). Our results
348 indicate that multimodality is more likely to occur for regional to continental-scale SADs, albeit not
349 exclusively. Some SADs selected as multimodal consist of local samples or plots, but all of these are
350 taxonomically diverse (between 12 and 76 families): ID3 consists of macrobenthos samples from the
351 Belgian Continental Shelf; IDs 95 and 96 of tropical forest plots in Malaysia, ID99 in Thailand, and
352 IDs 101 and 102 of tropical plots in Brazil and Colombia, respectively; and IDs 45, 92 and 108
353 consist of vegetation plots in the USA (desert, shortgrass steppe and dune vegetation, respectively).
354 This matches our regression analysis, for which local SADs with low family richness exhibited lower
355 prevalence of multimodality than it did at high family richness or broad spatial scales.

356

357 The explanatory variables we analysed here mirror the spatial and organizational scales suggested by
358 Levin (1992) as underpinning the variability of ecological patterns, and they support previous
359 explanations for multimodality. Multimodality has been proposed to arise as consequence of species
360 differences in ecological or functional characteristics (e.g. Magurran & Henderson 2003; Alonso *et al.*
361 2008) and of environmental heterogeneity (Dornelas *et al.*, 2009). Both of these explanations are
362 consistent with a greater prevalence of multimodality in communities with greater spatial extent or
363 taxonomic diversity. Our goal was not to develop a predictive model for multimodality, but to
364 quantify its prevalence and test its association with relevant ecological variables. We believe that
365 exploring in more detail the effects of environmental heterogeneity, functional diversity, and core-
366 transient species will prove a fruitful avenue to further understand what aspects of ecological
367 heterogeneity affect SAD shape and lead to multimodality.

368

369 An additional interesting research question is how temporal variability in the species abundances
370 might affect SADs' shapes over time. In the present study, we were interested in removing the
371 potential effect of temporal fluctuations of the relative abundances of species across years, to avoid
372 the possibility that multimodality could arise as an artefact of a single mode changing position over
373 time. In principle, it is also possible that pooling could reduce multimodality, if changes in the
374 position of modes over time make multiple modes more difficult to detect (for instance, if
375 multimodality arises as a transient feature of communities, as an effect of particular stochastic
376 environmental effects). Because the models we use implicitly account for sampling effects, and
377 require actual counts (number of individuals sampled), an investigation into the effects of temporal
378 averaging would require the development of an alternative statistical approach.

379

380

381 Rarity and commonness

382 SAD studies have often focused on the left-hand side of the distribution and on different theoretical
383 models' ability to accommodate the rarest species mode (e.g. Hubbell 2001; McGill 2003b), and
384 several studies have described the rarer mode as the one leading to a multimodal pattern (Magurran &
385 Henderson, 2003; Borda-de-Água *et al.*, 2012; Matthews *et al.*, 2014). Although a mode was often
386 fitted to the rarest species, some of the empirical SAD also exhibited modes for very abundant species
387 (e.g. IDs 30, 92, 99 and 108 in Fig. 3). This highlights the observation that communities characterized
388 by very high abundances of the most abundant species might not be accommodated within a single
389 lognormal SAD, and a multimodal distribution provides a better description, similarly to communities
390 with a very high prevalence of rare species. While the majority of species are rare and the universal
391 'hollow-curve' SAD is the definitive description of this, the few most common species
392 disproportionately dominate communities in terms of abundance and ecological processes (Gaston,
393 2010, 2011), and might also have considerable influence on SAD shape (e.g. Connolly *et al.* 2014).

394

395 Logseries distributions were selected as best model relatively frequently, despite all of the data
396 coming from intensely sampled communities. This suggests that, even for high sampling intensity,

397 some communities are characterized by a very high proportion of rare species. The logseries was
398 more often selected for communities encompassing smaller spatial scales, a finding consistent with
399 our regression analysis results. Additionally, visual inspection suggests that there was a slight
400 tendency for the logseries to be favoured when species richness was lower (not shown), and in our
401 analysis logseries was never the model with the best absolute fit (in terms of negative log-likelihood
402 values only; c.f. Baldrige *et al.*, 2015). Interestingly, none of the SADs selected as logseries had the
403 largest spatial extent, contrasting with the predictions of neutral theory with point-mutation speciation
404 (Hubbell, 2001), which predicts a logseries SAD for the metacommunity. On the other hand, the
405 maximum entropy theory of ecology (METE, Harte *et al.* 2008) predicts a logseries SAD, contrasting
406 with the support for multimodality we found, and with the effect of spatial scale and taxonomic
407 breadth on model frequency.

408

409

410 Model selection tools

411 The simulation study showed that the position of the modes, species richness and particularly σ values
412 greatly affected model selection, for both AIC_c and BIC. Additionally, species richness often had
413 contrary effects on the information criteria; this can be related to the high level of penalization exerted
414 by AIC_c as sample size decreases (Burnham & Anderson, 2002), while the opposite happens for BIC
415 (by definition), which can be problematic when testing for multimodality in SADs. As expected, BIC
416 was more conservative than AIC_c , reflected both in the very low false positive frequency and
417 particularly in the relatively high frequency of false negatives. While the former is a highly desirable
418 feature of a selection method, the latter suggests that BIC can be insensitive to deviations in SADs
419 indicative of multimodality.

420

421 On the other hand, although AIC_c overestimated the number of modes for some parameter
422 combinations, for a large number of empirical SADs with estimated parameters within that space, the
423 more parsimonious model was selected. This suggests that AIC_c is not overestimating the number of
424 modes generally, and that model selection criteria might be affected by parameter values in a

425 nondirectional fashion. As noted before for SADs, comparative measures of goodness of fit can often
426 produce conflicting results (McGill, 2003a; McGill *et al.*, 2007). We showed that additionally
427 calculating LRT frequency distributions further reduces the probability of erroneously selecting
428 multimodality when compared to using AIC_c alone.

429

430

431 Conclusions

432 Multimodal SADs occur at a non-negligible frequency. Larger spatial scale or higher taxonomic
433 breadth can yield multimodal SADs. Greater spatial scale and taxonomic breadth of the communities
434 imply higher ecological heterogeneity. In turn, this is expressed as different levels of species
435 abundance, thus being reflected in the SAD shape and informing on community structure. Here we
436 show that the dichotomy between logseries and lognormal as the sole adequate descriptors of SAD
437 should be expanded to include multimodal models. This will enhance our ability to use SADs to
438 detect the effects of ecological or functional mechanisms affecting the communities. Furthermore,
439 differences in SAD shape across different scales provide important insights to the current endeavour
440 of biodiversity scaling.

441 **Acknowledgements**

442

443 We thank Maria João Janeiro, Petr Keil, Felix May and another anonymous reviewer for their
444 thoughtful comments which considerably improved our manuscript. The authors are grateful to the
445 online repositories OBIS, Ecological Data Wiki and GBIF. We are also grateful to all the scientists
446 and participants in data collection, to all the data providers for making data publicly available and to
447 their funders (see Appendix S3 for the complete list of acknowledgements). We thank the University
448 of St Andrews MHD Cluster and the Bioinformatics Unit (Wellcome Trust ISSF grant
449 105621/Z/14/Z). LHA was supported by Fundação para a Ciência e Tecnologia, Portugal (POPH/FSE
450 SFRH/BD/90469/2012). AEM acknowledges the ERC (BioTIME 250189). MD acknowledges
451 funding from the Marine Alliance for Science and Technology Scotland (MASTS), funded by the
452 Scottish Funding Council (grant reference HR09011) and contributing institutions.

453

454

455 **Supporting information**

456

457 Appendix S1 – Log-likelihood functions (R code for fitting mixtures of 1, 2 and 3 PLN distributions).

458 Appendix S2 – S2 A: Simulation study details and results;

459 S2 B: Likelihood Ratio Test code and results;

460 Appendix S3 – S3 A: Empirical data selection criteria and non-multimodal SADs plots;

461 S3 B: Data sources and empirical datasets information & Δ AIC values and Akaike

462 weights for the empirical SADs (provided as csv files).

463

464

465 **Biosketch**

466 Laura Henriques Antão is interested in general patterns in ecology and macroecology, across taxa and
467 ecosystems, and also in biodiversity and the processes explaining species coexistence. Recently, her
468 research has been focusing on species abundance distributions.

469

470

471 **Author contributions**

472 LHA assembled the datasets and performed all the analyses and modeling work. SC and MD

473 contributed to the modeling development and analysis of the results. LHA wrote the first draft of the

474 manuscript; all authors have discussed the results and contributed extensively to improved revisions.

475 **References**

476

477 Alonso, D., Ostling, A. & Etienne, R.S. (2008) The implicit assumption of symmetry and the species
478 abundance distribution. *Ecology Letters*, **11**, 93–105.

479 Baldrige, E., Xiao, X. & White, E.P. (2015) An extensive comparison of species-abundance
480 distribution models. *bioRxiv*.

481 Barabás, G., D’Andrea, R., Rael, R., Meszéna, G. & Ostling, A. (2013) Emergent neutrality or hidden
482 niches? *Oikos*, **122**, 1565–1572.

483 Borda-de-Água, L., Borges, P.A.V., Hubbell, S.P. & Pereira, H.M. (2012) Spatial scaling of species
484 abundance distributions. *Ecography*, **35**, 549–556.

485 Bulmer, M. (1974) On fitting the Poisson lognormal distribution to species-abundance data.
486 *Biometrics*, **30**, 101–110.

487 Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference: a practical
488 information-theoretic approach*, 2nd edn. Springer, New York.

489 Burnham, K.P. & Anderson, D.R. (2004) Multimodel Inference: Understanding AIC and BIC in
490 Model Selection. *Sociological Methods & Research*, **33**, 261–304.

491 Connolly, S.R. & Dornelas, M. (2011) *Fitting and empirical evaluation of species abundance models.*
492 *Biological Diversity: Frontiers in Measurement and Assessment* (ed. by A.E. Magurran) and
493 B.J. McGill), pp. 123–140. Oxford University Press, Oxford.

494 Connolly, S.R., Dornelas, M., Bellwood, D.R. & Hughes, T.P. (2009) Testing species abundance
495 models: a new bootstrap approach applied to Indo-Pacific coral reefs. *Ecology*, **90**, 3138–49.

496 Connolly, S.R., MacNeil, M.A., Caley, M.J., Knowlton, N., Cripps, E., Hisano, M., Thibaut, L.M.,
497 Bhattacharya, B.D., Benedetti-Cecchi, L., Brainard, R.E., Brandt, A., Bulleri, F., Ellingsen,
498 K.E., Kaiser, S., Kroncke, I., Linse, K., Maggi, E., O’Hara, T.D., Plaisance, L., Poore, G.C.B.,
499 Sarkar, S.K., Satpathy, K.K., Schuckel, U., Williams, A. & Wilson, R.S. (2014) Commonness
500 and rarity in the marine biosphere. *Proceedings of the National Academy of Sciences of the
501 United States of America*, **111**, 8524–8529.

502 Connolly, S.R. & Thibaut, L.M. (2012) A comparative analysis of alternative approaches to fitting

503 species-abundance models. *Journal of Plant Ecology*, **5**, 32–45.

504 Dornelas, M. & Connolly, S.R. (2008) Multiple modes in a coral species abundance distribution.
505 *Ecology Letters*, **11**, 1008–1016.

506 Dornelas, M., Moonen, A.C., Magurran, A.E. & Bàrberi, P. (2009) Species abundance distributions
507 reveal environmental heterogeneity in modified landscapes. *Journal of Applied Ecology*, **46**,
508 666–672.

509 Fisher, R., Corbet, A. & Williams, C. (1943) The relation between the number of species and the
510 number of individuals in a random sample of an animal population. *The Journal of Animal
511 Ecology*, **12**, 42–58.

512 Gaston, K.J. (2011) Common Ecology. *BioScience*, **61**, 354–362.

513 Gaston, K.J. (2010) Valuing common species. *Science*, **327**, 154–155.

514 Gelman, A. & Hill, J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*,
515 1st ed. Cambridge University Press, Cambridge.

516 Gray, J.S., Bjørgesæter, A. & Ugland, K.I. (2006) On plotting species abundance distributions. *The
517 Journal of Animal Ecology*, **75**, 752–756.

518 Gray, J.S., Bjørgesæter, A. & Ugland, K.I. (2005) The impact of rare species on natural assemblages.
519 *Journal of Animal Ecology*, **74**, 1131–1139.

520 Green, J.L. & Plotkin, J.B. (2007) A statistical theory for sampling species abundances. *Ecology
521 Letters*, **10**, 1037–1045.

522 Grøtan, V. & Engen, S. (2008) poilog: Poisson lognormal and bivariate Poisson lognormal
523 distribution. R package version 0.4.

524 Hadfield, J.D. (2010) MCMC Methods for Multi-Response Generalized Linear Mixed Models: The
525 MCMCglmm R Package. *Journal of Statistical Software*, **33**.

526 Harte, J., Zillio, T., Conlisk, E. & Smith, A. (2008) Maximum entropy and the state-variable approach
527 to macroecology. *Ecology*, **89**, 2700–2711.

528 Henson, J.M., Reise, S.P. & Kim, K.H. (2007) Detecting Mixtures From Structural Model Differences
529 Using Latent Variable Mixture Modeling: A Comparison of Relative Model Fit Statistics.
530 *Structural Equation Modeling: A Multidisciplinary Journal*, **14**, 202–226.

531 Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*, Princeton
532 University Press, Princeton.

533 Knape, J. & de Valpine, P. (2012) Are patterns of density dependence in the Global Population
534 Dynamics Database driven by uncertainty about population abundance? *Ecology Letters*, **15**, 17–
535 23.

536 Levin, S. (1992) The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture.
537 *Ecology*, **73**, 1943–1967.

538 Magurran, A.E. (2004) *Measuring biological diversity*, Blackwell Science, Oxford.

539 Magurran, A.E. & Henderson, P.A. (2003) Explaining the excess of rare species in natural species
540 abundance distributions. *Nature*, **422**, 714–716.

541 Matthews, T.J., Borges, P.A. V & Whittaker, R.J. (2014) Multimodal species abundance distributions:
542 A deconstruction approach reveals the processes behind the pattern. *Oikos*, **123**, 533–544.

543 McGill, B. (2010) Matters of scale. *Science*, **328**, 575–576.

544 McGill, B. (2003a) Strong and weak tests of macroecological theory. *Oikos*, **102**, 679–685.

545 McGill, B.J. (2003b) Does Mother Nature really prefer rare species or are log-left-skewed SADs a
546 sampling artefact? *Ecology Letters*, **6**, 766–773.

547 McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas, M.,
548 Enquist, B.J., Green, J.L., He, F., Hurlbert, A.H., Magurran, A.E., Marquet, P.A., Maurer, B.A.,
549 Ostling, A., Soykan, C.U., Ugland, K.I. & White, E.P. (2007) Species abundance distributions:
550 moving beyond single prediction theories to integration within an ecological framework.
551 *Ecology Letters*, **10**, 995–1015.

552 McLachlan, G. & Peel, D. (2000) *Finite Mixture Models*, John Wiley & Sons, New York.

553 McLachlan, G.J. (1987) On Bootstrapping the Likelihood Ratio Test Statistic for the Number of
554 Components in a Normal Mixture. *Applied Statistics*, **36**, 318.

555 Pielou, E.C. (1969) *An Introduction to Mathematical Ecology*, Wiley-Interscience, New York.

556 Preston, F. (1948) The commonness, and rarity, of species. *Ecology*, **29**, 254–283.

557 R Core Team (2013) R: A language and environment for statistical computing. R Foundation for
558 Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.

- 559 Schwarz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461–464.
- 560 Ugland, K. & Gray, J. (1982) Lognormal distributions and the concept of community equilibrium.
- 561 *Oikos*, **39**, 171–178.
- 562 Vergnon, R., van Nes, E.H. & Scheffer, M. (2012) Emergent neutrality leads to multimodal species
- 563 abundance distributions. *Nature Communications*, **3**, 663.

564 **References of Data Sources used in this study**

565

566 de Abreu, L., Shackleton, N.J., Schönfeld, J., Hall, M. & Chapman, M. (2003) “Planktic foraminifera
567 counts of sediment core MD95-2040”. doi:10.1594/PANGAEA.66714, In Supplement to: de
568 Abreu *et al.* (2003): Millennial-scale oceanic climate variability off the Western Iberian margin
569 during the last two glacial periods. *Marine Geology*, 196(1-2), 1-20, doi:10.1016/S0025-
570 3227(03)00046-X. *PANGAEA*. Available at: [http://www.gbif.org/dataset/662510f4-f762-11e1-](http://www.gbif.org/dataset/662510f4-f762-11e1-a439-00145eb45e9a)
571 [a439-00145eb45e9a](http://www.gbif.org/dataset/662510f4-f762-11e1-a439-00145eb45e9a), accessed 2012.

572 Alvarez, M. “La Planada Forest Dynamics Plot, Colombia.” *The Center for Tropical Forest Science.*
573 *Smithsonian Tropical Research Institute*. Available at: <http://www.ctfs.si.edu/site/La+Planada/>,
574 accessed 2013.

575 Bakker, C., Herman, P. & Vink, M. (1994) A new trend in the development of the phytoplankton in
576 the Oosterschelde (SW Netherlands) during and after the construction of a storm-surge barrier.
577 *Hydrobiologia*, **282-283**, 79–100.

578 Bakker, K. & Herman, P. (1990) “Phytoplankton in the Oosterschelde before, during and after the
579 storm-surge barrier (1982-1990).” *Netherlands Institute of Ecology, Centre for Estuarine and*
580 *Marine Ecology, Netherlands*. Available at: [http://www.emodnet-biology.eu/data-](http://www.emodnet-biology.eu/data-catalog?module=dataset&david=1646)
581 [catalog?module=dataset&david=1646](http://www.emodnet-biology.eu/data-catalog?module=dataset&david=1646), accessed 2013.

582 Ballard, G., Herzog, M., Fitzgibbon, M., Moody, D., Jongsomjit, D. & Stralberg., D. (2008) “PRBO
583 Conservation Science - Point Counts.” *The California Avian Data Center. Petaluma, California.*
584 Available at: www.prbo.org/cadc, accessed 2012.

585 Balslev, H. “Aarhus University Palm Transect Database.” *Department of Bioscience, Aarhus*
586 *University*. Available at: <http://www.gbif.org/dataset/a9e763c8-f674-4492-94a8-4fd4eb9342a5>,
587 accessed 2013.

588 Battles, J.J., Fahey, T. & Cleavitt, N. (2003a) “Forest Inventory of a Whole Tree Harvest: Hubbard
589 Brook Experimental Forest Watershed 5, 1982, pre-harvest.” *The Hubbard Brook Ecosystem*
590 *Study LTER Program*. Available at: <http://www.hubbardbrook.org/data/dataset.php?id=36>,
591 accessed 2012.

592 Battles, J.J. & Fahey, T.J. "Tree inventory data for the Hubbard Brook Valley Plots, baseline data
593 collected 1995 - 1998." *The Hubbard Brook Ecosystem Study LTER Program*. Available at:
594 <http://www.hubbardbrook.org/data/dataset.php?id=125>, accessed 2012.

595 Battles, J.J., Johnson, C., Hamburg, S., Fahey, T., Driscoll, C. & Likens, G. (2003b) "Forest Inventory
596 of a Northern Hardwood Forest: Watershed 6 2002." The Hubbard Brook Ecosystem Study
597 LTER Program. Available at: <http://www.hubbardbrook.org/data/dataset.php?id=35>, accessed
598 2012.

599 Beck, R. (1996) "Nesting seabird census of Hog Island and Cobb Island of the Virginia Coast Reserve
600 1991." *Virginia Coast Reserve Long-Term Ecological Research Project*. Available at:
601 <http://www.vcrlter.virginia.edu/cgi-bin/showDataset.cgi?docid=knb-lter-vcrlter.18>, accessed 2013.

602 Bird Studies Canada (2012a) "BC Coastal Waterbird Survey (2004)." *NatureCounts, a node of the*
603 *Avian Knowledge Network*. Available at: <http://www.birdscanada.org/birdmon/>, accessed 2012.

604 Bird Studies Canada (2012b) "Marsh Monitoring Program." *NatureCounts, a node of the Avian*
605 *Knowledge Network*. Available at: <http://www.birdscanada.org/birdmon/>, accessed 2012.

606 Brockelman, W.Y. & Nathalang, A. "Mo Singto Forest Dynamics Plot, Thailand." *The Center for*
607 *Tropical Forest Science. Smithsonian Tropical Research Institute*. Available at:
608 <http://www.ctfs.si.edu/site/Mo+Singto/>, accessed 2013.

609 Brown, S.K.R., Zwanenburg, K. & Branton, R. (2005) "East Coast North America Strategic
610 Assessment Groundfish Atlas - ECNASAP." *OBIS Canada, Bedford Institute of Oceanography,*
611 *Dartmouth, Nova Scotia, Canada*. Available at: <http://iobis.org/>, accessed 2013.

612 Brush, G. (2007) "Permanent Plot Vegetation Sampling, 2003 Shrub and Vine Data." *Baltimore*
613 *Ecosystem Study LTER Program. Baltimore, MD, USA*. Available at:
614 http://www.beslter.org/metacat_harvest_attribute_level_eml/html_metadata/bes_414.asp,
615 accessed 2012.

616 Bunyavejchewin, S. "Huai Kha Khaeng Forest Dynamics Plot, Thailand." *The Center for Tropical*
617 *Forest Science. Smithsonian Tropical Research Institute*. Available at:
618 <http://www.ctfs.si.edu/site/Huai+Kha+Khaeng/>, accessed 2013.

619 Bunyavejchewin, S., Baker, P.J., LaFrankie, J. V. & Ashton, P.S. (2001) Stand structure of a seasonal

620 dry evergreen forest at Huai Kha Khaeng Wildlife Sanctuary, western Thailand. *Natural History*
621 *Bulletin of the Siam Society*, **49**, 89–106.

622 Bunyavejchewin, S., LaFrankie, J. V., Baker, P.J., Davies, S.J. & Ashton, P.S. (2009) Forest trees of
623 Huai Kha Khaeng Wildlife Sanctuary, Thailand: Data from the 50-hectare Forest Dynamic Plot.
624 *The National Parks, Wildlife and Plant Conservation Department*.

625 Bunyavejchewin, S., LaFrankie, J. V., Pattapong, P., Kanzaki, M., Itoh, A., Yamakura, T. & Ashton,
626 P.S. (1998) Topographic analysis of a large-scale research plot in seasonal dry evergreen forest
627 at Huai Kha Khaeng Wildlife Sanctuary, Thailand. *Tropics*, **8**, 45–60.

628 Carpenter, S., Kitchell, J., Kratz, T. & Magnuson, J. (2006a) “Biocomplexity at North Temperate
629 Lakes LTER; Coordinated Field Studies: Fish / Crayfish Abundance 2001 - 2004.” *North*
630 *Temperate Lakes Long Term Ecological Research Program, Center for Limnology, University of*
631 *Wisconsin-Madison*. Available at: [http://lter.limnology.wisc.edu/dataset/biocomplexity-north-](http://lter.limnology.wisc.edu/dataset/biocomplexity-north-temperate-lakes-lter-coordinated-field-studies-fish-crayfish-abundance-2)
632 [temperate-lakes-lter-coordinated-field-studies-fish-crayfish-abundance-2](http://lter.limnology.wisc.edu/dataset/biocomplexity-north-temperate-lakes-lter-coordinated-field-studies-fish-crayfish-abundance-2), accessed 2013.

633 Carpenter, S., Kratz, T., Cronon, W., Provencher, R. & Turner, M. (2006b) “Biocomplexity at North
634 Temperate Lakes LTER; Coordinated Field Studies: Riparian Plots 2001 - 2004.” *North*
635 *Temperate Lakes Long Term Ecological Research Program, Center for Limnology, University of*
636 *Wisconsin-Madison*. Available at: [http://lter.limnology.wisc.edu/dataset/biocomplexity-north-](http://lter.limnology.wisc.edu/dataset/biocomplexity-north-temperate-lakes-lter-coordinated-field-studies-riparian-plots-2001-2004)
637 [temperate-lakes-lter-coordinated-field-studies-riparian-plots-2001-2004](http://lter.limnology.wisc.edu/dataset/biocomplexity-north-temperate-lakes-lter-coordinated-field-studies-riparian-plots-2001-2004), accessed 2013.

638 Cavender-Bares, J. & Reich, P.B. (2012) Shocks to the system: community assembly of the oak
639 savanna in a 40-year fire frequency experiment. *Ecology*, **93**, S52–S69.

640 CCE LTER (a) "Bird and mammal counts for National Marine Fisheries cruises as part of the
641 Rockfish Recruitment Survey (ID117). *CalCOFI - Scripps Institution of Oceanography*.
642 *California Current Ecosystem (CCE) Long Term Ecological Research (LTER)*. Available at:
643 <http://oceaninformatics.ucsd.edu/datazoo/data/ccelter/datasets?action=summary&id=117>,
644 accessed 2012.

645 CCE LTER (b) “Bird and mammal counts for CalCOFI cruises off the west coast of the United States
646 (ID112).” *CalCOFI - Scripps Institution of Oceanography. California Current Ecosystem (CCE)*
647 *Long Term Ecological Research (LTER)*. Available at:

648 <http://oceaninformatics.ucsd.edu/datazoo/data/ccelter/datasets?action=summary&id=112>,
649 accessed 2012.

650 CCE LTER (c) “Bird and mammal counts for continuous plankton recorder cruises on the great circle
651 route from Vancouver, BC to Tokyo (ID118).” *CalCOFI - Scripps Institution of Oceanography*.
652 *California Current Ecosystem (CCE) Long Term Ecological Research (LTER)*. Available at:
653 <http://oceaninformatics.ucsd.edu/datazoo/data/ccelter/datasets?action=summary&id=118>,
654 accessed 2012.

655 Chuyong, G.B., Condit, R., Kenfack, D., Losos, E., Sainge, M., Songwe, N.C. & Thomas, D.W.
656 (2004) *Korup Forest Dynamics Plot, Cameroon. Forest Diversity and Dynamism: Findings from*
657 *a Large-Scale Plot Network* (ed. by E.C. Losos and E.G.J. Leigh), pp. 506–516. University of
658 Chicago Press, Chicago.

659 Clark, D. & Branton, B. (2007) “DFO Maritimes Research Vessel Trawl Surveys Fish Observations.”
660 *OBIS Canada Digital Collections. OBIS Canada, Bedford Institute of Oceanography,*
661 *Dartmouth, Nova Scotia, Canada*. Available at: <http://iobis.org/>, accessed 2013.

662 Condit, R. “Sherman Forest Dynamics Plot, Panama.” *The Center for Tropical Forest Science*.
663 *Smithsonian Tropical Research Institute*. Available at: <http://www.ctfs.si.edu/site/Sherman/>,
664 accessed 2013.

665 Condit, R. (1998) *Tropical forest census plots*, Springer-Verlag and R. G. Landes Company, Berlin,
666 Germany and Georgetown, Texas.

667 Coral Reef Ecosystem Division (2011) “CRED Rapid Ecological Assessments of Fish Belt Transect
668 Surveys and Fish Stationary Point Count Surveys in the Pacific Ocean 2000-2010.” *Coral Reef*
669 *Ecosystem Division (CRED), Pacific Island Fisheries Sciences Center, NOAA National Marine*
670 *Fisheries Service. Coral Reef Ecosystem Division, Honolulu, HI*. Available at:
671 <http://www.usgs.gov/obis-usa/>, accessed 2013.

672 CSIRO “CSIRO Marine Data Warehouse - OBIS Australia.” *CSIRO Division of Marine and*
673 *Atmospheric Research (CMAR), Australia*. Available at: <http://iobis.org/>, accessed 2013.

674 DATRAS (2010a) “Fish trawl survey: ICES Beam Trawl Survey for commercial fish species. ICES
675 Database of trawl surveys (DATRAS).” *The International Council for the Exploration of the*

676 *Sea, Copenhagen*. Available at: [http://www.emodnet-biology.eu/data-](http://www.emodnet-biology.eu/data-catalog?%3Fmodule=dataset&dasid=2761)
677 [catalog?%3Fmodule=dataset&dasid=2761](http://www.emodnet-biology.eu/data-catalog?%3Fmodule=dataset&dasid=2761), accessed 2013.

678 DATRAS (2010b) “Fish trawl survey: ICES French Southern Atlantic Bottom Trawl Survey for
679 commercial fish species. ICES Database of trawl surveys (DATRAS).” *The International*
680 *Council for the Exploration of the Sea, Copenhagen*. Available at: [http://www.emodnet-](http://www.emodnet-biology.eu/data-catalog?%3Fmodule=dataset&dasid=2759)
681 [biology.eu/data-catalog?%3Fmodule=dataset&dasid=2759](http://www.emodnet-biology.eu/data-catalog?%3Fmodule=dataset&dasid=2759), accessed 2013.

682 DATRAS (2010c) “Fish trawl survey: ICES North Sea International Bottom Trawl Survey for
683 commercial fish species. ICES Database of trawl surveys (DATRAS).” *The International*
684 *Council for the Exploration of the Sea, Copenhagen*. Available at: [http://www.emodnet-](http://www.emodnet-biology.eu/data-catalog?%3Fmodule=dataset&dasid=2763)
685 [biology.eu/data-catalog?%3Fmodule=dataset&dasid=2763](http://www.emodnet-biology.eu/data-catalog?%3Fmodule=dataset&dasid=2763), accessed 2013.

686 DATRAS (2010d) “Fish trawl survey: Irish Ground Fish Survey for commercial fish species. ICES
687 Database of trawl surveys (DATRAS).” *The International Council for the Exploration of the*
688 *Sea, Copenhagen*. Available at: [http://www.emodnet-biology.eu/data-](http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=2762)
689 [catalog?module=dataset&dasid=2762](http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=2762), accessed 2013.

690 DATRAS (2010e) “Fish trawl survey: Northern Irish Ground Fish Trawl Survey. ICES Database of
691 trawl surveys (DATRAS).” *The International Council for the Exploration of the Sea,*
692 *Copenhagen*. Available at: [http://www.emodnet-biology.eu/data-](http://www.emodnet-biology.eu/data-catalog?%3Fmodule=dataset&dasid=2764)
693 [catalog?%3Fmodule=dataset&dasid=2764](http://www.emodnet-biology.eu/data-catalog?%3Fmodule=dataset&dasid=2764), accessed 2013.

694 DATRAS (2010f) “Fish trawl survey: Scottish Rockall Survey for commercial fish species. ICES
695 Database of trawl surveys (DATRAS).” *The International Council for the Exploration of the*
696 *Sea, Copenhagen*. Available at: [http://www.emodnet-biology.eu/data-](http://www.emodnet-biology.eu/data-catalog?%3Fmodule=dataset&dasid=2767)
697 [catalog?%3Fmodule=dataset&dasid=2767](http://www.emodnet-biology.eu/data-catalog?%3Fmodule=dataset&dasid=2767), accessed 2013.

698 DATRAS (2010g) “Fish trawl survey: Scottish West Coast Survey for commercial fish species. ICES
699 Database of trawl surveys (DATRAS).” *The International Council for the Exploration of the*
700 *Sea, Copenhagen*. Available at: [http://www.emodnet-biology.eu/data-](http://www.emodnet-biology.eu/data-catalog?%3Fmodule=dataset&dasid=2766)
701 [catalog?%3Fmodule=dataset&dasid=2766](http://www.emodnet-biology.eu/data-catalog?%3Fmodule=dataset&dasid=2766), accessed 2013.

702 Day, F. (2010) “Long-term N-fertilized vegetation plots on Hog Island, Virginia Coastal Barrier
703 Islands, 1992-2014.” *Virginia Coast Reserve Long-Term Ecological Research Project*. Available

704 at: <http://www.vcrlter.virginia.edu/cgi-bin/showDataset.cgi?docid=knb-lter-vcr.106>, accessed
705 2013.

706 Day, F.P., Conn, C., Crawford, E. & Stevenson, M. (2004) Long-term effects of nitrogen fertilization
707 on plant community structure on a coastal barrier island dune chronosequence. *Journal of*
708 *Coastal Research*, **20**, 722–730.

709 Degraer, S., Wittoeck, J., Appeltans, W., Cooreman, K., Deprez, T., Hillewaert, H., Hostens, K.,
710 Mees, J., Vanden Berghe, E. & Vincx, M. (2006) “Macrobenthos: Long term trends in the
711 macrobenthos of the Belgian Continental Shelf.” *Oostende, Belgium*. Available at:
712 <http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=145>, accessed 2013.

713 Dillon, R., Johnson, P., Olden, J., Solomon, C. & Zanden, J. Vander (2007) “North Temperate Lakes
714 LTER: Snail Survey in Northern Wisconsin Lakes 2006.” *North Temperate Lakes Long Term*
715 *Ecological Research Program, Center for Limnology, University of Wisconsin-Madison*.
716 Available at: [http://lter.limnology.wisc.edu/dataset/north-temperate-lakes-lter-snail-survey-](http://lter.limnology.wisc.edu/dataset/north-temperate-lakes-lter-snail-survey-northern-wisconsin-lakes-2006)
717 [northern-wisconsin-lakes-2006](http://lter.limnology.wisc.edu/dataset/north-temperate-lakes-lter-snail-survey-northern-wisconsin-lakes-2006), accessed 2013.

718 Driscoll, C., Bailey, S., Blum, J., Buso, D., Eagar, C., Fahey, T., Fisk, M., Groffman, P., Johnson, C.,
719 Likens, G., Hamburg, S. & Siccama, T.G. (2003) “Forest Inventory of a Calcium Amended
720 Northern Hardwood Forest: Watershed 1, 1996.” *The Hubbard Brook Ecosystem Study LTER*
721 *Program*. Available at: <http://www.hubbardbrook.org/data/dataset.php?id=40>, accessed 2012.

722 Ellison, A. & Gotelli, N. (2009) “Ant Distribution and Abundance in New England since 1990.
723 Harvard Forest Data Archive: HF147.” *The Harvard Forest Long Term Ecological Research*
724 *Program*. Available at: <http://harvardforest.fas.harvard.edu:8080/exist/xquery/data.xq?id=hf147>,
725 accessed 2013.

726 Ernest, S.K.M., Valone, T.J. & Brown, J.H. (2009) Long-term monitoring and experimental
727 manipulation of a Chihuahuan Desert ecosystem near Portal, Arizona, USA. *Ecology*, **90**, 1708.

728 Fletcher, C. & Kassim, A.R. “Pasoh Forest Dynamics Plot, Malaysia.” *The Center for Tropical Forest*
729 *Science, Smithsonian Tropical Research Institute*. Available at:
730 <http://www.ctfs.si.edu/site/Pasoh/>, accessed 2013.

731 Foster, D., Francis, D. & Fuller, J. (2006a) “Influence of Little Ice Age on New England Vegetation

732 from 2000 BP to Present. Harvard Forest Data Archive: HF078.” *The Harvard Forest Long*
733 *Term Ecological Research Program*. Available at:
734 <http://harvardforest.fas.harvard.edu:8080/exist/xquery/data.xq?id=hf078>, accessed 2013.

735 Foster, D., Holle, B. Von & Parshall, T. (2006b) “Land Use on the Southern New England and New
736 York Coasts 1600-2001. Harvard Forest Data Archive: HF044.” *The Harvard Forest Long Term*
737 *Ecological Research Program*. Available at:
738 <http://harvardforest.fas.harvard.edu:8080/exist/xquery/data.xq?id=hf044>, accessed 2013.

739 Foster, D. & Motzkin, G. (2003) “Land Use and Forest Dynamics at Harvard Forest 1937-1995.
740 Harvard Forest Data Archive: HF015.” *The Harvard Forest Long Term Ecological Research*
741 *Program*. Available at: <http://harvardforest.fas.harvard.edu:8080/exist/xquery/data.xq?id=hf015>,
742 accessed 2013.

743 Gaiser, E. (2010) “Macrophyte count data collected from Northeast Shark Slough, Everglades
744 National Park (FCE) from September 2006 to Present”.
745 <http://dx.doi.org/10.6073/pasta/effd9e98134913af21b670febebd6233>. *Florida Coastal*
746 *Everglades LTER Program*. Available at:
747 http://fcelter.fiu.edu/data/core/metadata/EML/?datasetid=LT_PP_Gaiser_001, accessed 2012.

748 Gido, K.B. “Fish population on selected watersheds at Konza Prairie - CFP01.” *Konza Prairie LTER*
749 *Program*. Available at:
750 <http://www.konza.ksu.edu/KNZ/pages/data/Knzdsdetail.aspx?datasetCode=CFP01>, accessed
751 2012.

752 Gould, E. & Foster, D. (2000) “Vegetation Inventory of Harvard Forest 1986-1993. Harvard Forest
753 Data Archive: HF037.” *The Harvard Forest Long Term Ecological Research Program*.
754 Available at: <http://harvardforest.fas.harvard.edu:8080/exist/xquery/data.xq?id=hf037>, accessed
755 2012.

756 Grimm, N., Hope, D., Gries, C., Martin, C. & Burns, E. (2005) “Survey 200 long term study of
757 multiple sites in central Arizona-Phoenix.” *Central Arizona-Phoenix Long-Term Ecological*
758 *Research. Global Institute of Sustainability, Arizona State University*. Available at:
759 <https://caplter.asu.edu/data/data-catalog/?id=278>, accessed 2012.

760 Gunatilleke, C.V.S., Gunatilleke, I.A.U.N., Ashton, P.S., Ethugala, A.U.K., Weerasekera, N.S. &
761 Esufali, S. (2004) *Sinharaja Forest Dynamics Plot, Sri Lanka. Forest Diversity and Dynamism:
762 Findings from a Large-Scale Plot Network* (ed. by E.C. Losos and E.G.J. Leigh), pp. 599–608.
763 University of Chicago Press, Chicago.

764 Gunatilleke, N. & Gunatilleke, S. “Sinharaja Forest Dynamics Plot, Sri Lanka.” *The Center for
765 Tropical Forest Science. Smithsonian Tropical Research Institute*. Available at:
766 <http://www.ctfs.si.edu/site/Sinharaja/>, accessed 2013.

767 Hale, S.S., Hughes, M.M., Strobel, C.J., Buffum, H.W., Copeland, J.L. & Paul, J.F. (2002) Coastal
768 ecological data from the Virginian Biogeographic Province, 1990–1993. *Ecology*, **83**, 2942–
769 2942.

770 Harmon, M. & Franklin, J. (2012) “Long-term growth, mortality and regeneration of trees in
771 permanent vegetation plots in the Pacific Northwest, 1910 to present.” *Long-Term Ecological
772 Research. Forest Science Data Bank, Corvallis*. Available at:
773 <http://andrewsforest.oregonstate.edu/data/abstract.cfm?dbcode=TV010>, accessed 2012.

774 HMANA “Hawk Migration Association of North America (HMANA).” Available at:
775 <http://www.hmana.org/>, accessed 2012.

776 Hosley, N. (2003) “Vegetation Inventory of Harvard Forest 1937. Harvard Forest Data Archive:
777 HF039.” *The Harvard Forest Long Term Ecological Research Program*. Available at:
778 <http://harvardforest.fas.harvard.edu:8080/exist/xquery/data.xq?id=hf039>, accessed 2012.

779 Hubbell, S.P., Condit, R. & Foster, R.B. (2010) “Barro Colorado Island Forest Dynamics Plot,
780 Panama.” *The Center for Tropical Forest Science. Smithsonian Tropical Research Institute*.
781 Available at: <http://www.ctfs.si.edu/site/Barro+Colorado+Island/>, accessed 2013.

782 Hubbell, S.P., Foster, R.B., O’Brien, S.T., Harms, K.E., R., C., Wechsler, B., S.J. Wright & Lao, S.L.
783 de (1999) Light gap disturbances, recruitment limitation, and tree diversity in a neotropical
784 forest. *Science*, **283**, 554–557.

785 Johnson, S. & Farrand, A. (2014) “Aquatic insect sampling in Lookout Creek at the H.J. Andrews
786 Experimental Forest, 2001.” *Long-Term Ecological Research. Forest Science Data Bank,
787 Corvallis*. Available at: <http://andrewsforest.oregonstate.edu/data/abstract.cfm?dbcode=SA017>,

788 accessed 2012.

789 Jones, J. & Miller, J. (2005) "Spatial and temporal distribution and abundance of moths in the
790 Andrews Experimental Forest, 1994 to 2008." *H. J. Andrews Experimental Forest. Forest
791 Science Data Bank, Corvallis*. Available at:
792 <http://andrewsforest.oregonstate.edu/data/abstract.cfm?dbcode=SA015>, accessed 2012.

793 Kenfack, D., Thomas, D.W., Chuyong, G. & Condit, R. (2007) Rarity and abundance in a diverse
794 African forest. *Biodiversity and Conservation*, **16**, 2045–2074.

795 Kittredge, D., Foster, D. & McDonald, R. (2009) "Timber Harvesting Field Study in Western
796 Massachusetts 2004-2005. Harvard Forest Data Archive: HF127." *The Harvard Forest Long
797 Term Ecological Research Program*. Available at:
798 <http://harvardforest.fas.harvard.edu:8080/exist/xquery/data.xq?id=hf127>, accessed 2013.

799 Lathrop, R. (2000) "Madison Wisconsin Lakes Zooplankton 1976 - 1994." *North Temperate Lakes
800 Long Term Ecological Research Program, Center for Limnology, University of Wisconsin-
801 Madison*. Available at: [http://lter.limnology.wisc.edu/dataset/madison-wisconsin-lakes-
802 zooplankton-1976-1994](http://lter.limnology.wisc.edu/dataset/madison-wisconsin-lakes-zooplankton-1976-1994), accessed 2013.

803 Lauenroth, W.K. (2013) "SGS-LTER Disturbance intensity and above- and belowground herbivory
804 effects on long-term recovery of shortgrass steppe on the Central Plains Experimental Range,
805 Nunn, Colorado, USA 1977-1990." *Shortgrass Steppe (SGS) Long Term Ecological Research
806 Program*. Available at: http://sgslter.colostate.edu/dataset_view.aspx?id=grubr, accessed 2013.

807 Lee, H., Davies, S.J., LaFrankie, J. V., Tan, S., Itoh, A., Yamakura, T. & Ashton, P.S. (2002) Floristic
808 and structural diversity of 52 hectares of mixed dipterocarp forest in Lambir Hills National Park,
809 Sarawak, Malaysia. *Journal of Tropical Forest Science*, **14**, 379–400.

810 Lee, H.S., Ashton, P.S., Yamakura, T., Tan, S., Davies, S.J., Itoh, A., Chai, E.O.K., Okhubo, T. &
811 LaFrankie, J. V. (2005) *The 52-hectare Forest Research Plot at Lambir Hills, Sarawak,
812 Malaysia: Tree distribution maps, diameter tables and species documentation. Forest
813 Department Sarawak, The Arnold Arboretum-CTFS Asia Program, Smithsonian Tropical
814 Research Institute, Kuching, Sarawak, Malaysia*.

815 Manokaran, N., Seng, Q.E., Ashton, P.S., LaFrankie, J. V., Noor, N.S.M., Ahmad, W.M.S. & Okuda,

816 T. (2004) *Pasoh Forest Dynamics Plot, Peninsular Malaysia. Forest Diversity and Dynamism:*
817 *Findings from a Large-Scale Plot Network* (ed. by E.C. Losos and E.G.J. Leigh), pp. 585–598.
818 University of Chicago Press, Chicago.

819 Marine Resources Research Institute (2011) “Southeast Area Monitoring and Assessment Program
820 (SEAMAP) South Atlantic.” *SCDNR SEAMAP-SA Program. U.S. Geological Survey*. Available
821 at: <http://www.usgs.gov/obis-usa/>, accessed 2013.

822 McLarney, W.O., Meador, J. & Chamblee, J. (2010) “Upper Little Tennessee River Biomonitoring
823 Program Database.” *Coweeta Long Term Ecological Research Program*. Available at:
824 http://coweeta.uga.edu/dbpublic/dataset_details.asp?accession=LTWA_2010_06_01, accessed
825 2012.

826 Muldavin, E. (a) “Pinon Juniper Net Primary Production Quadrat Data from the Sevilleta National
827 Wildlife Refuge, New Mexico: 1999-2001.” *Sevilleta Long Term Ecological Research Program*.
828 Available at: <http://sev.lternet.edu/data/sev-187>, accessed 2013.

829 Muldavin, E. (b) “Pinon-Juniper (Core Site) Quadrat Data for the Net Primary Production Study at the
830 Sevilleta National Wildlife Refuge, New Mexico (2003-Present).” *Sevilleta Long Term*
831 *Ecological Research Program*. Available at: <http://sev.lternet.edu/node/1718>, accessed 2013.

832 NatureCounts (a) “Maritimes Breeding Bird Atlas (2006-2010): point count data.” *NatureCounts, a*
833 *node of the Avian Knowledge Network. Bird Studies Canada*. Available at:
834 <http://www.birdscanada.org/birdmon/>, accessed 2012.

835 NatureCounts (b) “Ontario Breeding Bird Atlas (2001-2005): point count data.” *NatureCounts, a node*
836 *of the Avian Knowledge Network. Bird Studies Canada*. Available at:
837 <http://www.birdscanada.org/birdmon/>, accessed 2012.

838 Nilon, C. (2010) “Biodiversity - Fauna - Bird Survey - Table 1 of 4 - Birds.” *Baltimore Ecosystem*
839 *Study LTER Program. Baltimore, MD, USA*. Available at:
840 http://beslter.org/metacat_harvest_attribute_level_eml/html_metadata/bes_543.asp, accessed
841 2012.

842 NIWA “South Western Pacific Regional OBIS Data Specify Subset (South Western Pacific OBIS).”
843 *National Institute of Water and Atmospheric Research*. Available at: <http://iobis.org/>, accessed

844 2013.

845 NOAA/NOS/NCCOS/CCMA (2007a) “La Parguera, Puerto Rico Fish Assessment and Monitoring
846 Data (2002 - Present).” *National Oceanic and Atmospheric Association (NOAA)/National Ocean
847 Service (NOS)/National Centers for Coastal Ocean Science (NCCOS)/Center for Coastal
848 Monitoring and Assessment (CCMA) - Biogeography Team. Silver Spring, MD. Available at:
849 <http://iobis.org/>, accessed 2013.*

850 NOAA/NOS/NCCOS/CCMA (2007b) “St. Croix, USVI Fish Assessment and Monitoring Data (2002
851 - Present).” *National Oceanic and Atmospheric Association (NOAA)/National Ocean Service
852 (NOS)/National Centers for Coastal Ocean Science (NCCOS)/Center for Coastal Monitoring
853 and Assessment (CCMA) - Biogeography Team. Silver Spring, MD. Available at:
854 <http://iobis.org/>, accessed 2013.*

855 NOAA/NOS/NCCOS/CCMA (2007c) “St. John, USVI Fish Assessment and Monitoring Data (2002 -
856 Present).” *National Oceanic and Atmospheric Association (NOAA)/National Ocean Service
857 (NOS)/National Centers for Coastal Ocean Science (NCCOS)/Center for Coastal Monitoring
858 and Assessment (CCMA) - Biogeography Team. Silver Spring, MD. Available at:
859 <http://iobis.org/>, accessed 2013.*

860 Northeast Fisheries Science Center (2005) “Northeast Fisheries Science Center Bottom Trawl Survey
861 Data (OBIS-USA).” *NOAA’s National Marine Fisheries Service (NMFS) Northeast Fisheries
862 Science Center. Woods Hole, Massachusetts, USA. Available at: <http://iobis.org/>, accessed 2013.*

863 NTL LTER (2012) “North Temperate Lakes LTER: Fish Abundance 1981 - current.” *North
864 Temperate Lakes Long Term Ecological Research Program, Center for Limnology, University of
865 Wisconsin-Madison. Available at: [http://lter.limnology.wisc.edu/dataset/north-temperate-lakes-
866 lter-fish-abundance-1981-current](http://lter.limnology.wisc.edu/dataset/north-temperate-lakes-
866 lter-fish-abundance-1981-current), accessed 2012.*

867 NTL LTER (2011) “North Temperate Lakes LTER: Zooplankton - Madison Lakes Area 1997 -
868 current.” *North Temperate Lakes Long Term Ecological Research Program, Center for
869 Limnology, University of Wisconsin-Madison. Available at:
870 [http://lter.limnology.wisc.edu/dataset/north-temperate-lakes-lter-zooplankton-madison-lakes-
871 area-1997-current](http://lter.limnology.wisc.edu/dataset/north-temperate-lakes-lter-zooplankton-madison-lakes-
871 area-1997-current), accessed 2013.*

872 Ohmart, R., Pearson, D., Hostetler, M., Katti, M. & Hulen, T. (2003) “Transect bird survey with data
873 synthesis from multiple transects in the central Arizona-Phoenix area: period 1998 to 2000.”
874 *Central Arizona-Phoenix Long-Term Ecological Research. Global Institute for Sustainability,*
875 *Arizona State University.* Available at: <https://caplter.asu.edu/data/data-catalog/?id=43>, accessed
876 2012.

877 Oliveira, A. de “Ilha do Cardoso, Brasil.” *The Center for Tropical Forest Science. Smithsonian*
878 *Tropical Research Institute.* Available at: <http://www.ctfs.si.edu/site/Ilha+do+Cardoso/>,
879 accessed 2013.

880 Paquette, A., Laliberté, E., Bouchard, A., Blois, S. de, Legendre, P. & Brisson, J. (2007) Lac Croche
881 understory vegetation data set (1998–2006). *Ecology*, **88**, 3209–3209.

882 Pardieck, K.L., Ziolkowski Jr., D.J. & Hudson, M.-A.R. (2015) “North American Breeding Bird
883 Survey Dataset 1966 - 2014, version 2013.0.” *U.S. Geological Survey, Patuxent Wildlife*
884 *Research Center.* Available at: <https://www.pwrc.usgs.gov/bbs/RawData/>, accessed 2012.

885 Ramesh, B.R., Swaminath, M.H., Patil, S. V., Dasappa, Pélissier, R., Venugopal, P.D., Aravajy, S.,
886 Elouard, C. & Ramalingam, S. (2010) Forest stand structure and composition in 96 sites along
887 environmental gradients in the central Western Ghats of India. *Ecology*, **91**, 3118–3118.

888 Reich, P., Wedin, D., Hobbie, S. & Davis, M. “Experiment 133 - Effect of Burning Patterns on
889 Vegetation in the Fish Lake Burn Compartments - Shrub Survey.” *Cedar Creek Ecosystem*
890 *Science Reserve.* Available at: <http://www.cedarcreek.umn.edu/research/data/experiment?e133>,
891 accessed 2012.

892 Reichert, M. (2009) “MARMAP Chevron Trap Survey 1990-2009.” *SCDNR/NOAA MARMAP*
893 *Program, SCDNR MARMAP Aggregate Data Surveys, The Marine Resources Monitoring,*
894 *Assessment, and Prediction (MARMAP) Program, Marine Resources Research Institute, South*
895 *Carolina Department of Natural Resources USA.* Available at: <http://www.usgs.gov/obis-usa/>,
896 accessed 2013.

897 Reichert, M. (2010a) “MARMAP Fly Net 1990-2009.” *SCDNR/NOAA MARMAP Program, SCDNR*
898 *MARMAP Aggregate Data Surveys, The Marine Resources Monitoring, Assessment, and*
899 *Prediction (MARMAP) Program, Marine Resources Research Institute, South Carolina*

900 Department of Natural Resources USA. Available at: <http://www.usgs.gov/obis-usa/>, accessed
901 2013.

902 Reichert, M. (2010b) "MARMAP Yankee Trawl 1990-2009." *SCDNR/NOAA MARMAP Program,*
903 *SCDNR MARMAP Aggregate data surveys, The Marine Resources Monitoring, Assessment, and*
904 *Prediction (MARMAP) Program, Marine Resources Research Institute, South Carolina*
905 *Department of Natural Resources USA.* Available at: <http://www.usgs.gov/obis-usa/>, accessed
906 2013.

907 Sarnthein, M., van Kreveld, S.A., Erlenkeuser, H., Grootes, P.M., Kucera, M., Pflaumann, U. &
908 Schulz, M. (2003) "Distribution of foraminifera of sediment core GIK23258-2".
909 doi:10.1594/PANGAEA.114682, In Supplement to: Sarnthein et al. (2003): Centennial-to-
910 millennial-scale periodicities of Holocene climate and sediment injections off western Barents
911 shelf, 75°N. *Boreas*, 32(3), 447-461, doi:10.1111/j.1502-3885.2003.tb01227.x. *PANGAEA.*
912 Available at: <http://www.gbif.org/dataset/8750386c-f762-11e1-a439-00145eb45e9a>, accessed
913 2012.

914 Shochat, E., Katti, M. & Warren, P. (2004) "Point count bird censusing: long-term monitoring of bird
915 distribution and diversity in central Arizona-Phoenix: period 2000 to 2011." *Central Arizona-*
916 *Phoenix Long-Term Ecological Research. Global Institute for Sustainability, Arizona State*
917 *University.* Available at: <https://caplter.asu.edu/data/data-catalog/?id=46>, accessed 2012.

918 Silveira, F.L. & Lopes, R.M. (2011) "Previous fisheries REVIZEE Program." *WSAOBIS.* Available
919 at: <http://iobis.org/>, accessed 2013.

920 Sipe, T., Bowden, R. & McClaugherty, C. (2009) "Tree Seedlings in CRUI Land Use Project at
921 Harvard Forest 1996. Harvard Forest Data Archive: HF141." *The Harvard Forest Long Term*
922 *Ecological Research program.* Available at:
923 <http://harvardforest.fas.harvard.edu:8080/exist/xquery/data.xq?id=hf141>, accessed 2012.

924 Stevens, M. (2010) "Trekvis - Migratory fishes in the river Scheldt." *Research Institute for Nature*
925 *and Forest (INBO).* Available at: [http://www.gbif.org/dataset/b2d0f29e-4614-4001-93c8-](http://www.gbif.org/dataset/b2d0f29e-4614-4001-93c8-f651878a86d2)
926 [f651878a86d2](http://www.gbif.org/dataset/b2d0f29e-4614-4001-93c8-f651878a86d2), accessed 2014.

927 Su, S., Chang-Yang, C., Lu, C., Tsui, C., Lin, T., Lin, C., Chiou, W., Kuan, L., Chen, Z. & Hsieh, C.

928 (2007) *Fushan subtropical forest dynamics plot: tree species characteristics and distribution*
929 *patterns*, Taipei.

930 Sukumar, R. “Mudumalai Forest Dynamics Plot, India.” *The Center for Tropical Forest Science.*
931 *Smithsonian Tropical Research Institute*. Available at: <http://www.ctfs.si.edu/site/Mudumalai/>,
932 accessed 2013.

933 Sukumar, R., Suresh, H.S., Dattaraja, H.S., John, R. & Joshi, N. V. (2004) *Mudumalai Forest*
934 *Dynamics Plot, India. Forest Diversity and Dynamism: Findings from a Large-Scale Plot*
935 *Network* (ed. by E.C. Losos and E.G.J. Leigh), pp. 551–563. University of Chicago Press,
936 Chicago.

937 Sun, I.F. & Hsieh, C.-F. “Fushan Forest Dynamics Plot, Taiwan.” *The Center for Tropical Forest*
938 *Science. Smithsonian Tropical Research Institute*. Available at:
939 <http://www.ctfs.si.edu/site/Fushan/>, accessed 2013.

940 Tan, S., Davies, S., Yamakura, T. & Itoh, A. “Lambir Hills Forest Dynamics Plot, Malaysia.” *The*
941 *Center for Tropical Forest Science. Smithsonian Tropical Research Institute*. Available at:
942 <http://www.ctfs.si.edu/site/Lambir/>, accessed 2013.

943 The Swedish University of Agriculture Sciences “National Forest Inventory (SLU).” *GBIF-Sweden.*
944 Available at: <http://www.gbif.org/dataset/c46708d0-12aa-11dd-9ff0-b8a03c50a862>, accessed
945 2013.

946 Thomas, D., Kenfack, D. & Chuyong, G. “Korup Forest Dynamics Plot, Cameroon.” *The Center for*
947 *Tropical Forest Science. Smithsonian Tropical Research Institute*. Available at:
948 <http://www.ctfs.si.edu/site/Korup/>, accessed 2013.

949 Thomas, D.W., Kenfack, D., Chuyong, G.B., Sainge, N.M., Losos, E.C., Condit, R.S. & Songwe,
950 N.C. (2003) *Tree Species of Southwestern Cameroon: Tree distribution maps, diameter tables*
951 *and species documentation of the 50-ha Korup Forest Dynamics Plot*. Center for Tropical Forest
952 *Science*, Washington, D.C.

953 Trexler, J. (2007) “Consumer Stocks: Fish, Vegetation, and other Non-physical Data from Everglades
954 National Park (FCE), South Florida from February 2000 to Present.” *Florida Coastal*
955 *Everglades LTER Program*. Available at:

956 http://fcelter.fiu.edu/data/core/metadata/EML/?datasetid=LT_CD_Trexler_001, accessed 2012.

957 U.S.D.A. Forest Service (2007) “U.S.D.A. Forest Service, Redwood Sciences Laboratory - Lamna
958 Point Count.” *Avian Knowledge Network*. Available at: [http://www.gbif.org/dataset/864da4c2-
959 f762-11e1-a439-00145eb45e9a](http://www.gbif.org/dataset/864da4c2-f762-11e1-a439-00145eb45e9a), accessed 2012.

960 USFS “Landbird Monitoring Program (UMT-LBMP).” *US Forest Service*. Available at:
961 <http://www.avianknowledge.net/>, accessed 2012.

962 Valencia, R. “Yasuni Forest Dynamics Plot, Ecuador.” *The Center for Tropical Forest Science*.
963 *Smithsonian Tropical Research Institute*. Available at: <http://www.ctfs.si.edu/site/Yasuni/>,
964 accessed 2013.

965 Valencia, R., Condit, R., Foster, R.B., Romoleroux, K., Muñoz, G.V., Svenning, J.-C., Magård, E.,
966 Bass, M., Losos, E.C. & Balslev, H. (2004) *Yasuni Forest Dynamics Plot, Ecuador. Forest
967 Diversity and Dynamism: Findings from a Large-Scale Plot Network* (ed. by E.C. Losos and
968 E.G.J. Leigh), pp. 609–620. University of Chicago Press, Chicago.

969 Vallejo, M.I., Samper, C., Mendoza, H. & Otero, J.T. (2004) *La Planada Forest Dynamics Plot,*
970 *Colombia. Forest Diversity and Dynamism: Findings from a Large-Scale Plot Network* (ed. by
971 E.C. Losos and E.G.J. Leigh), pp. 517–526. University of Chicago Press, Chicago.

972 Viereck, L.A., Van Cleve, K., Chapin, F.S.S., Ruess, R.W. & Hollingsworth, T.N. (2005) “Vegetation
973 Plots of the Bonanza Creek LTER Control Plots: Species Count (1975 - 2004).” *Bonanza Creek
974 LTER - University of Alaska Fairbanks*. Available at:
975 http://www.lter.uaf.edu/data_detail.cfm?datafile_pkey=175, accessed 2012.

976 Wade, E.J. (2011) “Snow crab research trawl survey database (Southern Gulf of St. Lawrence, Gulf
977 region, Canada) from 1988 to 2010.” *OBIS Canada Digital Collections*. *OBIS Canada, Bedford
978 Institute of Oceanography, Dartmouth, Nova Scotia, Canada*. Available at: <http://iobis.org/>,
979 accessed 2013.

980 Walker, J., Shochat, E., Katti, M. & Warren, P. (2004) “Point Count Bird Censusing Data Subset for
981 Paper ‘Effects of land use and vegetation cover on bird communities’ Walker et. al.” *Central
982 Arizona-Phoenix Long-Term Ecological Research*. *Global Institute for Sustainability, Arizona
983 State University*. Available at: <https://caplter.asu.edu/data/data-catalog/?id=394>, accessed 2012.

984 Warren, P., Kinzig, A., Martin, C. & Machabee, L. (2005) “Ecological and social Interactions in
985 urban parks: bird surveys in local parks in the central Arizona-Phoenix metropolitan area.”
986 *Central Arizona-Phoenix Long-Term Ecological Research. Global Institute for Sustainability,*
987 *Arizona State University.* Available at: <https://caplter.asu.edu/data/data-catalog/?id=256>,
988 accessed 2012.

989 Williams, D. “Pelagic Fish Observations 1968-1999.” *Australian Antarctic Data Centre.* Available at:
990 <http://www.gbif.org/dataset/85b0a82a-f762-11e1-a439-00145eb45e9a>, accessed 2012.

991 Woehler, E. (1999, updated 2015) “Distribution and abundance of seabirds in the Southern Indian
992 Ocean, 1978/1979+”. *Australian Antarctic Data Centre* - doi:10.4225/15/5643E8C0743C2.
993 Available at <http://www.iobis.org>, accessed 2013.

994 Woods, K.D. (2009) Multi-decade, spatially explicit population studies of canopy dynamics in
995 Michigan old-growth forests. *Ecology*, **90**, 3587–3587.

996 Zimmerman, J. “LFDP phenology plot seedlings – 16 ha plot.” *Luquillo Long Term Ecological*
997 *Research Program.* Available at: <http://luq.lternet.edu/data/luqmetadata175>, accessed 2014.

998 Zimmerman, J. & Brokaw, N. “Census of species, diameter and location at the Luquillo Forest
999 Dynamics Plot (LFDP), Puerto Rico.” *Luquillo Long Term Ecological Research Program.*
1000 Available at: <http://luq.lternet.edu/data/luqmetadata119>, accessed 2013.

1001 Zimmerman, J., Thompson, J. & Brokaw, N. “Luquillo Forest Dynamics Plot, Puerto Rico.” *The*
1002 *Center for Tropical Forest Science. Smithsonian Tropical Research Institute.* Available at:
1003 <http://www.ctfs.si.edu/site/Luquillo/>, accessed 2013.

1004 Zimmerman, J.K., Comita, L.S., Thompson, J., Uriarte, M. & Brokaw, N. (2010) Patch dynamics and
1005 community metastability of a subtropical forest: compound effects of natural disturbance and
1006 human land use. *Landscape Ecology*, **25**, 1099–1111.

1007 **List of tables**

1008 Table 1. Binomial and multinomial Bayesian generalised linear model fitting results, showing a
 1009 positive effect of spatial scale or higher taxonomic breadth on the prevalence of multimodality. For
 1010 the Bayesian GLM, the posterior mean estimates, the 95% credible intervals and the pMCMC
 1011 (MCMC p-values) values are shown. The parameter estimates were considered statistically significant
 1012 when pMCMC values < 0.05, and the 95% credible intervals did not include 0. The term
 1013 “Multimodality : SpatialExtent.Continental : NumberFamilies” refers to the estimation of
 1014 multimodality vs 1PLN at the Continental scale with the interaction with number of families.

Binomial GLM	Estimate	Std. Error	z value	Pr(> z)
SpatialExtent.Continental	-0.2207	1.0135	-0.2180	0.8277
SpatialExtent.Regional	-1.5511	1.2830	-1.2090	0.2267
SpatialExtent.Local	-3.7396	1.3940	-2.6830	0.0073
NumberFamilies	-0.0127	0.0154	-0.8230	0.4105
SpatialExtent.Regional : NumberFamilies	0.0060	0.0247	0.2430	0.8084
SpatialExtent.Local : NumberFamilies	0.0747	0.0260	2.8730	0.0041

Multinomial GLM	Posterior mean	Lower 95% CI	Upper 95% CI	pMCMC
Reference: 1PLN-SpatialExtent.Continental				
Multimodality : SpatialExtent.Continental	-0.0013	-2.4390	2.3690	0.9930
Multimodality : SpatialExtent.Regional	0.0183	-3.1700	3.2370	0.9944
Multimodality : SpatialExtent.Local	-3.7030	-6.9170	-0.5603	0.0194
Multimodality : SpatialExtent.Continental : NumberFamilies	-0.0248	-0.0669	0.0098	0.1597
Multimodality : SpatialExtent.Regional : NumberFamilies	-0.0124	-0.0805	0.0541	0.7079
Multimodality : SpatialExtent.Local : NumberFamilies	0.0897	0.0301	0.1519	0.0011
Logser : SpatialExtent.Continental	-186.0000	-336.7000	-0.4936	0.0164
Logser : SpatialExtent.Regional	188.1000	1.9480	338.0000	0.0092
Logser : SpatialExtent.Local	186.2000	1.4960	337.8000	0.0158
Logser : SpatialExtent.Continental : NumberFamilies	-1.7840	-4.5770	0.7772	0.3401
Logser : SpatialExtent.Regional : NumberFamilies	1.7330	-0.8160	4.5390	0.3565
Logser : SpatialExtent.Local : NumberFamilies	1.7650	-0.8011	4.5560	0.3454

1015 **List of figures**

1016

1017 Figure 1. Examples of a logseries distribution, a single Lognormal Poisson (1PLN), and mixtures of
1018 two and three Lognormal Poisson distributions (2PLN and 3PLN, respectively). For the logseries
1019 distribution, the single parameter is Fisher's alpha. For the PLN models, μ and σ are the mean and
1020 standard deviation of log-abundance for one of the underlying lognormal community abundance
1021 distributions (one pair of parameters for each mode), and ρ_n is the probability that a species comes
1022 from distribution n . The parameters used to generate the sampled abundance data for each distribution
1023 are shown, and the species pool size was $S=100$ (the code to generate the 2PLN example can be found
1024 in Appendix S2).

1025

1026 Figure 2. Map showing the 117 empirical SADs sampling locations and the model selected as best fit
1027 (each point corresponds to the mean latitude-longitude).

1028

1029 Figure 3. Species abundance distributions (SADs) of the empirical datasets selected as multimodal
1030 with high confidence, identified by the corresponding ID. All the fitting routines were run on non-
1031 binned data. SADs were plotted with bins representing true doubling classes of abundance, following
1032 Gray *et al.* (2006). For all SADs the yy axis is the number of species and the xx axis is the species
1033 abundance in \log_2 classes (the first bar represents species with abundance 1, the second one species
1034 with abundances 2-3, then 4-7, 8-15, etc). The fitted curves are red line for the logseries, bold blue
1035 line for 1PLN, dashed green line for 2PLN and dotted orange line for 3PLN. The leftmost column
1036 shows SADs for birds, the second column for terrestrial plants, and the taxon is identified for the
1037 remaining SADs. The symbols indicate SADs locations: ★ in Asia, ▲ in Europe, ◆ in North
1038 America and ● in South America.

1039

1040 Figure 4. Model selection frequency vs SADs classification for spatial extent and taxonomic breadth
1041 as number of families (the xx-axis was truncated at 80 for better visualization; the 4 SADs with higher

1042 number of families not shown were best fit by 1PLN). The absolute number of SADs per spatial
1043 extent is 11 Continental, 42 Regional and 64 Local.