

# Generalized Logarithmic Species-Area Relationship resolves the Arrhenius-Gleason debate

Mark Carey (✉ [mark.carey79@gmail.com](mailto:mark.carey79@gmail.com))

University of South Australia <https://orcid.org/0000-0002-5985-8007>

John Boland

University of South Australia

Gunnar Keppel

University of South Australia

---

## Research Article

**Keywords:** Biogeography, generalized logarithm, power law, species-area relationships

**Posted Date:** April 14th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1440321/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

1 Generalized Logarithmic Species-Area Relationship

2 resolves the Arrhenius-Gleason debate

3 Mark Carey<sup>1</sup>, John Boland<sup>2</sup>, Gunnar Keppel<sup>3</sup>

4 March 15, 2022

5 1. UniSA STEM , University of South Australia, Mawson Lakes Campus, GPO Box  
6 2471, Adelaide, South Australia 5001, Australia, E-mail: Mark.Carey@unisa.edu.au (cor-  
7 respondence author)

8  
9 2. UniSA STEM, University of South Australia, Mawson Lakes Campus, GPO Box 2471,  
10 Adelaide, South Australia 5001, Australia, E-mail: John.Boland@unisa.edu.au

11  
12 3. UniSA STEM & Future Industries Institute, University of South Australia, Mawson  
13 Lakes Campus, GPO Box 2471, Adelaide, South Australia 5001, Australia, E-mail: Gun-  
14 nar.Keppel@unisa.edu.au

## 16 **Abstract**

17 The species-area relationship (SAR) is widely applied in ecology. Mathematically it is  
18 usually expressed as either a semi-log or power-law relationship, with the former being  
19 introduced by Gleason and the latter by Arrhenius. We here resolve the dispute about  
20 which form of the SAR to prefer by introducing a novel model that smoothly transforms  
21 between the Gleason semi-log (GSL) and Arrhenius power law (APL) forms. The model  
22 introduced has the form of  $\ln_q(S) = a + z \ln A$ , with  $\ln_q$  being a generalized logarithmic  
23 function which is a linear map ( $y = x$ ) for  $q = 0$  and a logarithmic map ( $y = \ln x$ ) for  
24  $q = 1$  and  $q$  can take any intermediate value between 0 and 1. We applied this model to  
25 100 datasets (mostly islands), linking species richness to island area. The APL was the  
26 preferred model in 68% of head-to-head comparisons with the GSL. Both models were  
27 supported in 40% of cases. In just under half (44%) of the cases an intermediate model  
28 best explained the data. The results demonstrate the utility of a simple intermediate SAR  
29 model. Visualizing the profile of the range of model fits for all  $q \in [0, 1]$  allows us to gain  
30 extra insight into SARs not yielded by head-to-head comparisons of GSL and APL. The  
31 mathematics related to the generalized logarithmic function introduced here promises to  
32 have application to other areas of mathematical ecology, for instance in population biol-  
33 ogy models.

34

## 35 **Keywords**

36 Biogeography, generalized logarithm, power law, species-area relationships

# 1 Introduction

One of the earliest and persistently robust observations in the field of biogeography and ecology is that the diversity of species of a taxon increases in a predictable way with the area surveyed or the total area available (i.e., as on an oceanic island). This species-area relationship (SAR) has been described as the closest thing to a rule in ecology (Lomolino, 2000). SARs have become fundamental to the understanding of patterns of biodiversity and a critical tool for predicting biodiversity loss (Connor & McCoy (1979); Triantis *et al.* (2012); Dakare *et al.*, 2006). The form and parameterisation of SARs are known to be significantly affected by sampling scheme, spatial scale, and the types of organisms or habitats involved (Lomolino, 1989; Dakare *et al.*, 2006; Fattorini *et al.*, 2017; Shrader *et al.*, 2019).

First observed as a qualitative phenomenon by naturalists such as J. R. Forster (Forster, 1778), it was originally presented in mathematical form by Arrhenius (1921) as a power law:

$$S = cA^z \tag{1}$$

where  $S$  is species richness,  $A$  is area and  $c$  and  $z$  are parameters which are determined empirically. This was almost immediately challenged by Gleason (1922) who counter-proposed a SAR with a semi-log form (hereafter GSL):

$$S = a + b \ln A. \tag{2}$$

One of Gleason's main critiques of the Arrhenius' power law SAR (hereafter APL) was that it gave "impossibly high estimates" for large areas. In practice, this is not a such problem for oceanic islands, since island size is naturally bounded and usually relatively small (one of the largest oceanic islands is the large island of Hawaii which is approxi-

59 mately 10457 km<sup>2</sup>, but more typical are the islands of the Cape Verde archipelago which  
 60 range from around 2 to 1000 km<sup>2</sup>). This consideration has, however, led to later scale-  
 61 dependant variants of the SAR, for instance the persistence model of Plotkin *et al.* (2000)  
 62 for tropical forest plots (i.e., Plotkin, 2000) and the triphasic SAR model applicable across  
 63 many orders of magnitude of scale (see Hubbell, 2001). At the scales we are interested  
 64 in for this study, this consideration is largely irrelevant. In contrast, however, of par-  
 65 ticular relevance for us here is the observation that, for many field studies, the reported  
 66 data often falls somewhere between the power law and semi-log models (i.e., Simberloff &  
 67 Gotelli, 1984; Gitay *et al.*, 1991). This suggests that a simple head-to-head comparison  
 68 between the APL and GSL models may be inadequate.

69

70 Tjørve (2012) proposed a hybrid model SAR to better fit datasets which did not cleanly  
 71 fit an APL or GSL pattern. This model involved simply multiplying the two models  
 72 together while introducing parameters to “slide” between mixed states of the GSL and  
 73 APL i.e.:

$$S = \left( c_1 + b \log A \right)^{\frac{dA}{A+n}} \times \left( c_2 A^z \right)^{1 - \frac{dA}{A+n}} \quad (3)$$

74 where  $c_1$ ,  $c_2$ ,  $b$ ,  $d$  and  $n$  are parameters.

75

76 This model has one immediate major problem, however, which is the number of free  
 77 parameters it now contains. Tjørve resolves this problem by fixing four of the six pa-  
 78 rameters,  $b$ ,  $z$ ,  $c_1$  and  $c_2$ , so that only two parameters,  $d$  and  $n$ , remained to be fitted by  
 79 regression. However, this fixing of parameters causes Equation 3 to fail to be equivalent  
 80 to the APL and GSL in the cases where one would expect it to be (see Appendix 1 in the  
 81 Supplementary Materials). This model has the advantage that it is capable of generating  
 82 sigmoid curves. Some SARs based on the contiguous expansion of sample area have been

83 observed to be sigmoid in log-log space (see Rosenzweig, 1995; Tjørve & Tjørve, 2008),  
84 although we are not concerned with those here.

85

86 In this study we present a novel method of constructing SARs intermediate between the  
87 APL and GSL forms which does not suffer from the drawbacks resulting from the inclusion  
88 of the many extra parameters necessary to define and fit the Tjørve model of Equation 3.  
89 Our approach utilises the generalised logarithmic function which has not previously been  
90 part of the mathematical ecologist's toolkit and is likely to be useful in other contexts  
91 (see Section ??). The resulting model contains only one additional free parameter and is  
92 identical to the regular APL and GSL in the limit cases ( $q = 1$  and  $q = 0$ ). We test the  
93 new approach on 100 datasets gathered from the literature, including the original datasets  
94 of Arrhenius and Gleason.

95

## 96 **2 Methods**

### 97 **2.1 The generalised logarithmic transform**

98 For a conceptual picture of this function imagine you set out walking on an infinite plain  
99 with a very large bag of pegs. After a metre you put a peg in the ground and then two  
100 metres further on you put another peg in the ground and then four metres after that you  
101 put down another peg and then eight metres, sixteen and so on and so on off into the  
102 distance as long as you have pegs and time (theoretically indefinitely, this being a thought  
103 experiment). Now imagine, having done this, you summon a Djinn to compress all of space  
104 itself so all of your pegs become 1m apart. This is a logarithmic transformation of that  
105 space (base 2 in this illustration). In imagining the last part of this thought experiment  
106 you likely visualised the space being continuously compressed. If you could stop the  
107 process at any point between the pegs being increasing powers of two metres apart and

108 being equidistant then that is what we mean by a generalised logarithmic transformation.  
 109 The parameter  $q$ , used throughout the remainder of this paper, refers conceptually to  
 110 where you halt the compression of space. Mathematically such a function (in base e) can  
 111 be defined as:

$$\ln_q(x) = \int_q^x \frac{dt}{t^q} \quad (4)$$

112 Equation 4 is a modified form of the generalised logarithmic function as given in Tsalis  
 113 (2016) with the bottom limit changed from 1 to  $q$ . This makes the transformation exact  
 114 rather than approximate (see also Martinez, 2008). This then evaluates as

$$\ln_q x = \begin{cases} \frac{x^p - q^p}{p} & \text{if } q \neq 1 \\ \ln x & \text{if } q = 1 \end{cases} \quad (5)$$

115 Where  $p = 1 - q$  and  $q \in [0, 1]$  (derivation of this given in Appendix 2). This function  
 116 smoothly transforms between a null transform ( $\ln_0 x = x$ ) and a natural logarithmic  
 117 transform ( $\ln_1 x = \ln x$ ). Equation 5 is a modification of the Box Cox transform which  
 118 is already approximately what we seek (the Box Cox transform goes from  $y = \ln x$  to  
 119  $y = x + 1$  instead of to  $y = x$  so it just slightly “misses the mark”). This is fixed by the  
 120 introduction of the  $q^p$  term. The relation of Equation 4 to the Box Cox transformation is  
 121 given in detail in Appendix 2 in the Supplementary Materials, along with a discussion of  
 122 the additional mathematical properties of this function.

## 123 **2.2 A new hybrid SAR model**

124 The GSL model can be expressed as a log-transformed version of the APL i.e.,

$$S = \ln(cA^z) = \ln c + z \ln A. \quad (6)$$

125 Therefore a function such as that defined in Equation 5 can smoothly transition from a

126 linear map to a logarithmic map, producing a very simple hybrid SAR model i.e.,

$$S = \ln_q(cA^z). \quad (7)$$

127 This hybrid SAR only has a single free parameter,  $q$ , which smoothly and exactly trans-  
128 forms between the APL and the GSL forms. Having a single free parameter which tracks  
129 how we “slide” from the APL to the GSL also has the distinct advantage that we can  
130 then plot on a simple two dimensional graph, the  $q$  value against the goodness of fit of  
131 the model (i.e., AIC or BIC), which gives us a way to visualise the model fits against the  
132 data in a novel way.

133

## 134 **2.3 Data**

135 We analysed 100 datasets for which at least one of either the APL or GSL were statis-  
136 tically significant. Data was collated from GIFT database and from datasets previously  
137 published in the literature. A full list of the sources of the data used is provided in Ap-  
138 pendix 3 in the Supplementary Materials. For Arrhenius and Gleason’s original datasets  
139 (Arrhenius, 1921; Gleason, 1922) we present the analysis in detail, since these were the  
140 data that were the historical context for the original APL vs GSL debate. Arrhenius used  
141 species counts for plant associations of different types lying in the islands of Stockholm,  
142 sampled areas increasing by square decimetres up to 100, except for weed association  
143 species where the maximum area was 300 square decimetres. Gleason, by contrast, used  
144 species counts of a series of scattered and contiguous quadrats for aspen associations in  
145 North Michigan. In Arrhenius’s data set, the *Pinus* woody species were aggregated, sim-  
146 ilar to Tjørve (2012), and the weed association species considered separately.

147

## 148 2.4 Analysis

149 The form of Equation 7 which we actually test against the data is

$$\ln_q S = z \ln A + d \quad (8)$$

150 where  $d = \ln c$ . This model (hereafter referred to as the SqA model, or SqA family of  
151 models) can be fitted by simple linear regression. It is worth noting that the  $q$  here is not  
152 the same  $q$  as in Equation 7 but they are related. See Appendix 2 in the Supplementary  
153 Materials for a full discussion of this along with additional details on the algebra of the  
154 generalized logarithmic function. Henceforth when we refer to the  $q$  value of an SqA  
155 model we are referring to the  $q$  in Equation 8.

156

157 Models were compared using the second-order bias correction to the Akaike Information  
158 Criterion ( $AICc$ ) recommended for small samples (Burnham & Anderson, 2003). When  
159 using  $AICc$ , the model with the lowest value is considered to be the best, that is closest  
160 to the unknown “true” model. Models within an absolute difference ( $\Delta AICc$ ) of one or  
161 two units are usually considered to be indistinguishable in statistical power (Sakamoto *et*  
162 *al.* 1986; Triantis *et al.* 2003). Thus we consider a model to be equally well supported by  
163 the data as another model if it had  $\Delta AIC < 2$  following Triantis *et al.* (2003).

164

165 In Equation 8 we are continuously transforming the response variable as we vary  $q$ . This  
166 poses a novel problem if we want to compare models by  $AICc$ , since we are comparing the  
167 residuals of a model in spaces which are being differently deformed relative to each other  
168 so  $AICc$  values will not be directly comparable in any case except where the  $q$  values of the  
169 models are equal. There is a way we can correct for this problem and recover useful  $AICc$   
170 comparisons between models in this context, however. This is by employing a Jacobian  
171 term as advised by Akaike (1978). This Jacobian transformation, the same one used when

172 changing variables in calculus, is multiplied by the likelihood function when comparing  
173 models in different co-ordinate systems by likelihood. Since AIC is constructed from the  
174 log-likelihood, however, this turns out to be a simple additive correction:

$$AICcQ = AICc + 2q \left( \sum_i \ln s_i \right). \quad (9)$$

175 where  $s_i$  is the number of species in the  $i$ th sample or island and  $q$  is the  $q$  in Equation  
176 8. The novel part here is how the  $q$  term is incorporated. Fortunately this turns out to be  
177 as simple an intuitive as we could have wished it to be (i.e., we just multiply the usual  
178 Jacobian correction by  $q$ ).

179

180 Having established this criterion for model comparison, we considered the interval  $q \in$   
181  $[0, 1]$ , divided this interval into 100 segments and ran a regression for Equation 8 for each  
182 101  $q$  values in this range (i.e., including  $q = 0$  which corresponds to the GSL model)  
183 using *fitlm* in MATLAB R219b. For each regression we extracted the AICc and the  
184 p-values of the slope and the intercept from model structure outputs. We then plotted  
185 the  $q$  values for models against the AICc values for only those values of  $q$  which generated  
186 a statistically significant model. A model was deemed statistically significant if both the  
187 slope and the intercept were statistically significant at the 95% level in a space where the  
188 model is a linear fit to the data (i.e., log-log space for the APL and log-linear space for  
189 the GSL). In addition, for each dataset we stored the  $q$  value for the optimal SqA model,  
190 as well as the range of  $q$  values which represent the spread of models which were equally  
191 well supported by AICc. The graphs thus generated i.e., in Figures 1 and 2 give the AICc  
192 profile for all statistically significant SqA models as a continuous curve interpolated from  
193 101 equidistant model fits.

194

195 In the Results we give overall statistics for all model fits. The  $q$  versus AICc graphs are  
196 only given there for the Arrhenius and Gleason datasets. However, these graphs are given

197 for the remainder of the model fits in Appendix 4 in the Supplementary Materials.

198

## 199 **3 Results**

### 200 **3.1 Arrhenius Stockholm dataset**

201 For the (aggregated) *Pinus* wood species in Arrhenius' study we have the optimal SqA  
202 model at  $0.5 < q < 0.8$ . All SqA models are statistically significant  $\forall q$ . For the weed  
203 association species, the optimal SqA model was in the range  $0.75 < q < 0.95$ . There was  
204 a statistical significance cutoff at around  $q = 0.4$ , that is no SqA models with  $q < 0.4$   
205 were statistically significant. If only the APL and GSL were compared, the APL would  
206 be preferred for both datasets.

207

### 208 **3.2 Gleason North Michigan dataset**

209 The scattered quadrats clearly show a semi-log SAR ( $q = 0$ ). For the contiguous quadrats  
210 the power law is actually preferred on a head to head comparison, although the optimum  
211 model was  $q = 0.79$  which is better supported ( $\Delta AICc > 2$ ) statistically than the APL  
212 model.

213

### 214 **3.3 Example model fit graphs**

215 Examples of the model fits against the data are shown in Figure 3, where we see how the  
216 APL, the GSL and the optimum SqA SAR fit the real data. In the case of Arrhenius'  
217 *Pinus* woody species (Figure 3 A), the raw data more strongly supported the APL. For  
218 Gleason's contiguous quadrats of aspens the data more strongly supported an interme-  
219 diate model, although the GSL would still have registered as statistically significant if it

220 was the only model tested. In both cases, the optimal SqA model was a better fit for the  
221 data than either the APL or the GSL.

222

### 223 **3.4 Statistics for model fits over the 100 datasets**

224 The APL was statistically significant for 90 % of the datasets and the GSL was statisti-  
225 cally significant for 50 % of the datasets (the datasets were chosen such that either the  
226 APL or the GSL were statistically significant for all 100 datasets). On head to head  
227 comparisons (i.e., if the intermediate models are not considered), the APL is preferred in  
228 the majority (68 %) of cases. The GSL was preferred in 26 % of cases. In 6 % of cases the  
229 APL and GSL models were equally well supported by  $AICc$  ( $\Delta AICc < 2$ ). Both models  
230 were at least statistically significant (if not preferred) in 40 % of cases.

231

232 The ranges of  $q$  values for statistically equally well supported models are shown in Figure  
233 4. We see here that there are a majority of datasets where an intermediate model would  
234 be preferred but also that the APL would be preferred in the majority of cases in a head-  
235 to-head comparison the GSL. There are a number of cases for which we have  $q = 1$  as the  
236 optimal model but hardly any for which we have  $q = 0$  as the optimal model. Overall,  
237 higher  $q$  values predominated (mean 0.7, median 0.76). The range of  $q$  values representing  
238 SqA models of  $\Delta AICc < 2$  versus the optimum model was, however, rather large (mean  
239 0.4, median 0.3). This is illustrated in Figure 5. In 44 % of cases neither the APL or  
240 the GSL models were contained within the  $\Delta AICc < 2$  interval around the optimal SqA  
241 model.

242

## 243 4 Discussion

244 When comparing the APL and GSL head-to-head, two results suggest that an interme-  
245 diate model might perform better than either form of the SAR in some cases. Firstly,  
246 both models were statistically significant for 40 % of the datasets, and these cases could  
247 therefore count as support for either model, if only the APL and the GSL were consid-  
248 ered. Secondly, in only a few (6 %) of the cases were the APL and GSL equally well  
249 supported by AICc. The first point is well illustrated by fits for the Aspen association  
250 data of Gleason. If taken together, the contiguous and scattered quadrats show statistical  
251 support for the GSL. However, the contiguous quadrats are actually better fit by the APL.

252

253 The results of the model fits for the 100 datasets fall into three main categories of interest:

- 254 1. where either the APL or GSL would be clearly preferred on a head-to-head com-  
255 parison and one of them is the optimal model.
- 256 2. where both are equally well supported by AICc.
- 257 3. where an intermediate model is clearly preferred.

258 In just under half (44%) of cases, a hybrid model ( $q \neq 1, q \neq 0$ ) was preferred and  
259 neither the GSL or the APL were equally well supported as the optimal model by AICc  
260 ( $\Delta AICc < 2$ ). Thus we find that we concur with Tjørve (2012) in that, for a large  
261 proportion of the datasets in the literature, the best fit is a model somewhere between  
262 the APL and GSL. Indeed, 3 of the 4 original datasets used to derive the APL and  
263 GSL models are better described by intermediate models. If we take averages over the  
264 datasets, however, then we do see broad support for the general practise of using the APL  
265 as the default SAR model. The mean of  $q = 0.7$  together with a wide average range of  
266  $\Delta q \approx 0.4$  of models which were considered as good as the optimum model ( $\Delta AICc < 2$ )  
267 indicates that the APL should at least show no substantial lack of fit in most cases. This

268 is consistent with the findings of Conner and McCoy (1979) who found that of the 100  
269 datasets they examined the APL showed no substantial lack of fit for 75 of those datasets.

270

## 271 **4.1 Intermediate species-area relationships and distributions and** 272 **the relationship to underlying relative abundance distribu-** 273 **tions**

274 Beginning with Preston (1962) and summarised by May (1975), various scholars have  
275 noted that there is a theoretical mathematical relationship between underlying abun-  
276 dance distributions (SADs) and the emergent SARs. Key to these arguments is that one  
277 assumes that the same underlying SAD applies at different scales and the SAR is then  
278 obtained by regressing across data at these different scales (all SARs are inferred by re-  
279 gression across samples or islands of different sizes). Using this assumption the log-series  
280 SADs and semi-log SARs and lognormal type SADs and power law SARs can be con-  
281 nected, in that the SAR can be taken as a manifestation of the underlying SAD(s) at a  
282 larger but more coarse grained scale. But what if the underlying distributions on different  
283 scales are mixed (some log-series-like and some lognormal-like)? One interpretation of the  
284  $q$  value of a SAR is that it might represent the degree of mixture of underlying abundance  
285 distributions which we are regressing across to determine our SAR. This is illustrated in  
286 Figure 6.

287

## 288 **4.2 A remark on scattered versus contiguous quadrats**

289 It is the scattered quadrats, rather than the contiguous quadrats, which furnish the clear-  
290 est justification for the GSL in Gleason's dataset. In Arrhenius's 1923 reply to Gleason's  
291 original paper (Arrhenius, 1923) he noted that Gleason's contiguous quadrats are well

292 fitted by his own model, as we independently verified in this study. It is also relevant that  
293 Gleason (1922) stated that he expected the accumulation of species with area to have a  
294 different character for scattered rather than contiguous quadrats.

295

296 Following, and expanding on, Gleason's logic, we can expect of contiguous quadrats that  
297 they should result in a relatively slow but long-continued increase in species, whereas  
298 with scattered quadrats rarer species throughout the whole sample should appear earlier  
299 in the count, resulting in a rapid initial rise followed by an early decline in increase on new  
300 species counted. In its most general form the argument is as follows: With contiguous  
301 quadrats, as surveyed area increases then rare species in the first few quadrats which may  
302 also be present in the next few quadrats have already been counted (since the contiguous  
303 quadrats are necessary close together the first few are likely to share species both rare and  
304 common). More distant rare species are only encountered later. The situation is different  
305 for scattered quadrats. Gleason spaced his scattered quadrats as far apart as he could  
306 within his experimental design. This is most clearly stated in his 1925 paper following  
307 Arrhenius's 1923 reply (Gleason, 1925). The first few quadrats which contribute to the  
308 species count in a scattered quadrat regime are less likely to share rare species, so the  
309 species count will initially increase faster. Counting in this manner (more similar to ran-  
310 dom sampling) we are more likely to more quickly converge representative estimate of the  
311 overall number of species in the surveyed areas. Thus, the observed increase of species  
312 with area, and hence the apparent form of the SAR, will change with the sampling design.  
313 Figure 7 illustrates the expected increase of species with area under the GSL versus the  
314 APL. From these curves and the above arguments, we can see how the progression of the  
315 species count (with increasing area) under a regime of scattered quadrats will more likely  
316 lead us to infer the GSL, whereas the progression of the species count under a regime of  
317 contiguous quadrats will more likely lead us to infer the APL.

318

### 319 **4.3 Application of generalised logarithms beyond species-area** 320 **relationships**

321 The inverse of the generalized logarithmic function is the generalized exponential function:

$$\exp_q = (px + q^p)^{1/p}. \quad (10)$$

322 this can be used to construct a random variable based on an exponential random vari-  
323 able but with a variable bias i.e., a random variable which smoothly transforms from an  
324 exponential random variable to a uniform random variable based on the value of a single  
325 parameter ( $q$ ) :

$$X = \frac{\exp_q(U)}{\exp_q(1)} \quad (11)$$

326 (or  $X = 1 - \frac{\exp_q(U)}{\exp_q(1)}$ ) where  $U$  is the uniform distribution.

327

328 This has applications in stochastic modelling. A specific application of this which relates  
329 to population biology is the unification of niche apportionment models (Carey, 2021).  
330 Robert MacArthur (MacArthur, 1957, 1960) introduced a simple stochastic model of how  
331 the total niche space could be divided stochastically to model species abundance distri-  
332 butions. Tokeshi (1990, 1996) provided an expanded classification of such models which  
333 could reproduce observed abundance distributions. The adjustable random variable de-  
334 fined in Equation 11 can be inserted into these stochastic models such that they all become  
335 aspects of a single model (Carey, 2021).

336

337 **5 Conclusion**

338 If we just compare the APL and the GSL models head to head then the APL will be pre-  
 339 ferred for the majority, but by no means all of cases. The GSL is preferred or equally well  
 340 supported in enough cases to make it a useful alternative model if we were just comparing  
 341 the two models head-to-head. In just under half of cases however, we have an intermediate  
 342 model ( $q \neq 0, 1$ ) that is better supported statistically than either the GSL or APL. This  
 343 strongly implies the utility of an intermediate SAR model. This is also theoretically ne-  
 344 cessitated, if we consider that regressing on underlying species abundance distributions of  
 345 different types should yield intermediate SARs. The particular intermediate SAR model  
 346 introduced here improves functionally and aesthetically on previous integration attempts.  
 347 In addition, the generalised logarithmic function that we introduce here and the algebra  
 348 associated with it should have much wider application to other problems in mathematical  
 349 ecology and biology.

350

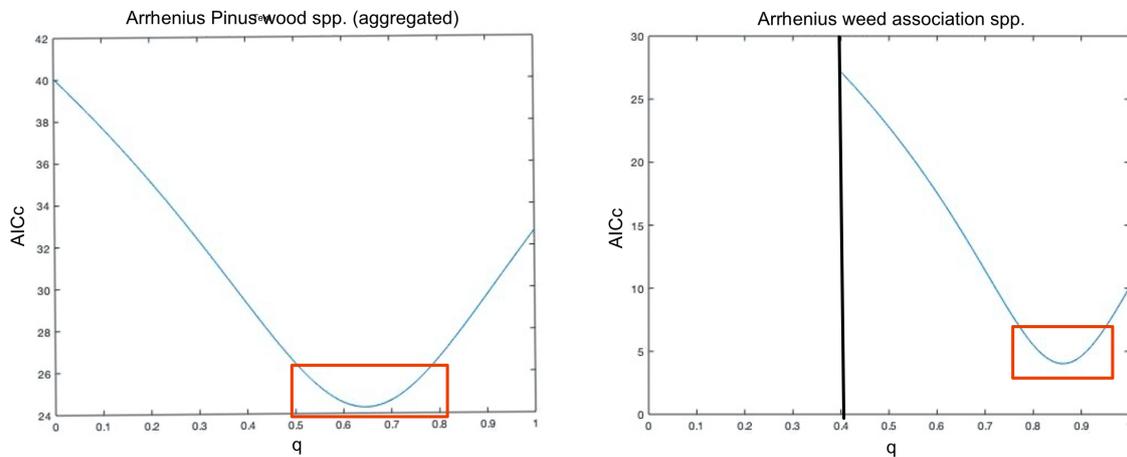


Figure 1: SqA model fits for *Pinus* wood species and weed association species from Arrhenius’s 1921 study of sites located in the islands of Stockholm. The red boxes indicate the range of SqA models which were within  $\Delta AICc < 2$  of the optimal model. The black line on right hand side graph of weed association species indicates the cut off for statistically significant models, with no SqA models statistically significant for with  $q < 0.4$

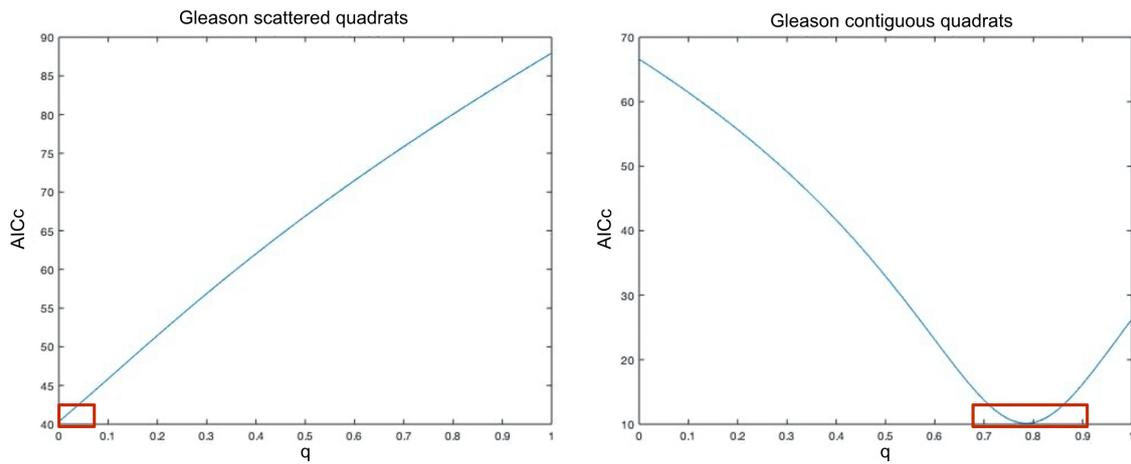


Figure 2: SqA model fits for scattered and contiguous quadrats for aspen associations in North Michigan from Gleason’s 1922 study. The red boxes indicate the range of models which were within  $\Delta AICc < 2$  of the optimal model.

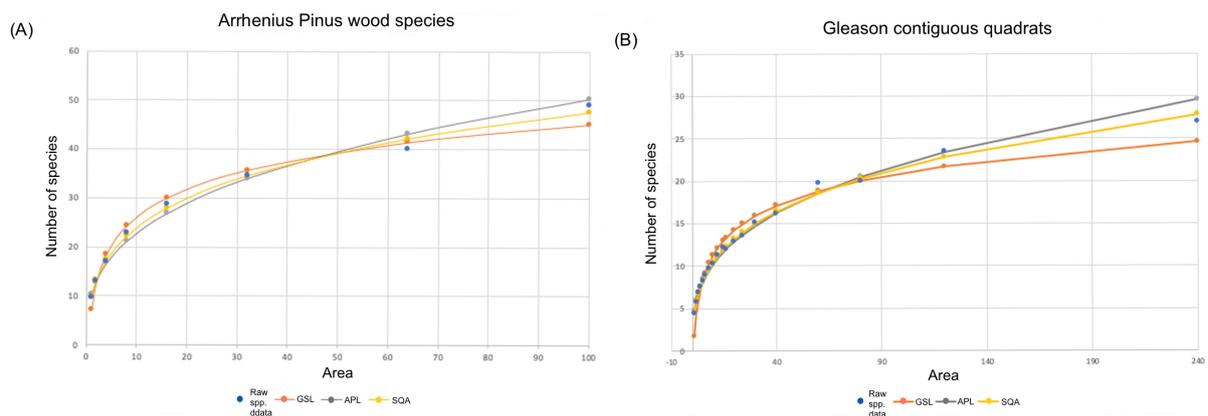


Figure 3: Graphs illustrating the model fits for the Arrhenius power law (APL = grey), Gleason semi-log (GSL = orange) and optimal  $\ln_q S = d + \ln A$  (SqA = yellow) SAR model fits against the raw data (blue dots) for (A) Arrhenius’s *Pinus* woody species data and (B) Gleason’s contiguous quadrats aspen association data. For the Arrhenius *Pinus* woody species, in linear space, residual sum of squares (RSS) for the  $GSL = 31.411$ ,  $APL = 18.627$ , optimal  $SqA = 8.074$ . For the Gleason contiguous quadrats, RSS for  $GSL = 28.821$ ,  $APL = 8.925$ , optimal  $SqA = 3.070$ .

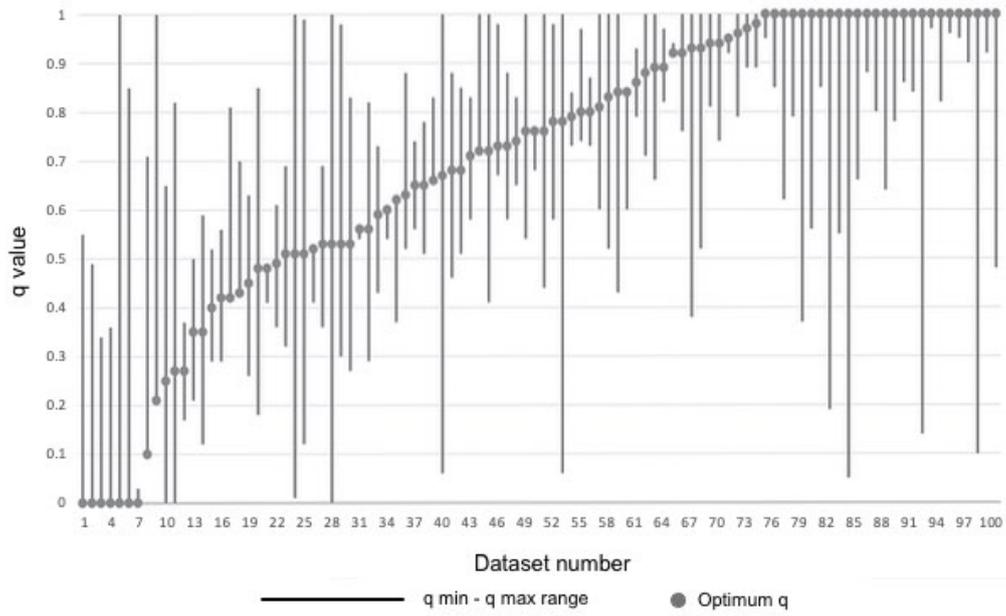


Figure 4: Optimum  $q$  values and  $q$  ranges for models equally supported by AICc ( $\Delta AICc < 2$  versus the optimal model) for models of the type  $\ln_q S = d + z \ln A$  for 100 datasets gathered from the literature. Datasets arranged in order of increasing optimum  $q$  value.

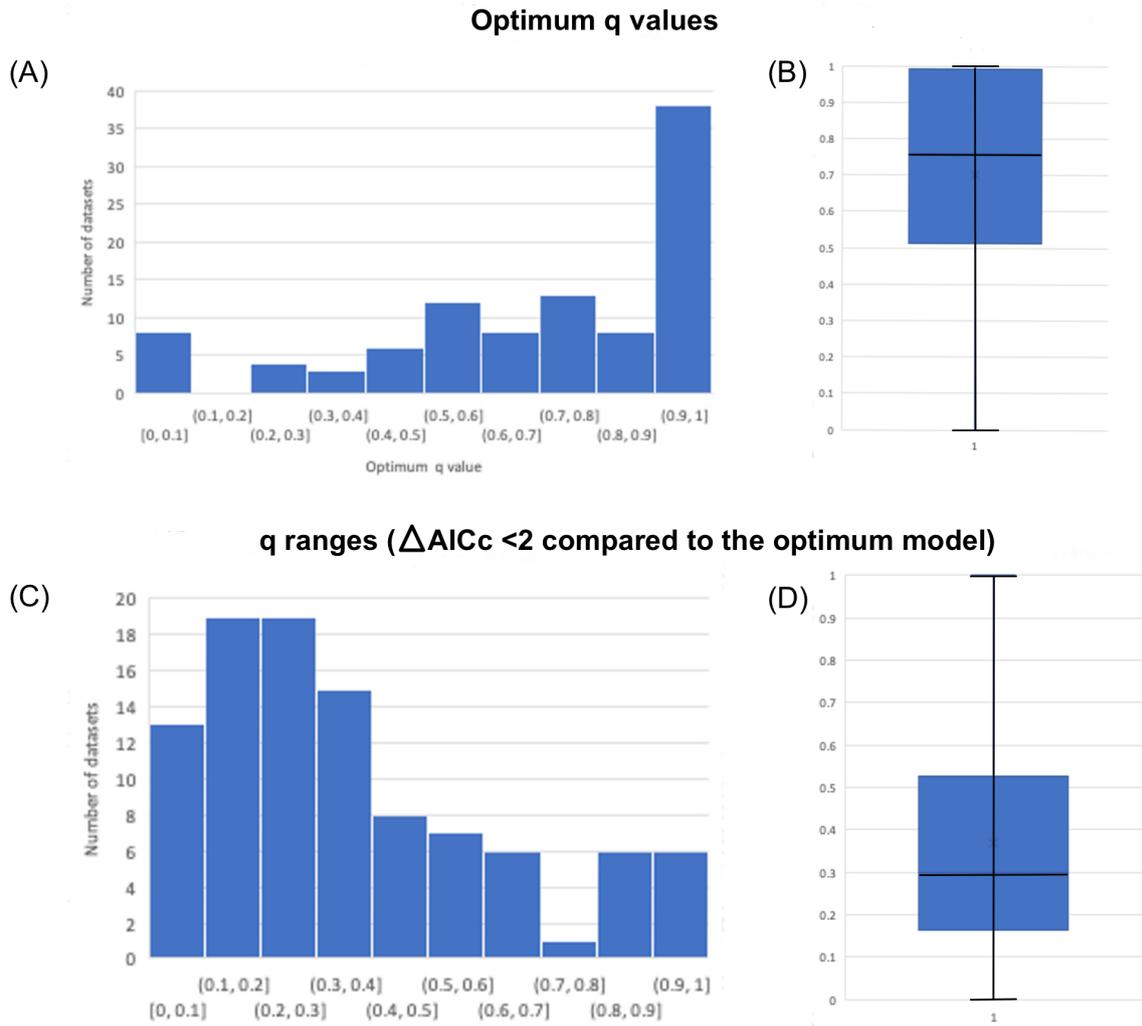


Figure 5: Graphs of summary statistics for models of the type  $\ln_q S = d + z \ln A$  for 100 datasets gathered from the literature. (A) and (B) show the distribution of  $q$  values for the optimal SqA models (lowest AICc) and (C) and (D) show the distribution ranges of  $q$  values for SqA models which were statistically equivalent ( $\Delta AICc < 2$  versus the optimal model). For the optimal  $q$  values we have Mean = 0.70, Standard Deviation = 0.30, Median = 0.76 Q1 = 0.52, Q3 = 0.99 and Interquartile Range = 0.47. The ranges of  $q$  values for “equally good” models ( $\Delta AICc < 2$ ) around the optimum model had Mean = 0.37, Standard Deviation = 0.27, Median = 0.30, Q1 = 0.17, Q3 = 0.52 and Interquartile Range = 0.36.

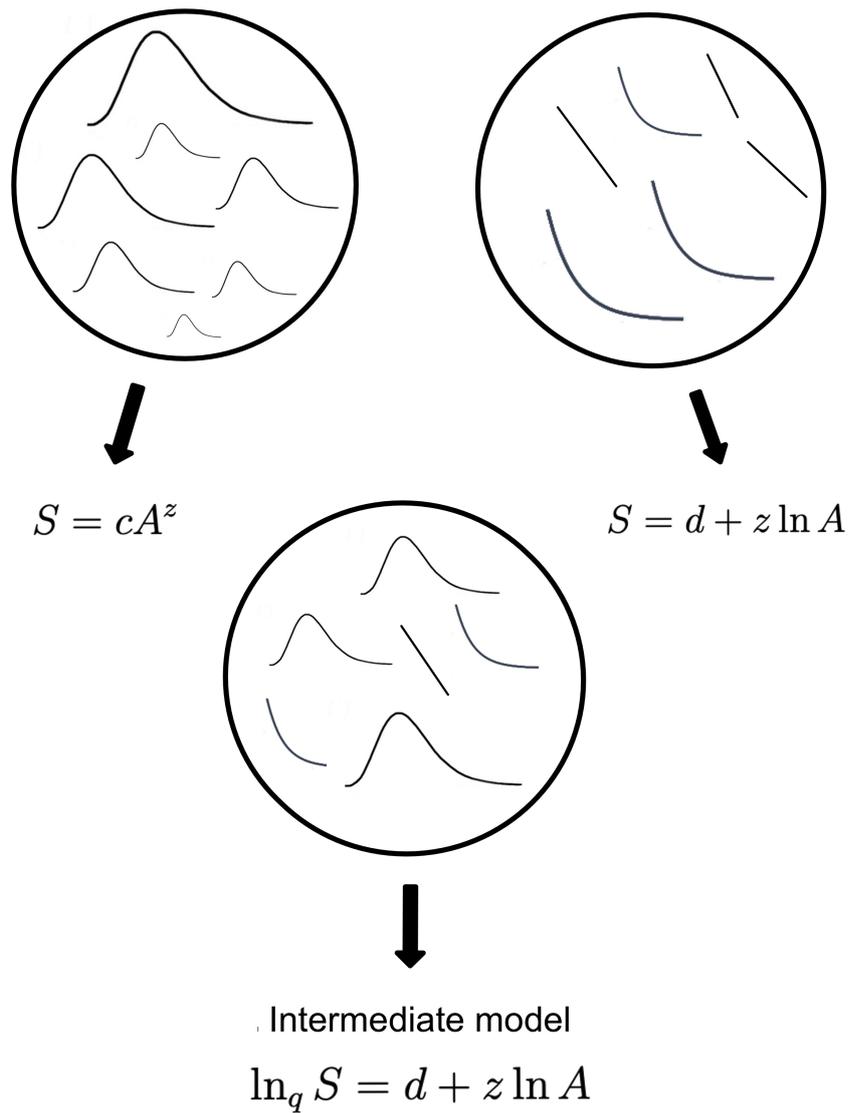


Figure 6: Log-series type species abundance distributions (SADs), when scaled across islands of different areas imply a semi-log species area relationship (SAR) and lognormal distributions scaled across islands of different area imply a power law SAR. A mix of different underlying SADs across different scales would imply an intermediate SAR somewhere between the power law and the semi-log SAR. The straight lines and downward curved lines represent geometric series and log-series distributions respectively and the positively skewed hump-shaped shapes represent lognormal distributions.

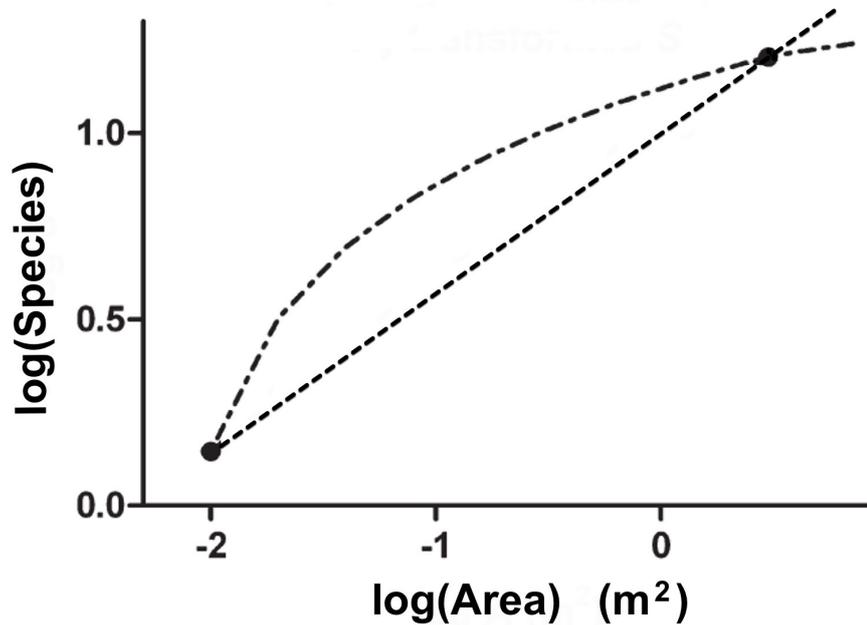


Figure 7: A comparison of a power law (dashed line) and a semi-log SAR (dotted and dashed line) on a log-log plot. Note that the semi-log SAR initially accumulates more species than the power law and then later less. This is relevant to the interpretation of Gleason's results using scattered versus contiguous quadrats (see main text). Figure adapted from Tjørve (2011)

## 351 Author Declarations

## 352 Funding

353 The authors declare that no funds, grants, or other support were received during the  
 354 preparation of this manuscript.

## 355 Competing Interests

356 The authors have no relevant financial or non-financial interests to disclose.

357 **Author Contributions**

358 All authors contributed to the study design based on an initial concept by MC. Material  
359 preparation, data collection and analysis were performed by MC. The first draft of the  
360 manuscript was written by MC. and all authors commented on previous versions of the  
361 manuscript. All authors read and approved the final manuscript.

362 **Ethics Approval**

363 Not applicable

364 **Consent to participate**

365 Not applicable

366 **Consent to publish**

367 Not applicable

368 **Data availability**

369 The datasets generated during and/or analysed during the current study are available  
370 from the corresponding author on reasonable request. Sources for the data are are given  
371 in Appendix 3 of the Supplementary Materials.

372 **Code availability**

373 All analysis was done using freely available *R* packages. No specialized code was generated.

## 374 References

- 375 Akaike, H. (1978). On the likelihood of a time series model. *Journal of the Royal Statistical*  
376 *Society: Series D (The Statistician)*, 27(3-4), 217–235.
- 377 Arrhenius, O. (1921). Species and area. *Journal of Ecology*, 9(1), 95–99.
- 378 Arrhenius, O. (1923). On the relation between species and area. – a reply. *Ecology*, 4(1),  
379 90–91.
- 380 Burnham, K. P., & Anderson, D. R. (2003). *Model selection and multimodel inference: a*  
381 *practical information-theoretic approach*. Springer Science & Business Media.
- 382 Carey, M. (2021). *Statistical analysis of dynamic patterns in insular biogeography* (Un-  
383 published doctoral dissertation). University of South Australia. UniSA STEM.
- 384 Connor, E. F., & McCoy, E. D. (1979). The statistics and biology of the species-area  
385 relationship. *The American Naturalist*, 113(6), 791–833.
- 386 Drakare, S., Lennon, J. J., & Hillebrand, H. (2006). The imprint of the geographical,  
387 evolutionary and ecological context on species–area relationships. *Ecology Letters*, 9(2),  
388 215–227.
- 389 Fattorini, S., Borges, P. A., Dapporto, L., & Strona, G. (2017). What can the parameters  
390 of the species–area relationship (sar) tell us? insights from mediterranean islands.  
391 *Journal of Biogeography*, 44(5), 1018–1028.
- 392 Forster, J. R., Thomas, N., & Guest, H. (1778). *Observations made during a voyage*  
393 *round the world*. University of Hawaii Press.
- 394 Gitay, H., Roxburgh, S. H., & Wilson, J. B. (1991). Species-area relations in a new  
395 zealand tussock grassland, with implications for nature reserve design and for commu-  
396 nity structure. *Journal of Vegetation Science*, 2(1), 113–118.

- 397 Gleason, H. A. (1922). On the relation between species and area. *Ecology*, 3(2), 158–162.
- 398 Gleason, H. A. (1925). Species and area. *Ecology*, 6(1), 66–74.
- 399 Hubbell, S. P. (2001). The unified neutral theory of species abundance and diversity.  
400 *Princeton University Press, Princeton, NJ. Hubbell, SP (2004) Quarterly Review of*  
401 *Biology*, 79, 96–97.
- 402 Lomolino, M. V. (1989). Interpretations and comparisons of constants in the species-area  
403 relationship: an additional caution. *The American Naturalist*, 133(2), 277–280.
- 404 Lomolino, M. V. (2000). Ecology’s most general, yet protean pattern: the species-area  
405 relationship. *Journal of Biogeography*, 27(1), 17–26.
- 406 MacArthur, R. (1960). On the relative abundance of species. *The American Naturalist*,  
407 94(874), 25–36.
- 408 Martinez, A. S., González, R. S., & Terçariol, C. A. S. (2008). Generalized expo-  
409 nential function and some of its applications to complex systems. *arXiv preprint*  
410 *arXiv:0812.3071*.
- 411 May, R. M. (1975). Patterns of species abundance and diversity. *Ecology and Evolution*  
412 *of Communities*, 81–120.
- 413 Plotkin, J. B., Potts, M. D., Douglas, W. Y., Bunyavejchewin, S., Condit, R., Foster, R.,  
414 ... others (2000). Predicting species diversity in tropical forests. *Proceedings of the*  
415 *National Academy of Sciences*, 97(20), 10850–10854.
- 416 Preston, F. W. (1962). The canonical distribution of commonness and rarity: Part i.  
417 *Ecology*, 43(2), 185–215.
- 418 Rosenzweig, M. L., et al. (1995). *Species diversity in space and time*. Cambridge University  
419 Press.

- 420 Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). Akaike information criterion statis-  
421 tics. *Dordrecht, The Netherlands: D. Reidel*, 81.
- 422 Schrader, J., Moeljono, S., Keppel, G., & Kreft, H. (2019). Plants on small islands revis-  
423 ited: The effects of spatial scale and habitat quality on the species-area relationship.  
424 *Ecography*, 42(8), 1405–1414.
- 425 Simberloff, D., & Gotelli, N. (1984). Effects of insularisation on plant species richness in  
426 the prairie-forest ecotone. *Biological Conservation*, 29(1), 27–46.
- 427 Tjørve, E. (2012). Arrhenius and gleason revisited: new hybrid models resolve an old  
428 controversy. *Journal of Biogeography*, 39(4), 629–639.
- 429 Tjørve, E., & Tjørve, K. M. C. (2008). The species–area relationship, self-similarity, and  
430 the true meaning of the z-value. *Ecology*, 89(12), 3528–3533.
- 431 Tokeshi, M. (1990). Niche apportionment or random assortment: species abundance  
432 patterns revisited. *The Journal of animal ecology*, 1129–1146.
- 433 Tokeshi, M. (1996). Power fraction: a new explanation of relative abundance patterns in  
434 species-rich assemblages. *Oikos*, 543–550.
- 435 Triantis, K., Mylonas, M., Lika, K., & Vardinoyannis, K. (2003). A model for the species–  
436 area–habitat relationship. *Journal of Biogeography*, 30(1), 19–27.
- 437 Triantis, K. A., Guilhaumon, F., & Whittaker, R. J. (2012). The island species area  
438 relationship: biology and statistics. *Journal of Biogeography*, 39(2), 215–231.
- 439 Tsallis, C. (2016). Inter-occurrence times and universal laws in finance, earthquakes and  
440 genomes. *Chaos, Solitons & Fractals*, 88, 254–266.
- 441 Weigelt, P., König, C., & Kreft, H. (2020). Gift—a global inventory of floras and traits for  
442 macroecology and biogeography. *Journal of Biogeography*, 47(1), 16–43.