UNIVERSITY OF
BIRMINGHAM

University of Birmingham
Research at Birmingham

The two-parameter Weibull distribution as a universal tool to model the variation in species relative abundances

Ulrich, Werner; Nakadai, Ryosuke; Matthews, Thomas J.; Kubota, Yasuhiro

Link to publication on Research at Birmingham portal

The two-parameter Weibull distribution as a universal tool to model the variation in species relative abundances

Werner Ulrich[1*], Ryosuke Nakadai[2], Thomas J. Matthews[3,4,5], Yasuhiro Kubota[6]

[1]Department of Ecology and Biogeography, Nicolaus Copernicus University, Toruń, Poland,

[2]Faculty of Science, University of the Ryukyus, Nishihara, Japan

[3]School of Geography, Earth and Environmental Sciences, University of Birmingham, UK

[4]CE3C – Centre for Ecology, Evolution and Environmental Changes/Azorean Biodiversity Group and Univ. dos Açores – Depto de Ciências e Engenharia do Ambiente, PT-9700-042, Angra do Heroísmo, Açores, Portugal

[5]Birmingham Institute of Forest Research, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

[6]Marine and Terrestrial Field Ecology, Tropical Biosphere Research Center, University of the Ryukyus, Nishihara, Japan

E-mails: WU: ulrichw@umk.pl; RN: r.nakadai66@gmail.com; TM: T.J.Matthews@bham.ac.uk; YK: kubota.yasuhiro@gmail.com

* Corresponding author

Running title: Weibull fits to species relative abundance

Author contribution

WU developed the theoretical background, analyzed the data, and wrote the first draft. YK and RN provided the forest data. TM contributed theoretical background. All authored contributed significantly to the final text version.

Abstract

The study of species abundance distributions (SADs) needs a precise modelling of their drivers and ecological implications. We introduce the two-parameter Weibull distribution as a versatile tool to fit various kinds of observed SADs and to compare observed and theoretically expected values at the species level. We show that the shape and the scale parameter of this distribution have precise ecological interpretations, the first being a measure of the excess of either rare or common species, and the second as a quantification of the proportion of persistent species in the focal community. Applying the Weibull model to 534 global tree communities we demonstrate that plots of the parameters of the Weibull distribution demark ecologically impossible species abundance distributions. This promises new insight into the ecological constraints on community assembly.

1. Introduction

1.1 Theoretical Background

Within ecological assemblages, species generally differ widely in abundances (Magurran, 2004; Matthews and Whittaker, 2015). Often, the dominant species exceed the least abundant species by more than five orders of magnitude (Ulrich et al., 2010). Since Motomura (1932) formally introduced the concept of the species relative abundance distribution (SAD), the question of which ecological processes are responsible for this large variance in abundance has been controversially discussed (Tokeshi, 1999; Hubbell, 2001; McGill et al. 2007; Locey and White, 2013). Whilst early SAD models focused on the role of niche (Sugihara, 1980) vs. stochastic processes (May, 1975), recent discussion has largely centered on the impact on the SAD of dispersal (Hubbell, 2001; Dexter et al., 2017) vs. persistence (Magurran and Henderson, 2003). These discussions have sparked the development of a large number of different SAD models, each based on a specific set of assumptions about community assembly (Magurran, 2005; McGill et al., 2007). Irrespective of the ecological and theoretical background, SADs are commonly used for biodiversity assessment and monitoring (Matthews and Whittaker, 2015).

Models of relative abundances need to be fitted to observed distributions, a task that is far

66   from being straightforward (Ulrich et al., 2010; Mathews and Whittaker, 2014; Baldridge et

67   al., 2016). To be of ecological value these fits need to be compared among different

68   assemblages, for instance to infer environmental or geographical gradients (Ulrich et al.,

69   2016a, b) and variation in abiotic conditions and biotic interactions (Erlén and Morris, 2015).

70   Traditionally, SAD models have mostly been fitted to the statistical distribution (Matthews

71   and Whittaker, 2014), where abundances are classified into $\log_2$ bins prior to fitting (Preston,

72   1948, Fig. 1 left panels). As the selection of the logarithmic base and the way in which

73   abundances are binned are arbitrary decisions, different binning procedures might have major

74   impacts on model fitting (Nekola et al., 2008; Connolly and Dornelas, 2011). Further, any

75   sufficient estimate of class frequencies needs an appropriate number of species, making

76   frequency distribution fits reliable only for larger communities (Wilson et al., 1993). In

77   contrast, plots that use ranked abundances for all species (Whittaker plots: Whittaker, 1975;

78   Bazzaz, 1975) do not loose information due to data binning (Fig. 1 right panels) or the

79   pooling of species to observed numbers of individuals. As a consequence, fits become reliable

80   at much lower species richness. Ulrich et al. (2010) recommended ten species as the lower

81   boundary for model fit, whereas Wilson (1993) reported that even 40 species may not be

82   enough to reliably identify particular types of statistical distributions. Importantly, deviation

83   of the fits from the original rank – abundance plot can be directly traced down to single

84   species and possibly to the underlying processes that influence the abundance of these

85   species. However, rank – abundance fits might fail if an excess of few very abundant or very

86   rare species biases the metric used for fitting (often ordinary least squares in combination with

87   Akaike information maximization). Further, the octaves of the SAD models that are based on

88   a statistical distribution (e.g. the lognormal distribution) and not on an algorithm that directly

89   generates for each species the expected abundance (e.g. most niche division based models,

90   Tokeshi, 1996) need to be interpolated to species abundances prior to fitting, introducing a

91   degree of subjectivism.

92   As observed SADs differ widely in shape and scale (the range in abundances), current

93   models, that are based on one (shape, for instance the lognormal and most niche division

94   models) or two (shape and scale, particularly the log-series) parameters, have limited

95   variation in shape. Most are not able to accurately mimic all of the main different SAD types

96   and do not cover the whole possible SAD space (i.e. all of the different empirically observed

97   SAD forms). For instance, in recent global comparisons of dryland plant (Ulrich et al., 2016a)

98   and forest tree (Ulrich et al., 2016b) communities, the latitudinal variation in SAD shape

99   required separate fits of two different models. In this situation it became difficult to identify

100   any gradual latitudinal trend. In this respect, Baldridge et al. (2016) reported common

101   distribution based SAD models had weak discrimination power, making any ecological

102   inference challenging.

103   These problems regarding model fitting and comparison, the biases introduced by the

104   binning of distributions, and the need for multiple model fits call for the development of

105   flexible statistical SAD descriptors that are able to mimic various SAD shapes. In this respect,

106   Ulrich et al. (2010) demonstrated that SADs can generally be classified into three basic

107   shapes. The first shape is a lognormal type statistical distribution characterized by a larger

108   number of species with intermediate abundance and fewer species with high and low

109   abundance (Fig. 1a). In empirical SADs, there is often an excess of rare species than predicted

110   by a lognormal distribution (Fig. 1b). The second shape is equivalent to a log-series sample

111   distribution (Fisher et al., 1943), which is characterized by a few abundant and a larger

112   number of relatively rare species (Fig. 1c). A small number of assemblages, particularly

113   arthropod samples, follow a third shape, a power function SAD characterized by a heavy tail

114   of rare species (Pueyo, 2006; Ulrich et al., 2010) (Fig. 1d). Here, we argue that a versatile

115   SAD model must be able to fit these three basic shapes.

116   Hughes (1986) was the first to develop a flexible SAD model. However, fitting this

117   model is not straightforward and the model has received little attention. Tokeshi (1996)

118   described a one parameter resource division model that is able to fit the symmetric and

119   skewed lognormal, but not other SAD shapes (Fig 1). The dynamic model of Dewdney (2000)

120   provides good fits to log-series shaped distributions and possibly also to power functions.

121   Ugland et al. (2007) developed the Gambin model, based on a discrete version of the

122   statistical gamma distribution. These authors argued that Gambin is able to mimic several

123   observed distribution shapes by variation of a single parameter (α); small values of α

124   characterize log-series SAD shapes, while higher values indicate lognormal curve shapes. As

125   Gambin is intended to be a descriptor only, the parameter has no clear ecological

126   interpretation, rather, it is simply a measure of the shape of the SAD. In addition, Gambin is

127   based on a statistical distribution (as used by Preston 1948) and involves binning the

128   abundance data into octaves before fitting. There is no straightforward way to rescale the

129   Gambin distribution to species abundances.

130   Of course, neutral, ecological drift models (Hubbell, 2001) provide a mechanistic

131   interpretation of observed abundance distributions with ecologically well-defined parameters.

132   Depending on the probability of dispersal, speciation rates, local abundances, and meta-

133   community size they provide predictions (Fig. 1) that are close to those of either the skewed

134   lognormal or log-series (Hubbell, 2001). These models do not predict power function SADs

135   (Fig. 1), commonly observed in arthropods (Siemann, 1999; Borda-de-Água et al., 2017). In

136   addition, neutral models are notoriously difficult to fit and certain models require information

137   on the structure of the underlying meta-community (Rosindell et al., 2010). Maximum

138   likelihood fitting of neutral models requires complex assumptions about macroevolutionary

139   processes (Etienne et al., 2007) that can result in unfeasible (or unmeasurable) parameter

140   values, such as speciation rates (Ricklefs, 2003) and long-distance dispersal (Rosindell and

141   Cornell, 2009). These issues have limited the application of neutral models for the prediction

142   of species abundances.

143   Here, we strongly argue that a flexible descriptive SAD model needs to predict the

144   abundances of each species directly and that abundance – rank orders are superior to

145   distribution approaches. Only such species-focused models make it possible to trace

146   deviations from observation and possible ecological drivers directly to the species level. There

147   is also clearly a need to link a SAD model to species functional traits or phylogenetic

148   relationships, for instance to infer how traits influence the dominance order of species

149   abundances (Jones et al., 2017). Consequently, a flexible SAD model should ideally be based

150   on Whittaker plot data.

151   Stauffer (1979) was apparently the first to propose the Weibull distribution (Weibull,

152   1951) as a model to explain observed species abundance distributions in forest trees.

153   However, his derivation of the model and parameter interpretation were based on the broken

154   stick model (Mac Arthur, 1957), a model that was subsequently found to be unrealistic. As his

155   approach was distribution based the work did not receive the attention it deserves. Recently,

156   Storch et al. (2018) used the one-parameter Weibull distribution to mimic the log-series. The

157   R package SADs (Prado et al. 2017) provides functionality to fit the Weibull model to the

158   statistical distribution, but not to rank abundances.

159   Here, we reintroduce the Weibull distribution as a flexible descriptive model that meets

160   the above defined requirements. We show that this distribution is able to fit the most

161   commonly observed SAD shapes. This ability enables us to compare the respective parameter

162   values across SADs from different studies within the same modelling framework. Although

163   Weibull distributions result from several stochastic processes (Rinne, 2008), we do not claim

164   that the model is directly linked to ecological processes. We also demonstrate that the shape

165   and the scale parameters of the model have straightforward ecological interpretations that can

166   be used in ecological analyses. As a case study, we illustrate our approach using a set of

167   global forest tree data.

168

169   1.2 The Weibull distribution as a SAD model

170   　　The empirical Weibull distribution (Weibull, 1951; Rinne, 2008) is an extension of the

171   exponential family of distributions and is widely used in survival analyses (Lawless, 2003)

172   and extreme value forecasting (Carter and Challenor, 1983). Its two-parameter form has the

173   probability density function (pdf)

174    $p(x > 0; \eta; \lambda) = \frac{\eta}{\lambda} \left(\frac{x}{\lambda}\right)^{\eta-1} e^{-\left(\frac{x}{\lambda}\right)^{\eta}}$　　　　　　　　　　(1)

175   where η is the shape and λ the scale parameter. When applied to species abundances the

176   random variate x must contain log-transformed values. Here, we use the ln-transform as a

177   standard.

178   　　Both parameters, η and λ, have clearly defined ecological interpretations. The scale

179   parameter is given by

180    $\lambda^2 = \sigma^2 \left[\left(\Gamma\left(1 + \frac{2}{\eta}\right) - \left(\Gamma\left(\left(1 + \frac{1}{\eta}\right)\right)^2\right]^{-1}$　　　　　(2)

181   where Γ denotes the gamma function and σ^2 the variance. Therefore, λ increases with

182   increasing variance in abundance and provides a measure of the range in ln-transformed

183   species abundances. We note that the abundance range is closely connected to the concept of

184   evenness. Evenness measures the variance in abundance (Smith and Wilson 1996). Therefore,

185   the wider the range in abundance is, the lower is the degree of evenness. The λ parameter can

186   therefore be interpreted as a measure of SAD shape specific evenness. The shape parameter η

187   is connected to the excess of either highly abundant species (low η) or rare species (high η).

188   　　The Weibull distribution is most often applied to assess prospected survival and failure

189   times in demography and industry. If x denotes the time to failure the quantity

190    $T = \lambda\Gamma(\frac{1}{\eta} + 1)$　　　　　　　　　　　　　(3)

191   is an estimate of the expected average time to failure (extinction in ecological terms). This

192   interpretation can be extended to species abundances. Abundant species should be highly

193   competitive or adapted to a focal habitat. Under this interpretation, the average value T

194   divides those species that are more competitive (having higher reproductive output) from

195   those that are less competitive (lower reproductive output). The position of T along the

196   abundance axis in Preston plots (Fig. 1, left panels) indicates therefore the proportion of

197   species that are relatively competitive. Interpreting the parameters of the Weibull distribution

198   in terms of reproductive output and therefore fitness implies that high values of T indicate

199 increased proportions of species with high fitness.

200   Using Weibull distributed random numbers (Press et al., 1986) rescaled to relative

201 abundances, our approach predicts directly the abundances of each species without the need

202 for retransformation. Further, Tokeshi (1999) advised the use of such stochastic models for

203 which goodness of fit can be assessed in terms of standard errors for the predicted abundance

204 of each species. As our model involves randomly assigned abundances, we here obtain such

205 errors from 100 fits to each empirical SAD and assess goodness of fit from the proportion of

206 species falling outside the 95% confidence limits of the model.

207

208 **2. Methods**

209 2.1 Fitting the Weibull model to empirical species abundance distributions

210   Simple but effective maximum likelihood estimators for the two-parameter Weibull

211 distributions already exist (Nwobi and Ugomma, 2014). However, predicting the abundances

212 of each species requires interpolation of the distribution, making the resultant fits less reliable.

213 Here we use a twofold approach to fitting the Weibull distribution to SAD data that directly

214 generates expected species abundances without the need of back-binning the distribution data.

215 We iteratively encapsulate the parameter values of η to find the value that minimizes a

216 goodness of fit metric defined by the reduced major axis value

217 $fit = \frac{\sum_{i=1}^{S}\left(lnA_{i,obs}-lnA_{i,pred}\right)^2+\min\left[(j-i)^2\right]}{S}$ (4)

218   where $A_{i,obs}$ and $A_{i,pred}$ are the respective observed and predicted relative abundances of

219 species i in the community of S species, and j runs over all S species. At each step, this fitting

220 process involves estimating the expected abundances $A_{i,pred}$ obtained from Weibull distributed

221 random numbers (f_{W}, see Press et al., 1986)

222 $f_{W;k} = \left(\frac{-1}{\lambda^{-\eta_{k-1}}}\ln(1-f_l)\right)^{1/\eta_{k-1}}$ (5)

223 where f_l is a linear random number between 0 and 1. Eq. 5 uses an initial maximum likelihood

224 estimate of λ, which is given by

225 $\lambda = \left(\frac{1}{S}\sum_{i=1}^{S}x_i^{\eta}\right)^{1/\eta}$ (6)

226 where η comes from the solution of

227 $\frac{1}{S}\frac{\sum_{i=1}^{S}x_i^{\eta}lnx_i}{\sum_{i=1}^{S}x_i^{\eta}} - \frac{1}{\eta} - \frac{\sum_{i=1}^{S}lnx_i}{S} = 0$ (7)

228 S is the observed number of species (Cohen, 1965). Eq. 7 can easily be solved using common

229 numerical methods. The values

230     $A_i = e^{x_i}$ (8)

231   are then the Weibull abundance estimates of each species i.

232   In contrast to Ulrich et al. (2016a, b), here we have used reduced major axis fits rather

233   than OLS fits as the former places similar weight on all species irrespective of relative

234   abundance, whilst the latter approach places greater weight on abundant and rare species

235   (Connolly and Dornelas, 2011). We note that OLS fits to the same data returned qualitatively

236   identical results. *fit* < 0.05 indicates an excellent fit while *fit* > 0.3 is poor. The Fortran code

237   used for fitting has already been published in Ulrich et al. (2016a, b) and is freely available

238   from WU on request.

239

240   2.2 Case study

241   To illustrate our fitting approach, we fitted the Weibull distribution to four artificial

242   communities of 50 species each that are presented in Figure 1. These represent the major SAD

243   shapes observed in nature. In a case study using empirical data, we fitted the Weibull

244   distribution to a set of 534 fully censused tree communities. These datasets were compiled by

245   restricting the global compilation of tree communities published in Ulrich et al. (2016) and

246   Kubota et al. (2018) to 534 fully censused communities ranging between 10 and 100 species

247   and containing between 20 and 38902 individuals. These communities span a wide range of

248   different abundance distributions (Ulrich et al. 2016). Evenness *J* of these communities was

249   calculated from $J = H/\ln(S)$, where *H* denotes the Shannon diversity.

250

251   **3. Results**

252   3.1 Performance of the Weibull model

253   The Weibull model provided a nearly perfect fit (*fit* = 0.004) to the symmetrical

254   lognormal data (Fig. 2a) and very good approximations to the power function data (*fit* =

255   0.007) (Fig. 2d). The fits to the left skewed lognormal data with an excess of rare species (*fit*

256   = 0.05) (Fig. 2b) and the log-series data (*fit* = 0.03) (Fig. 2c) were only marginally weaker.

257   Importantly, using the parameters of the Weibull model it was clearly possible to separate

258   these four common types of SAD (Table 1). For the lognormal (Fig. 2a), the log-series (Fig.

259   2c) and the power function data (Fig. 2c) all of the observed species abundances in Fig. 2 A

260   (left panels) were within the 95% confidence limits of the model fits, whilst for the skewed

261   lognormal data (Fig. 2b) three of the 50 species (6.0%) deviated from the fitted value. The

262   rescaling of the distributions and the fits to $\log_2$ frequency distributions (Fig. 2 B) confirmed

263   the good performance of the Weibull distribution.

264   The 95% bootstrap confidence limits of the basic parameters of the fitted distribution

265   (skewness, kurtosis, coefficients of variation in log-abundance), as well as the Weibull model

266   parameters λ and η included in all but three cases the observed values (Tab. 1). In addition,

267   the parameters significantly differed (one-way ANOVA, $P < 0.001$) between the four shapes

268   (Tab. 1), demonstrating the excellent discriminatory power of the model.

269

270   3.2 Abundance distributions of global tree communities

271   Ulrich et al. (2018) provide fits to all 534 distributions in a figshare database

272   (10.6084/m9.figshare.5975098) demonstrating the excellent performance of the model.

273   Weibull fits to 144 of the 534 communities (27.0%) were excellent (*fit* < 0.05, Fig. 3), while

274   only 20 fits (3.7%) were comparatively poor (fit > 0.3, Fig. 3 and Ulrich et al. 2018).

275   Goodness of fit was independent of species richness (Fig. 3a) but moderately increased with

276   increasing total abundance (Fig. 3b), abundance range (Fig. 3c) and model parameters (Figs 3

277   d, e). Goodness of fit decreased with increasing skewness, that is, the excess of abundant

278   species (Fig. 3f).

279   These results are corroborated by the fact that observed skewness and modelled skewness

280   were strongly positively correlated (Fig. 4a). As expected from eq. 2, the scale parameter λ

281   was positively correlated to the observed abundance range (Fig. 4b). Low η values were

282   linked to an excess of rare species, and high η values to an excess of abundant species (Fig.

283   4c). Values of λ and η were linearly correlated (Fig. 4d), defining areas of not realized

284   parameter combinations. Although λ and η are related to the variance and skewness of the

285   SAD, respective skewness - variance plots performed worse in identifying realized and

286   forbidden SAD shapes (Fig. 5e).

287

288   **4. Discussion**

289   Our fits to 534 global tree communities showed that the two-parameter Weibull

290   distribution is an appropriate tool to mimic a wide variety of observed species abundance

291   distributions (Ulrich et al. 2018). We note that the model does not provide a mechanistic

292   explanation for these distributions although several stochastic processes are known to be

293   Weibull distributed (Rinne, 2008). The data presented in Ulrich et al. (2018) also show that

294   the model might provide weaker fits in communities dominated by a small number of highly

295   abundant species and in communities with a marked excess in very rare species. We note that

296   this fitting problem also applies to other SAD models designed to fit specific stochastic niche

9

297 partitioning processes (Tokeshi 1999) and statistical distributions (Ulrich et al. 2010).

298   An important finding of the present study regards the relationship of the shape and the

299 scale parameters of the Weibull distribution that for the first time enable us to define limits to

300 the observed shapes of empirical abundance distributions (Fig. 4d). Respective plots of

301 skewness and variance (Fig. 4e) and skewness and evenness (Fig. 4f) did not recover these

302 limits. With few exceptions, communities with $\eta > 3$ did not exist. As λ increases with

303 abundance range (Fig. 4b), communities with $\lambda > 6$ might exist for instance in invertebrate

304 assemblies with abundance ranges > six orders of magnitude. High η is linked to the excess of

305 relatively rare species (Tab. 1). The theoretical power fraction SAD with a high excess of such

306 species (Figs. 1 and 2) had $\eta > 3$. Therefore, our result strongly indicates that η cannot exceed

307 a certain limit in natural communities and that such extreme SADs do not exist in global

308 forest tree communities.

309   What do such 'forbidden' communities look like? Fig. 5 shows that low η in combination

310 with low λ (Fig. 5a) predicts communities with an initial steep decline in abundance and a

311 short 'heavy tail' of rare species. High η and low λ (Fig. 5b) generate SADs very similar to

312 the popular, but discredited, broken stick distribution (MacArthur, 1957). Our results question

313 whether such communities are realized. Indeed, empirical evidence for broken stick

314 communities is very limited (Smart, 1976). Higher values of λ and η generate communities

315 for which the proportion of very rare species exceeds 40% (Fig. 5c). As already noted, such

316 distributions do not exist for species rich communities.

317   The fact that η was bounded to values between 1 and 3 implied that $\Gamma(1/\eta+1)$ takes

318 values between 0.89 and 1, making the failure time T to be a nearly linear function of λ (eq.

319 3). This fact gives the scale parameter λ yet another interpretation. It denotes the relative

320 proportion of adapted and possibly persistent species in a community. In the forest data, T was

321 significantly linearly correlated with the total abundance of all species ($r^2 = 0.54$) and

322 abundance range ($r^2 = 0.58$), but not with species richness ($r^2 = 0.01$). We speculate that

323 communities with high overall abundance also contain a higher proportion of species not

324 endangered by local extinction.

325   The approximately linear relationship between η and λ makes the respective quotient q =

326 η/λ a potentially new metric of community composition that catches the relationship between

327 range in abundance (Fig. 4b) and the excess of either rare or abundant species (Fig. 4c). It is

328 beyond the scope of the present paper to apply this metric to environmental and trait data but

329 we note that in the forest tree communities of this study, q was positively correlated to

330 community evenness ($r^2 = 0.37$) and consequently negatively to the range of abundances ($r^2 =$

331 0.71). This opens the possibility of partitioning dissimilarity in abundances into two parts, one

332 part linked to SAD shape (skewness) and a second part linked to community evenness.

333   Finally, the good fits of the Weibull distribution to observed SADs (Fig. 3 and Ulrich et

334 al. 2018) makes it possible to take a species based approach to SADs to compare model

335 prediction and observation. In this respect, distributions of functional traits and phylogenetic

336 relationships that characterize the relative abundance of realized species characteristics in a

337 community have come into focus (Gross et al. 2017; Jones et al., 2017). Being explicitly

338 based on species – rank order abundances, they allow for a direct comparison of species traits

339 among communities and between realized and theoretically expected values (Cornwall and

340 Ackerly, 2009). As neutral model fitting is challenging in most cases (Gotelli and McGill,

341 2006; Rosindell et al., 2010), the application of the Weibull distribution offers a valuable

342 alternative, particularly for the identification of outliers in abundance and the analysis of how

343 such outlying species influence community trait space and functional performance. In this

344 respect, Ulrich et al. (2016a) and Matthews et al. (2017) have recently shown that

345 comparative analyses of SAD shape and parameters can reveal important biogeographic

346 patterns (e.g. latitudinal and climate gradients). Our approach not only provides a flexible

347 model but offers ecologically interpretable parameters that can be related to species trait and

348 environmental data. We argue that the Weibull distribution might form a statistical standard

349 (similar to a null model) to which observed species relative abundances can be compared.

350

351

358

359 **Data accessibility**

360 Raw data of the forest plots used in this study are contained in Ulrich et al. (2018).

361

362 **References**

363   1. Baldridge, E., Harris, D.J., Xiao, X., White, E.P., 2016. An extensive comparison of

364   species-abundance distribution models. PeerJ 4, e2823.

365   2. Bazzaz F.A., 1975. Plant species diversity in old-field successional ecosystems in

366   southern Illinois. Ecology 56, 485–488.

367   3. Borda-de-Água, L., Whittaker, R.J., Cardoso, P., Rigal, F., Santos, A.M.C., Amorim,

368   I.R., Parmakelis, A., Triantis, K.A., Pereira, H.M., Borges, P.A.V., 2017. Dispersal

369   ability determines the scaling properties of species abundance distributions: a case

370   study using arthropods from the Azores. Sci. Rep. 7, 3899.

371   4. Carter, D.J.T., Challenor, P.G., 1983. Application of extreme value analysis to Weibull

372   data. Quart. J. R. Meteorolog. Soc 109, 429-433.

373   5. Cohen, A.C., 1965. Maximum likelihood estimation in the Weibull distribution based

374   on complete and on censored samples. Technometrics 7, 3.

375   6. Connolly, S.R., Dornelas, M., 2011. Fitting and empirical evaluation of models for

376   species abundance distributions. In: Magurran, A.E. and McGill, B.J. (eds.) Biological

377   diversity: frontiers in measurement and assessment, Oxford University Press, pp. 123–

378   140.

379   7. Cornwall, W., Ackerly, D., 2009. Community assembly and shifts in plant trait

380   distributions across an environmental gradient in coastal California. Ecol. Monogr. 79,

381   109-126.

382   8. Dewdney, A.K., 2000. A dynamical model of communities and a new species–

383   abundance distribution. Biol. Bull. 198, 152–165.

384   9. Dexter, K.G., Lavin, M., Torke, B.M., Twyford, A.D.D., Kursar, T.A., Coley, T.D.,

385   Drake, C., Hollands, R., Pennington, R.T., 2017. Dispersal assembly of rain forest tree

386   communities across the Amazon basin. Proc. Natl. Acad. Sci. USA 114, 2645-2650.

387   10. Erlén, J., Morris, F.W., 2015. Predicting changes in the distribution and abundance of

388   species under environmental change. Ecol. Lett. 18, 303-314.

389   11. Etienne, R.S., Apol, M.E.F., Olff H., Weissing, F.J., 2007. Modes of speciation and the

390   neutral theory of biodiversity. Oikos 116, 241-258.

391   12. Fisher, R.A., Corbet, A.S., Williams, C.B., 1943. The relation between the number of

392   species and the number of individuals in a random sample of an animal population. J.

393   Animal Ecol. 12, 42–58.

394   13. Hubbell, S.P., 2001. The unified theory of biogeography and biodiversity. Princeton

395   University Press.

396   14. Hughes, R.G., 1986. Theories and models of species abundance. Am. Nat. 128, 879–

397   899.

398   15. Gotelli, N.J., McGill, B.J., 2006. Null versus neutral models: what's the difference?
399   Ecography 29, 793-800.

400   16. Gross, N., Le Bagousse-Pinguet, Y., Liancourt, P., Berdugo, M., Gotelli, N.J., Maestre,
401   F.T., 2017. Functional trait diversity maximizes ecosystem multifunctionality. Nature
402   Ecol. Evol. 1, 0132.

403   17. Jones, S.K., Ripplinger, J., Collins, S.L., 2017. Species reordering, not changes in
404   richness, drives long-term dynamics in grassland communities. Ecol. Lett. 20, 1556-
405   1575.

406   18. Kubota Y., Kusumoto B., Shiono T., Ulrich W., 2018. Multiple filters affect tree
407   species assembly in mid-latitude forest communities. Oecologia, in press.

408   19. Lawless, J. F., 2003. Statistical Models and Methods for Life time Data. 3rd Edition,
409   John Wiley and Sons, New York.

410   20. Locey, K.J., White, E.P., 2013. How species richness and total abundance constrain the
411   distribution of abundance. Ecol. Lett. 16, 1177–1185.

412   21. MacArthur, R.H., 1957. On the relative abundance of bird species. Proc. Natl. Acad.
413   Sci. USA 43, 293–295.

414   22. Magurran, A.E., 2004. Measuring Biological Diversity. Blackwell Science, Oxford.

415   23. Magurran, A., 2005. Species abundance distributions: pattern or process? Funct. Ecol.
416   19, 177-181.

417   24. Magurran, A.E., Henderson, P.A., 2003. Explaining the excess of rare species in
418   natural species abundance distributions. Nature 422, 714–716.

419   25. Matthews, T.J., Whittaker, R.J., 2014. Fitting and comparing competing models of the
420   species abundance distribution: assessment and prospect. Front. Biogeogr. 6, 67–82.

421   26. Matthews, T.J., Whittaker, R.J., 2015. On the species abundance distribution in
422   27. applied ecology and biodiversity management. J. Appl. Ecol. 52, 443–454.

423   28. Matthews, T.J., Borges, P.A.V., de Azevedo, E.B., Whittaker, R.J., 2017. A
424   biogeographical perspective on species abundance distributions: recent advances and
425   opportunities for future research. J. Biogeogr. 44, 1705–1710

426   29. May, R.M., 1975. Patterns of species abundance and diversity. In: Cody, M.L.,
427   Diamond, J.M. (eds.) Ecology and Evolution of Communities, Cambridge University
428   Press, pp. 81–120.

429   30. McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K.,
430   Dornelas,M., Enquist, B.J., Green, J.L., He, F., Hurlbert, A.H., Magurranm A.E.,
431   Marquet, P.A., Maurer, B.A., Ostling, A., Soykan, C.U., Ugland, K.I., White, E.P.,

432   2007. Species abundance distributions: moving beyond single prediction theories to

433   integration within an ecological framework. Ecol. Lett. 10, 995–1015.

434 31. Motomura, I., 1932. A statistical treatment of associations. Japan. J. Zool. 44, 379–383

435   (in Japanese).

436 32. Nekola, J.C., Šizling, A.L., Boyer, A.G., Storch, D., 2008. Artifactions in the log-

437   transformation of species abundance distributions. Folia Geobot. 43, 259–268.

438 33. Nwobi, F.N., Ugomma, C.A., 2014. A Comparison of Methods for the Estimation

439 34. of Weibull Distribution Parameters. Metodologya Zvezki 11, 65-78.

440 35. Prado, P.I., Dantas Miranda, M., Chalom, A. 2017. Package sads.

441   http://piLaboratory.github.io/sads.

442 36. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1986. Numerical

443   Recipes, The Art of Scientific Computing, Cambridge University Press.

444 37. Preston, F.W., 1948. The commonness, and rarity, of species. Ecology 29, 254–283.

445 38. Pueyo, S., 2006. Self-similarity in species–area relationship and in species abundance

446   distribution. Oikos 112, 156–162.

447 39. Ricklefs, R. E., 2003. A comment on Hubbell's zero-sum ecological drift model. Oikos

448   100, 185–192.

449 40. Rinne, H., 2008. The Weibull Distribution. A Handbook. CRC Press.

450 41. Rosindell, J., Cornell, S.J., 2009. Species–area curves, neutral models, and long-

451   distance dispersal. Ecology 90, 1743–1750.

452 42. Rosindell, J., Cornell, S.J., Hubbell, S.P., Etienne, R.S., 2010. Protracted speciation

453   revitalizes the neutral theory of biodiversity. Ecol. Lett. 13, 716–727.

454 43. Smart, J.S., 1976. Statistical tests of the broken-stick model of species-abundance

455   relations. J. Theor. Biol. 59, 127-139.

456 44. Siemann, E., Tilman, D., Haarstad, J., 1999. Abundance, diversity and body size:

457   Patterns from a grassland arthropod community. J. Animal Ecol. 68, 824-835.

458 45. Smith, B, Wilson, J.B., 1996. A consumer's guide to evenness indices. Oikos 76, 70-

459   82.

460 46. Stauffer, H.B., 1979. A derivation for the Weibull distribution. J. Theor. Biol. 81, 55-

461   63.

462 47. Storch, D., Bohhalková, E., Okie, J., 2018. The more-individuals hypothesis revisited:

463   the role of community abundance in species richness regulation and the productivity–

464   diversity relationship. Ecol. Lett. 21, 920-937.

465 48. Tokeshi, M., 1996. Power fraction: a new explanation for species abundance patterns

14

466   in species-rich assemblages. Oikos 75, 543–550.

467   49. Tokeshi, M., 1999. Species coexistence: ecological and evolutionary perspectives.

468   Blackwell Scientific, Oxford.

469   50. Ugland, K.I., Lambshead, P.J.D., McGill, B., Gray, J.S., O'Dea, N., Ladle, R.J.,

470   Whittaker, R.J., 2007. Modelling dimensionality in species abundance distributions:

471   description and evaluation of the Gambin model. Evol. Ecol. Res. 9, 313-324.

472   51. Ulrich W., Ollik M., Ugland, K.I., 2010. A meta-analysis of species - abundance

473   distributions. Oikos 119, 1149-1155.

474   52. Ulrich, W., Kusumoto, B., Shiono, T., Kubota, Y., 2016a. Climatic and geographical

475   correlates of global forest tree species abundance distributions and community

476   evenness. J. Veg. Sci. 27, 295-305.

477   53. Ulrich, W., Soliveres, S., Thomas, A.D., Dougill, A.J., Maestre, F.T., 2016b.

478   Environmental correlates of species rank – abundance distributions in global drylands.

479   Persp. Plant Ecol. Evol. Syst. 20, 56-64.

480   54. Ulrich, W., Nakadai, R., Matthews, T., Kubota, Y. 2018. Weibull fits to global tree

481   species abundance distributions. figshare

482   https://figshare.com/articles/Weibull_fits/5975098.

483   55. Weibull, W., 1951. A Statistical Distribution of wide Applicability. J. Appl. Mech. 18,

484   239–296.

485   56. Whittaker, R.H., 1975. Communities and Ecosystems, 2nd edn. MacMillan, New

486   York.

487   57. Wilson, J.B., 1993. Would we recognise a broken-stick community if we found one?

488   Oikos 67, 181-183.

489

490 **Software**

491 Fortran source code and the updated stand-alone application RAD 2.0 (Ulrich et al. 2010,

492 2016a, b) are freely available from W.U. by request.

493

494 **Data availability**

495 Weibull fits to global tree species abundance distributions.
496 figshare https://figshare.com/articles/Weibull_fits/5975098.

497

498 **Table 1.** Coefficients of variation (CV) and moments of the ln-transformed relative

499 abundances of the communities in Fig. 2, and the respective fits of the Weibull distribution

500 (given are shape and scale parameters of the best fit) together with upper and lower two-sided

501 95% confidence limits (CL) of 1000 bootstrap samples. Shape and scale of the data refer to

502 maximum likelihood Weibull parameters taken directly from the ln-transformed abundances.

503

| Model	Statistics	CV	Standard deviation	Skewness	Kurtosis	Shape	Scale
lognormal	Data	0.45	1.41	0.10	-0.21	2.51	2.45
	fit	0.53	1.35	0.36	-0.26	2.45	2.06
	lower CL	0.43	1.12	-0.10	-1.32	-	-
	upper CL	0.64	1.61	0.87	0.50	-	-
skewed lognormal	Data	**0.38**	2.00	**-0.50**	0.05	3.22	4.28
	fit	**0.54**	1.87	**0.12**	-0.39	3.16	3.23
	lower CL	**0.43**	1.58	**-0.35**	-1.25	-	-
	upper CL	**0.65**	2.23	**0.62**	0.25	-	-
log-series	Data	**0.98**	1.27	0.75	-0.39	1.11	0.93
	fit	**0.80**	1.14	1.03	0.90	1.09	0.94
	lower CL	**0.65**	0.91	0.47	-1.35	-	-
	upper CL	**0.96**	1.42	1.72	2.78	-	-
power function	Data	0.91	1.21	1.20	0.89	1.15	0.96
	fit	0.92	1.23	1.66	3.40	1.12	0.96
	lower CL	0.73	0.87	0.93	-0.71	-	-
	upper CL	1.14	1.62	2.66	7.33	-	-

504

16

Figure 1. Four typical shapes of species abundance distributions (Whittaker species - rank
order plots). a) a lognormal distribution, b) left skewed lognormal with an excess of rare
species, c) a log-series distribution, d) a power function. Each community contains 50 species.
Arrows show which types of distributions the selected flexible SAD models can mimic.
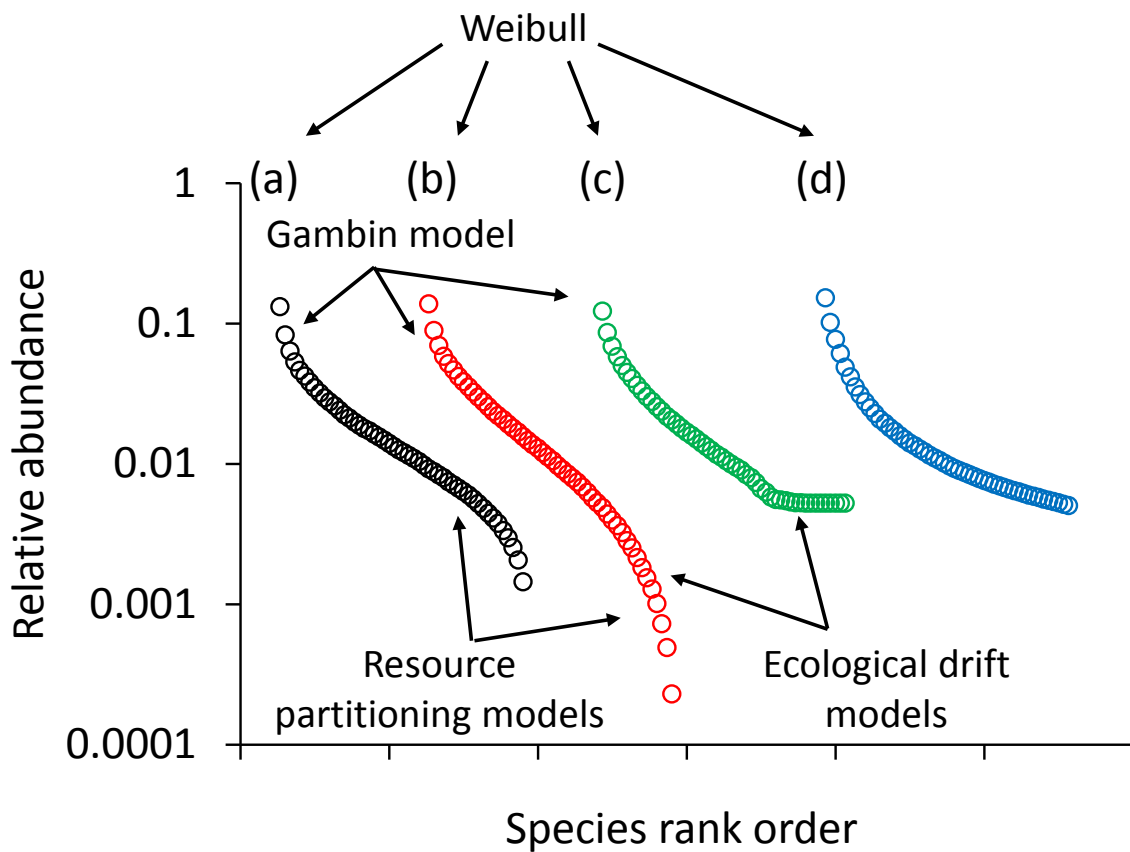
514   **Figure 2**. Reduced major axis fits of the Weibull model to rank abundance (left side) and the

515   respective $\log_2$-binned distribution data (right side) of the four typical shapes of species rank

516   order –abundance distributions, of 50 species each shown in Fig. 1. a) a lognormal

517   distribution, b) left skewed lognormal with an excess of rare species, c) a log-series

518   distribution, and d) a power function. Error bars denote one standard deviation of the Weibull

519   model fits. Parameter values are given in Tab. 1

520
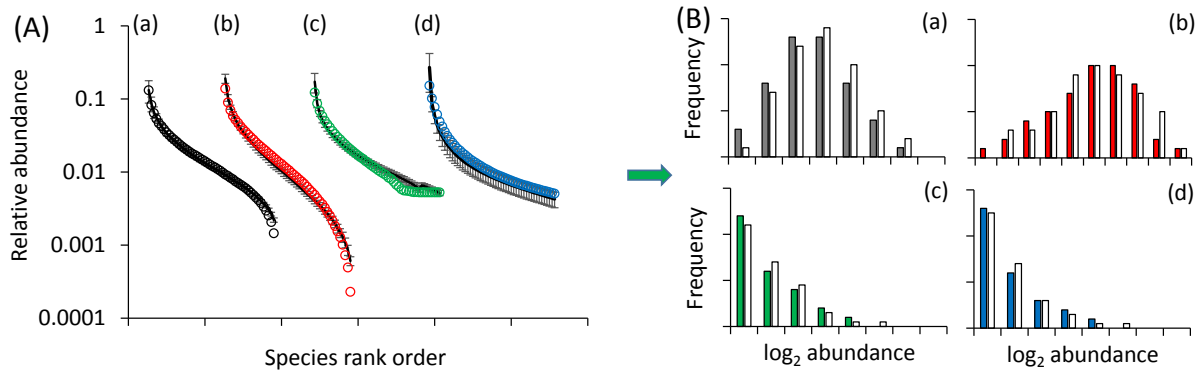
521
522
523

524   **Figure 3**. Goodness of fit of 534 empirical global forest tree communities in relation to

525   species richness (a), total abundance (b), range in abundance (c), Weibull fit parameters λ (d),

526   and η (e), and observed skewness (f). Given are the r^2 values from exponential (a, d, e, f) and

527   power function (b, c) OLS regressions.
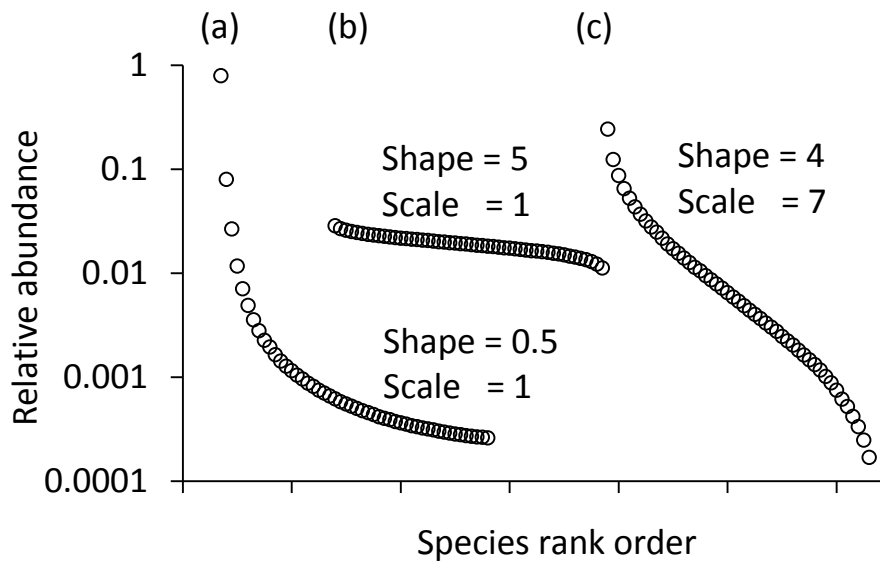
528

529

530

531

532
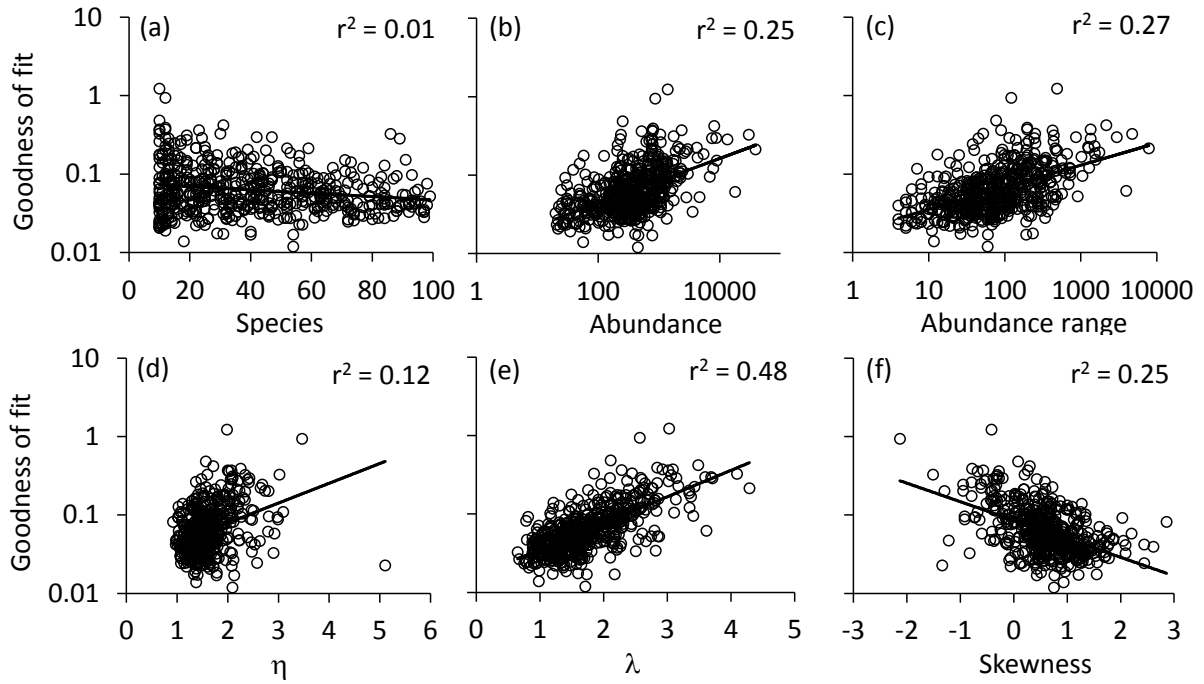
Figure 4. Relationships between observed and expected skewness (a), λ and observed
abundance range (b), η and observed skewness (c), η and λ (d), observed skewness and
variance (e), and evenness (f) from the fits of the Weibull distribution to 534 empirical global
tree communities. Given are the r^2 values of linear (a, c, d, e) and logarithmic (b) OLS
regressions. Approximate upper and lower boundaries of η are given in (d).
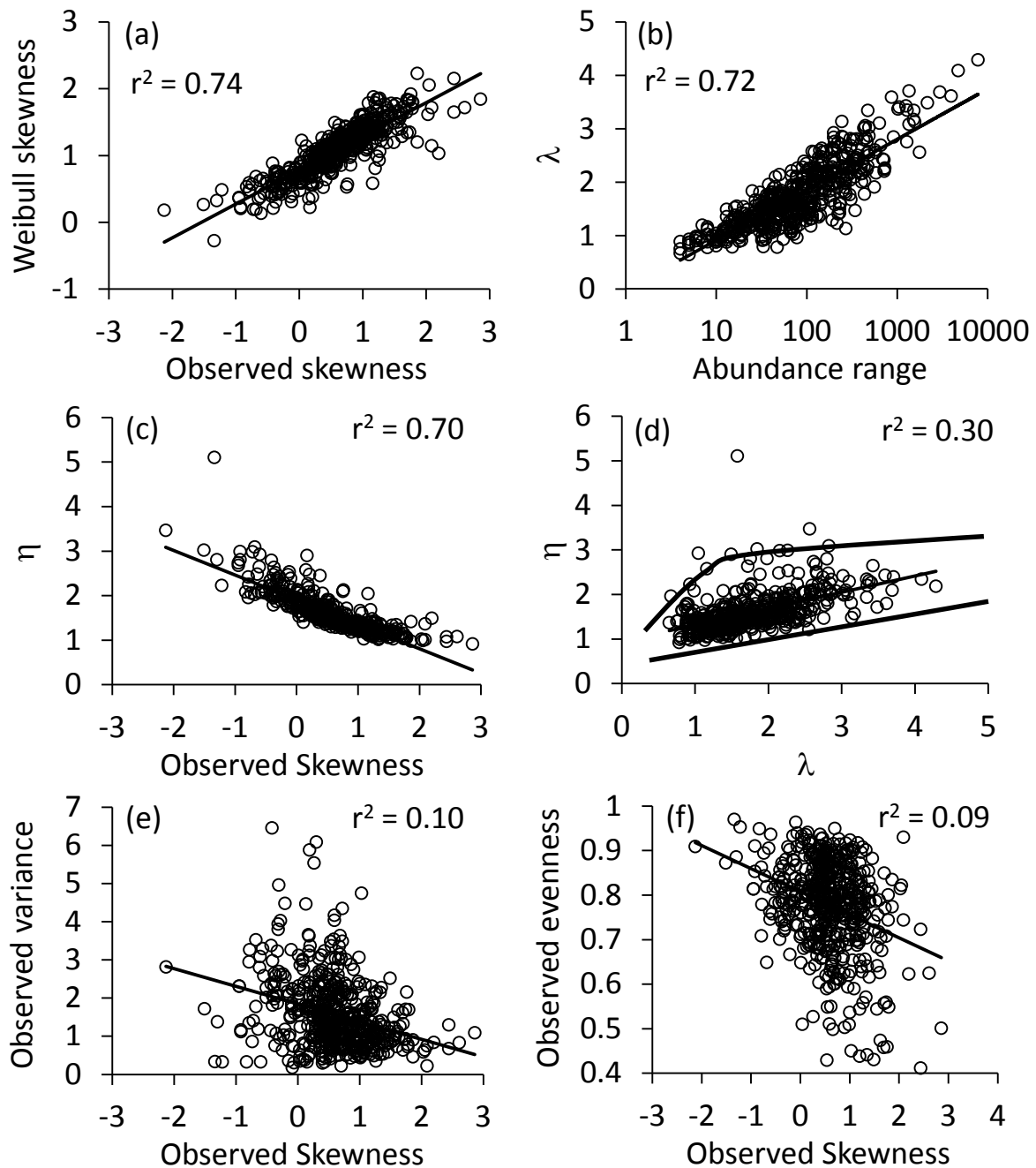
Figure 5. Three examples of communities (50 species each) generated by Weibull

distributions that are not realized according to Figure 4d.