Fuzzy quantification of common and rare species in ecological communities (FuzzyQ)

Juan Antonio Balbuena, Clara Monlleó-borrull, Cristina Llopis-belenguer, Isabel Blasco-costa, Volodimir Sarabeev, Serge Morand

HAL Id: hal-03188037

https://hal.inrae.fr/hal-03188037v1

Submitted on 17 Nov 2024

1   **Fuzzy Quantification of Common and Rare Species in Ecological Communities**

2   **(FuzzyQ)**

3   Juan A. Balbuena[1], Clara Montlleó[1], Cristina Llopis-Belenguer[1], Isabel Blasco-Costa[2],

4   Volodimir L. Sarabeev[3], Serge Morand[4,5]

5   [1] Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia,

6   Valencia, Spain

7   [2] Department of Invertebrates, Natural History Museum of Geneva, Geneva,

8   Switzerland

9   [3] Department of Biology, Zaporizhzhia National University, Zhaporizhzhia, Ukraine

10   [4] Institut des Sciences de l'Evolution, CNRS, CC065, Université de Montpellier,

11   Montpellier Cedex, France; CIRAD ASTRE

12   [5] Faculty of Veterinary Technology, Kasetsart University, 10400 Bangkok, Thailand

13   **Correspondence:** Juan A. Balbuena Email: j.a.balbuena@uv.es

14

15   **Running headline:** Fuzzy Commonness and Rarity in Communities

16 **Abstract**

17 1. Most species in ecological communities are rare whereas only a few are common.

18 This distributional paradox has intrigued ecologists for decades but the interpretation of

19 species abundance distributions remains elusive.

20 2. We present Fuzzy Quantification of Common and Rare Species in Ecological

21 Communities (FuzzyQ) as an R package. FuzzyQ shifts the focus from the prevailing

22 species-categorization approach to develop a quantitative framework that seeks to place

23 each species along a rare-commonness gradient. Given a community surveyed over a

24 number of sites, quadrats, or any other convenient sampling unit, FuzzyQ uses a fuzzy

25 clustering algorithm that estimates a probability for each species to be common or rare

26 based on abundance-occupancy information. Such as probability can be interpreted as a

27 commonness index ranging from 0 to 1. FuzzyQ also provides community-level metrics

28 about the coherence of the allocation of species into the common and rare clusters that

29 are informative of the nature of the community under study.

30 3. The functionality of FuzzyQ is shown with two real datasets. We demonstrate how

31 FuzzyQ can effectively be used to monitor and model spatio-temporal changes in

32 species commonness, and assess the impact of species introductions on ecological

33 communities. We also show that the approach works satisfactorily with a wide range of

34 communities varying in species richness, dispersion and abundance currencies.

35 4. FuzzyQ produces ecological indicators easy to measure and interpret that can give

36 both clear, actionable insights into the nature of ecological communities and provides a

37 powerful way to monitor environmental change on ecosystems. Comparison among

38 communities is greatly facilitated by the fact that the method is relatively independent

39 of the number of sites or sampling units considered. Thus, we consider FuzzyQ as a

40 potentially valuable analytical tool in community ecology and conservation biology.

41   Key-Words: Rarity, commonness, abundance-occupancy distributions, fuzzy clustering,

42   community ecology, assembly rules, environmental monitoring.

43   **Introduction**

44   Ecological communities are formed by species that differ widely in abundance. Almost

45   invariably the observation is that most species are rare, whereas a few are common

46   (Magurran & Henderson, 2011). This pervasive pattern has intrigued ecologists for

47   decades but, despite the large literature on the topic, the interpretation of species

48   abundance distributions remains elusive (Werner et al., 2014; Enquist et al., 2019). The

49   assumption often made is that underlying factors, such as immigration, succession and

50   competition, eventually determine differences in establishment and persistence of each

51   species in the community (McGill et al., 2007; McGill, 2011; Alroy, 2015; Calatayud et

52   al., 2019).

53   A quantitative framework for species commonness and rarity amenable to

54   hypothesis testing and statistical modelling would facilitate evaluating the roles played

55   by demographic variables and species traits, thereby illuminating assembly rules in

56   ecological communities. Such a framework would also be extremely valuable for

57   conservation biology in at least three important aspects: (1) Species rarity is an

58   important predictor of extinction risk because the impact of environmental disturbances

59   is expected to be higher on small populations (Davies, Margules, & Lawrence, 2000)

60   and since budgets for biodiversity conservation are limited, quantifying the rarity of

61   species would facilitate prioritizing some over the others. (2) From a functional

62   perspective, the role played by common and rare species in providing ecosystem

63   services is currently under scrutiny. Although intuitively it might be assumed that most

64   of the ecosystem functionality should reside in the former, the contribution of rare

65   species is still poorly understood and, in fact, might be substantial (Leitão et al., 2016;

66 Violle et al., 2017; Dee et al., 2019). So assessing rarity could also be justified in terms

67 of identifying species that provide essential ecosystem services (Flather & Sieg, 2007;

68 Violle et al., 2017; Dee et al., 2019) or stabilize ecological communities (Calatayud et

69 al., 2019). (3) Monitoring variation of commonness-rarity patterns over time or along

70 geographical and environmental gradients provides a simple way to obtain crucial

71 information on ecosystem changes (McGill, 2011). For instance, if common species

72 become increasingly rare in response to environmental disturbances, it might have a

73 cascading effect on the rest of the community (Gaston & Fuller, 2008).

74   Thus, metrics of commonness and rarity at species and community level would

75 be extremely useful to unveil the architecture of ecological communities, assess the

76 likelihood of extinction of rare species, correlate commonness or rarity with functional

77 distinctiveness and monitor environmental change. However, a universal quantitative

78 framework is currently lacking. A great deal of effort has been put on establishing the

79 distribution patterns emerging from the categorization of species as common or rare

80 (Gray, Bjørgesæter, & Ugland, 2005; McGill et al., 2007; Antão, Connolly, Magurran,

81 Soares, & Dornelas, 2017). However, a major problem of fitting models to species

82 abundance distributions has been adjusting the data to a suitable theoretical distribution

83 (Williamson & Gaston, 2005; McGill et al., 2007; Alroy, 2015). To some extent this is

84 because the border between common and rare species is often blurred (Magurran &

85 Henderson, 2011), which has led authors to propose additional subcategories of rarity

86 (Hanski, 1991; Yu & Dobson, 2000; Arnan, Gaucherel, & Andersen, 2011).

87   Herein we propose shifting the focus from species categorization to a

88 quantitative approach that seeks to place each species along a rare-commonness

89 gradient. Fuzzy Quantification of Common and Rare Species in Ecological

90 Communities (FuzzyQ) is based on the analysis of abundance-occupancy relationships

4

91   (AORs), which assumes a positive relationship between local abundance and occupancy

92   (Gaston et al., 2000; Gaston & He, 2011). Given a community surveyed over a number

93   of sites, quadrats, or any other convenient sampling unit, FuzzyQ applies a fuzzy

94   clustering algorithm (Kaufman & Rousseeuw, 1990) that estimates a probability for

95   each species to be common or rare based on its AOR.

96   Although widely used, we acknowledge at the onset that abundance and/or

97   occupancy are not the only criteria to assess species commonness and rarity (Gaston,

98   1994, 1997). However, the key point is that regardless of the data used, we can always

99   use fuzzy clustering to quantify the degree of belonging of each species to the common

100   or rare categories (or any other pre-established categorization for that matter).

101   We show herein that FuzzyQ produces ecological indicators easy to measure and

102   interpret that are amenable to hypothesis testing and statistical modelling. In addition,

103   FuzzyQ is distribution free, i.e. no a priori assumption about the distribution of species

104   abundances is required. We illustrate the capabilities of the framework with two real-

105   world examples involving each related and unrelated (i.e. not sharing species)

106   communities and evaluate the effect of sample size on the estimation of commonness

107   and rarity.

108   **Overview of FuzzyQ**

109   FuzzyQ is provided as an R package (R Core Team, 2020), available at

110   https://github.com/Ligophorus/FuzzyQ, which depends on algorithms implemented in

111   package `cluster` (Maechler, Rousseeuw, Struyf, Hubert, & Hornik, 2019). We first

112   illustrate application of FuzzyQ with a dataset of ant species (ants_Darwin_A in

113   Calatayud el al., 2019) comprising the abundance of 46 species in 100, 18×18 m plots

114   sampled in the Northern Territory, Australia (Arnan et al., 2011).

115

116

117 **TABLE 1** Overview of the functions implemented in package FuzzyQ

Function	Description	Output
fuzzyq	Determines the abundance-occupancy per species of a site × species matrix. Performs fuzzy clustering of common and rare species based on abundance occupancy.	An object of class `list` and `fuzzyq` of three objects: $A_O Fraction of sites occupied and the mean abundance across sites per species $Diss Object of class `dist` with pairwise dissimilarities among species based on their abundance and occupancy $spp Three metrics for each species: 1. Cluster membership, where 0 and 1 denote allocation to the rare or common category, respectively 2. Silhouette width 3. Commonness index $global Community-level metrics: 1. Average silhouette widths per cluster and globally 2. Mean commonness indices per cluster 3. Normalized Dunn coefficient
fuzzyqBoot	Produces *N* replicates bootstrapping the site × species matrix by site and applies `fuzzyq` to each replicate.	A matrix of either species commonness indices (level = "spp") or community-level metrics (level = "global") of each bootstrap replicate
fuzzyqCI	Computes confidence intervals of the parameters computed with `fuzzyqBoot`. Three methods are available: percentile, bias corrected, and bias corrected and accelerated.	A matrix of lower and upper bound limits at a given confidence level (default 95%)
AOplot	Plots the abundance occupancy relationship of a `fuzzyq` object.	A scatter plot of fraction of sites occupied by each species vs mean abundance per site. Common and rare species are distinguished by respective convex hulls
sortClus	Sorts species data of a matrix (columns) by the cluster allocation of a `fuzzyq` object. Useful to sort the output of `fuzzyqBoot` or `fuzzyqCI` by the original `fuzzyq` object.	A matrix sorted by cluster matriz. Species are arranged by cluster and by increasing silhouette width within cluster

118   Table 1 provides an overview of the functions in package FuzzyQ. Function

119   fuzzyq takes a given site-by-species abundance matrix and performs a fuzzy clustering

120   algorithm that evaluates all pairwise dissimilarities among species in terms of their

121 AORs to allocate each species into two clusters of common and rare species,

122 respectively. Since occupancy and abundance are in different scales and can come in

123 different units (for instance, the former can be reported as either number or fraction of

124 sites occupied), `fuzzyq` uses by default Gower's (1971) dissimilarities, which are

125 appropriate for such mixed data. Clustering is subsequently performed with function

126 `fanny` in `cluster`, which aims to minimize the objective function

$$\sum_{v=1}^{k} \frac{\sum_{i,j=1}^{n} u_{iv}^{a}\, u_{jv}^{a}\, d(i,\ j)}{2 \sum_{j=1}^{n} u_{jv}^{a}}, \tag{1}$$

127 where u_{iv} and u_{jv} are the membership coefficients of observations (species in our case)

128 i and j to cluster v, n is the number of observations, k is the number of clusters (herein 2:

129 common and rare), a is a membership exponent (we set $a = 2$ as in the original

130 formulation of Kaufman & Rousseeuw 1990) and $d(i,\ j)$ is the dissimilarity between

131 observations i and j (Maechler et al., 2019). Fig. 1a displays the `fuzzyq` allocation of

132 species of the ant dataset to the rare and common clusters.

133   In fuzzy clustering, each observation can be assigned to several clusters with a

134 different level of certainty. So u_{iv} in (1) above represents the probability of the ith

135 observation belonging to cluster v (Kaufman & Rousseeuw, 1990). We re-interpret

136 these probabilities as indices of commonness (C_i) and rarity (R_i) for species i, so that

137 each species is classified simultaneously as common and rare with a certain level of

138 certainty. (Given that $C_i = 1 - R_i$, we will only report C_i). `fuzzyqBoot` generates and

139 applies `fuzzyq` to bootstrap replicates by site of the species abundance matrix and

140 `fuzzyqCI` computes confidence intervals of C_i based on these replicates (Fig. 1b).
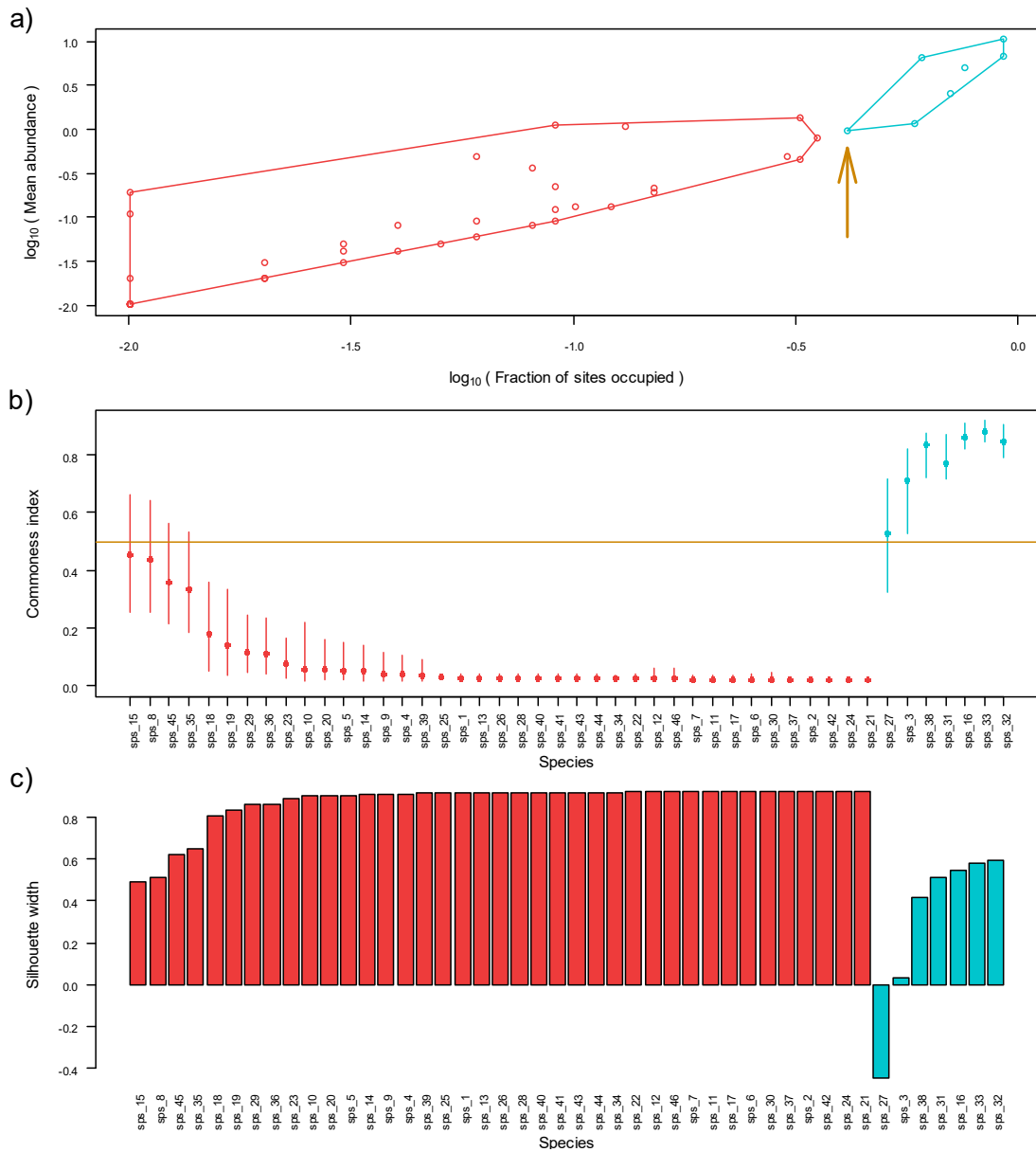
FIGURE 1 Fuzzy quantification of common and rare species in a community of 46 ant species in 100 plots (Arnan et al., 2011, Calatayud el al., 2019). (a) Abundance occupancy relationship of species. The arrow points to the position of Species 27. (b) Commonness indices of species. Error bars represent bias-corrected and accelerated 95% confidence intervals (error bars) (Efron & Tibshirani, 1994) computed with 1,000 replicates bootstrapping the plots of the abundance matrix. The horizontal line marks the 0.5 threshold separating rare and common species. (c) Silhouette plot of the 46 ant species. The negative value of Species 27 suggests a poor fit to the cluster of common species. Light blue and light red denote common and rare species, respectively.

141   In addition, `fuzzyq` computes silhouette widths, which are measures of how

142 similar abundance and occupancy of each species are to its own cluster relatives and to

143 these of species in the other cluster, as follows:

144
$$s_i = \frac{c_i - r_i}{max(c_i,\ r_i)}\ ,$$

145 where c_i and r_i are the mean dissimilarity between species i and all other species in the

146 clusters of common and rare species, respectively (Kaufman & Rousseeuw, 1990). Fig.

147 1c shows the species silhouettes of the ant database. Silhouettes can range between -1

148 and +1. The high positive values of most rare ant species indicate that they are well

149 matched to its own cluster. Common ant species showed smaller silhouette widths,

150 suggesting a weaker cluster. In particular, the negative silhouette of species 27 indicates

151 a poor fit to the common-species group (Fig. 1c), which conforms to its position in the

152 AOR plot and its $C_i \approx 0.5$ (Fig. 1a, b).

153   `fuzzyq` also computes community-level metrics that measure the coherence of

154 the common- and rare-species clusters, and the strength of overall classification. The

155 former is assessed by the average silhouettes' widths of the common and rare species

156 ($\bar{S}_C$ and $\bar{S}_R$, respectively) and, alternatively, by the corresponding average commonness

157 coefficients ($\bar{C}_C$ and $\bar{C}_R$). The latter can be appraised by the average silhouette width of

158 the whole community ($\bar{S}$) and the normalized Dunn's partition coefficient (D')

159 (Kaufman & Rousseeuw, 1990). The Dunn's coefficient is computed as

160
$$D = \sum_{i=1}^{n}\sum_{v=1}^{k} u_{iv}^2 / n\ ,$$

161 where n is the number of observations (i.e., species). D is subsequently normalized to

162 vary between 0 (complete fuzziness) and 1 (hard clusters). When $k = 2$, as in our case,

163 the normalized Dunn's coefficient is $D' = 2D - 1$ (Kaufman & Rousseeuw, 1990).
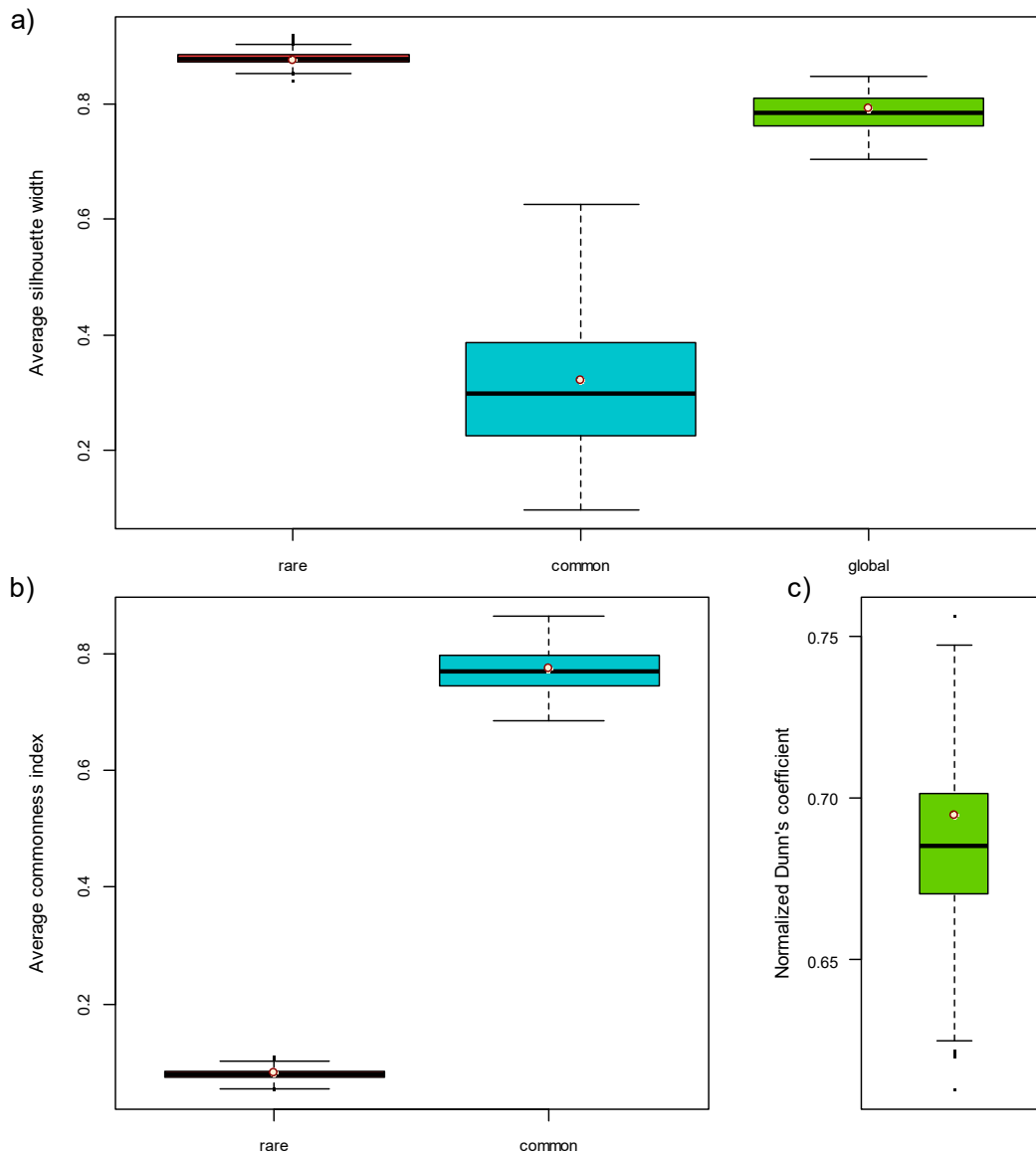
FIGURE 2 Community-level estimates (points) and their variation computed with 1,000 replicates bootstrapping the plots of the abundance matrix (boxplots) of 46 ant species in 100 plots (Arnan et al., 2011, Calatayud el al., 2019). (a) Average silhouette widths of rare, common and all species. (b) Average commonness indices of rare and common species. (c) Normalized Dunn's coefficient.

164   Fig. 2 displays the global metrics of the ant species database. $\bar{S}$ and D' were 0.79

165 and 0.69, which suggests a strong clustering structure separating common and rare

166 species (Kaufman & Rousseeuw, 1990). We assessed the variation of the global

167 estimates by bootstrapping the sites of the sites × species matrix with `fuzzyqBoot`.

168 Compared with common ones, rare species showed a higher average silhouette width,

169 and showed lower variation in both silhouettes and commonness indices, indicating that

170 they form a harder cluster (Fig. 2).

Worked-out examples

172 We demonstrate the new method and its capabilities, with two real datasets involving

173 the comparison of related and unrelated communities, respectively. When comparing

174 several communities, one must consider how to deal with species absences. Absences

175 may be due to (a) eco-evolutionary constraints (structural absence), (b) sampling

176 variability (random absence), or (c) methodological errors (false absences) (Blasco-

177 Moreno, Pérez-Casany, Puig, Morante, & Castells, 2019). Although `fuzzyq` cannot

178 deal with (c), it would produce different metrics in (a) and (b) situations and researchers

179 should make an informed decision based on the nature of their system. The logical

180 argument `rm.absent` in `fuzzyq` specifies whether species absences are to be treated

181 as structural or random.

Example 1. Mammal Data from Powdermill Biological Station 1979-1999

183 To illustrate how to monitor changes in species commonness in a community, we used a

184 long-term (1979-1999) time series of small mammal abundances from the Powdermill

185 Biological Station in Pennsylvania, USA (Merrit, 2013). Mammals were captured in a

186 1-ha live trapping grid consisting of 10×10 quadrats of trap stations at 10-m intervals.

187 Trapping was conducted bimonthly from September 1979 to October 1999. For the sake

188 of demonstration, the abundance of each mammal species was aggregated per quadrat

189 and per year in order to capture the annual variation in commonness of each species and

190 we assumed that the pool of species did not change over the study period (random
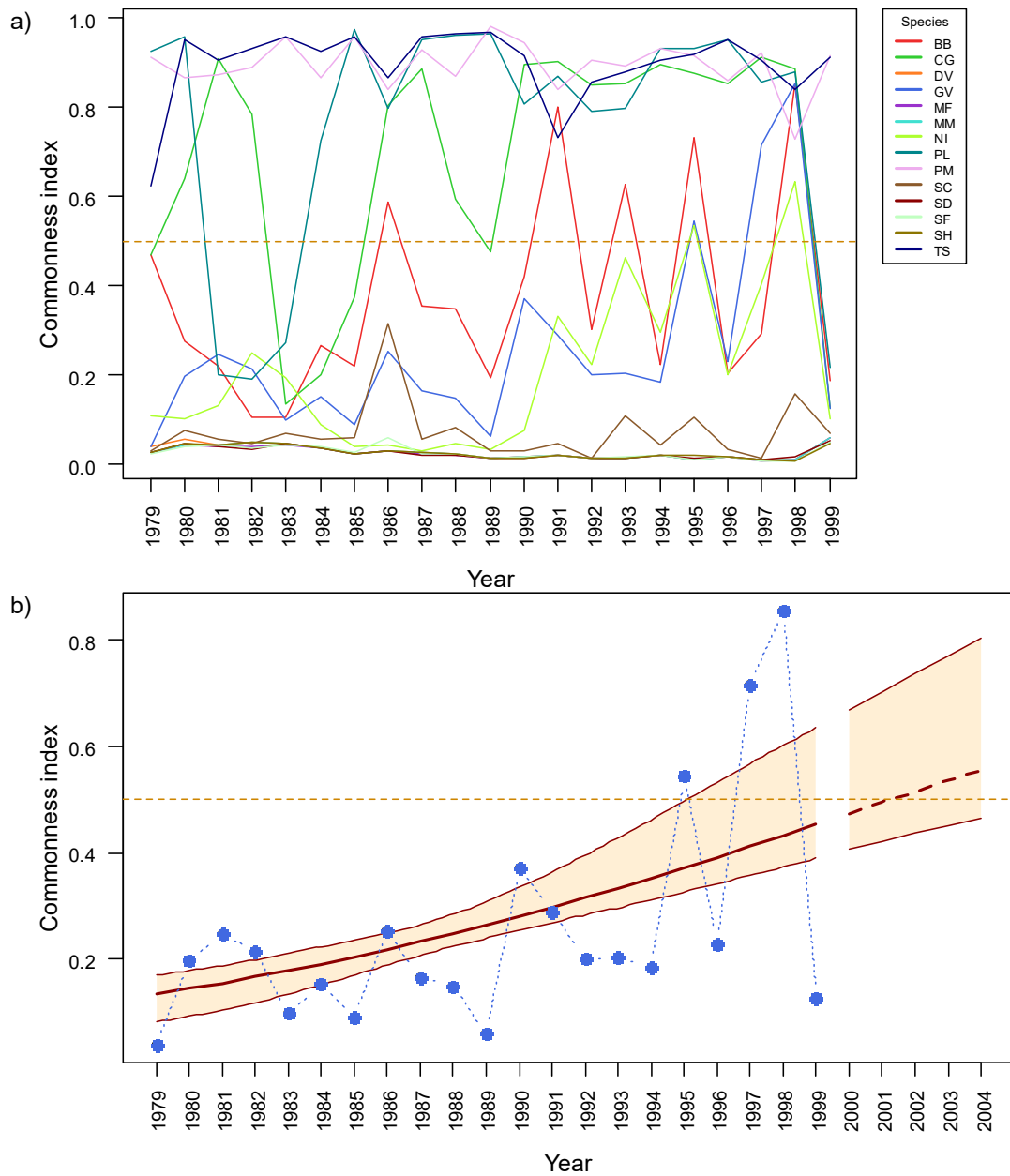
191 absences).

11

FIGURE 3 Fuzzy quantification of common and rare species in a community of 14 mammal species sampled at the Powdermill Biological Station from 1979 to 1999 (Merrit, 2013). (a) Variation of commonness indices in the study period. (b) Generalized Additive Model describing the variation in commonness of *Glaucomys volans* (GV) in the study period and predicted change (2001-2004). Blue points: observed values. Thick red line: fitted and predicted model (continuous and stippled lines). Thin red line: 95% confidence interval of the model. Stippled orange line: 0.5 threshold between rare and common species. Species abbreviations: **BB**, *Blarina brevicauda*; **CG**, *Clethrionomys gapperi*; **DV**, *Didelphis virginiana*; **GV**, *Glaucomys volans*; **MF**, *Mustela frenata*; **MM**, *Marmota monax*; **NI**, *Napaeozapus insignis*; **PL**, *Peromyscus leucopus*; **PM**, *Peromyscus maniculatus*; **SC**, *Sorex cinereus*; **SD**, *Sorex dispar*; **SF**, *Sorex fumeus*; **SH**, *Sorex hoyi*; **TS**, *Tamias striatus*.

12

192   The C_is indicated that two and seven of the 14 species could be categorized

193   consistently as common and rare, respectively, throughout the study period. The C_is of

194   the remaining five varied considerably over the years (Fig. 3a). We modelled the change

195   in commonness of a species in the latter group, *Glaucomys volans* (GV), by fitting a

196   Generalized Additive Model. Bootstrap replicates to fit 95% confidence intervals to the

197   model were generated with `fuzzyqBoot`. The fitted model was also used to predict the

198   change in C_i five years ahead. (Details on model fitting are given in an accompanying R

199   script. See Data Availability below.) The model suggests a progressive increase in C_i of

200   *G. volans* over the years and predicts a similar increase rate in the following years (Fig.

201   3b).

202   ***Example 2. Parasite communities of the so-iuy mullet in native and introduced areas***

203   We compared the patterns of commonness and rarity of helminth communities of the

204   so-iuy mullet (*Planiliza haematocheilus* in its native (Sea of Japan) and introduced (Sea

205   of Azov and Black Sea) areas (Llopis-Belenguer, Blasco-Costa, Balbuena, Sarabeev, &

206   Stouffer, 2020). We used here 12 and 7 surveys in the introduced and native areas,

207   respectively, in which the number of fish sampled was ≥ 20, totalling 378 and 192 fish,

208   respectively. Based on biogeographical evidence, species absences within the native and

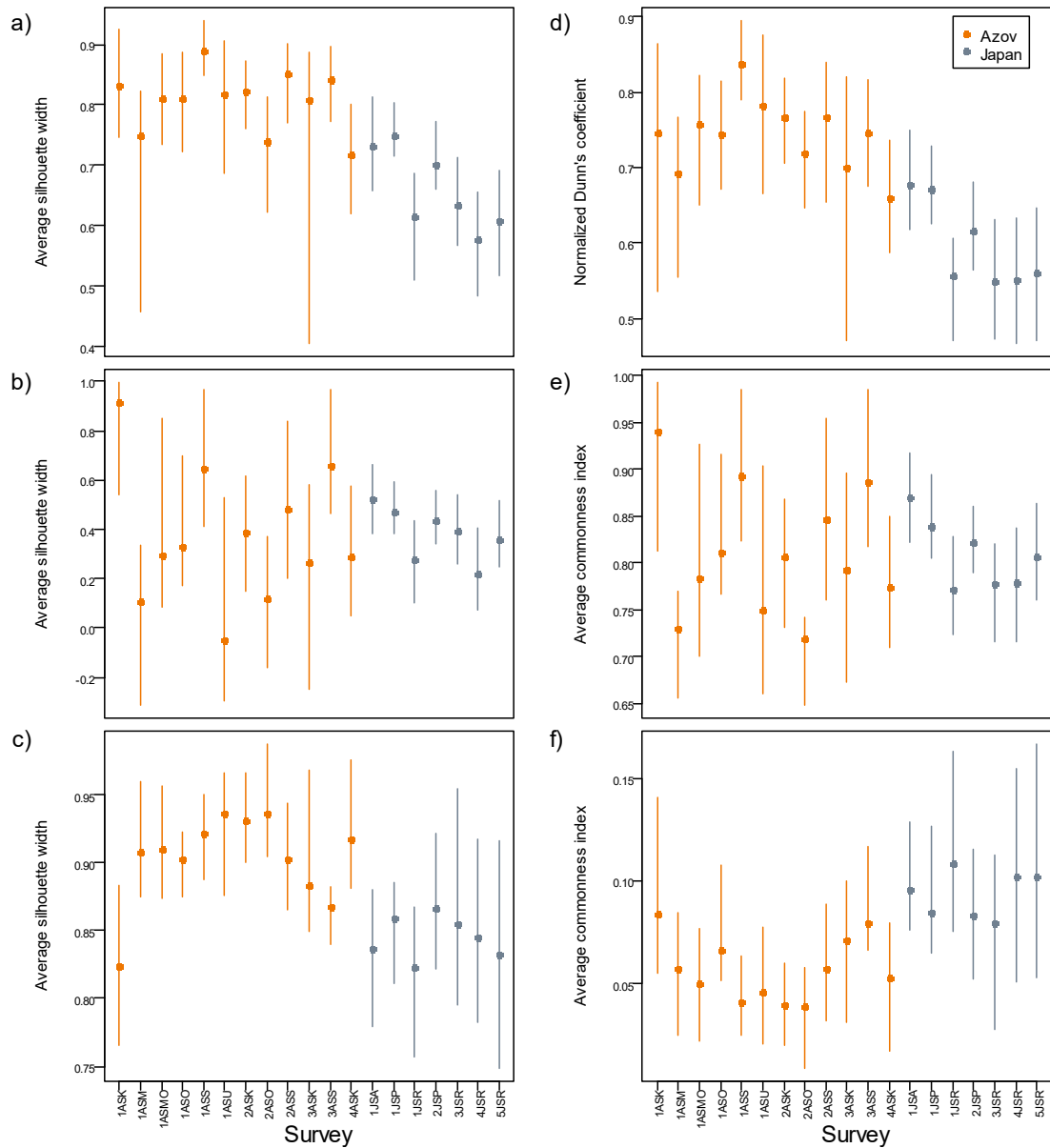209   introduced areas were treated as random zeros (Kostadinova, 2008).

FIGURE 4 Community-level estimates (points) and bias-corrected and accelerated 95% (Efron & Tibshirani, 1994) confidence intervals (error bars) computed with 1,000 replicates bootstrapping the plots of the abundance matrix of helminth communities of *Planiliza haematocheilus* in 7 native (Japan Sea) and 12 introduced (Azov and Black Seas) surveys. (a) Average silhouette widths of all species. (b) Idem common species. (c) Idem rare species. (d) Normalized Dunn's coefficient. (e) Average commonness indices of common species. (f) Idem rare species.

210   We used fuzzyq to compute $\bar{S}$, $\bar{S}_C$, $\bar{S}_R$, $\bar{C}_C$, $\bar{C}_R$ and D' of each survey and

211   fuzzyqBS and fuzzyqCI to estimate their 95% confidence intervals (Fig. 4).

212   Differences in these metrics between surveys in the native and introduced areas were

213   evaluated by Mann-Whitney tests. In the introduced area, rare species had significantly

14

214   higher $\bar{S}_R$ and lower $\bar{C}_R$ than in the native one ($p = 0.0012$ and $p = 0.0003$, respectively).

215   Differences in $\bar{S}$ and D' were also significant ($p = 0.0003$ and $p = 0.0002$, respectively),

216   indicating a clearer distinction between common and rare species in the introduced area

217   than in the native one. By contrast, there was no evidence for significant differences

218   between areas in $\bar{S}_C$ and $\bar{C}_C$ ($p = 0.71$ and $p = 0.97$, respectively). These results conform

219   to previous work that indicates that the introduction of the mullet so-iuy in the new area

220   entailed a deep structural change in its helminth communities (Sarabeev, Balbuena, &

221   Morand, 2017; Llopis-Belenguer et al., 2020). Most native species were lost and only

222   two *Ligophorus* spp. common in the native area were co-introduced and remained

223   common in the introduced area (Figs. S1, S2 in Supporting Information). So the

224   majority of species in the introduced area were acquired from local grey mullet species

225   (Sarabeev et al., 2017). Since newly acquired parasite species are expected to lack

226   specific adaptations to the new host, this would account for their pronounced rarity

227   compared to rare species in the native area (Sarabeev, Balbuena, & Morand, 2018).

228   **Effect of number of sites**

229   A key question for users interested in comparing different communities is whether the

230   community-level estimates depend on the number of sites sampled. We examined this

231   issue using 20 datasets compiled in Calatayud et al. (2019) and (Jeliazkov et al., 2020)

232   involving 87+ sites. These included communities varying widely in taxonomic

233   composition, species richness and spatial scale. In addition, different abundance

234   currencies were employed. (See Table S1 in Supporting Information for details.) Being

235   N the total number of sites of a given dataset, global metrics ($\bar{S}, \bar{S}_C, \bar{S}_R, \bar{C}_C, \bar{C}_R$ and D')

236   were computed in most cases dropping successively 1, 2, 3, ..., N-10 sites randomly

237   drawn (without replacement) from the dataset. (In species-poor communities or

238   communities with very sparsely distributed species, the series was 1, 2, 3, ..., N-20).

15

239 Species absences in each draw were treated as structural, because our goal was to

240 examine how incomplete species coverage resulting from low sample sizes affected the
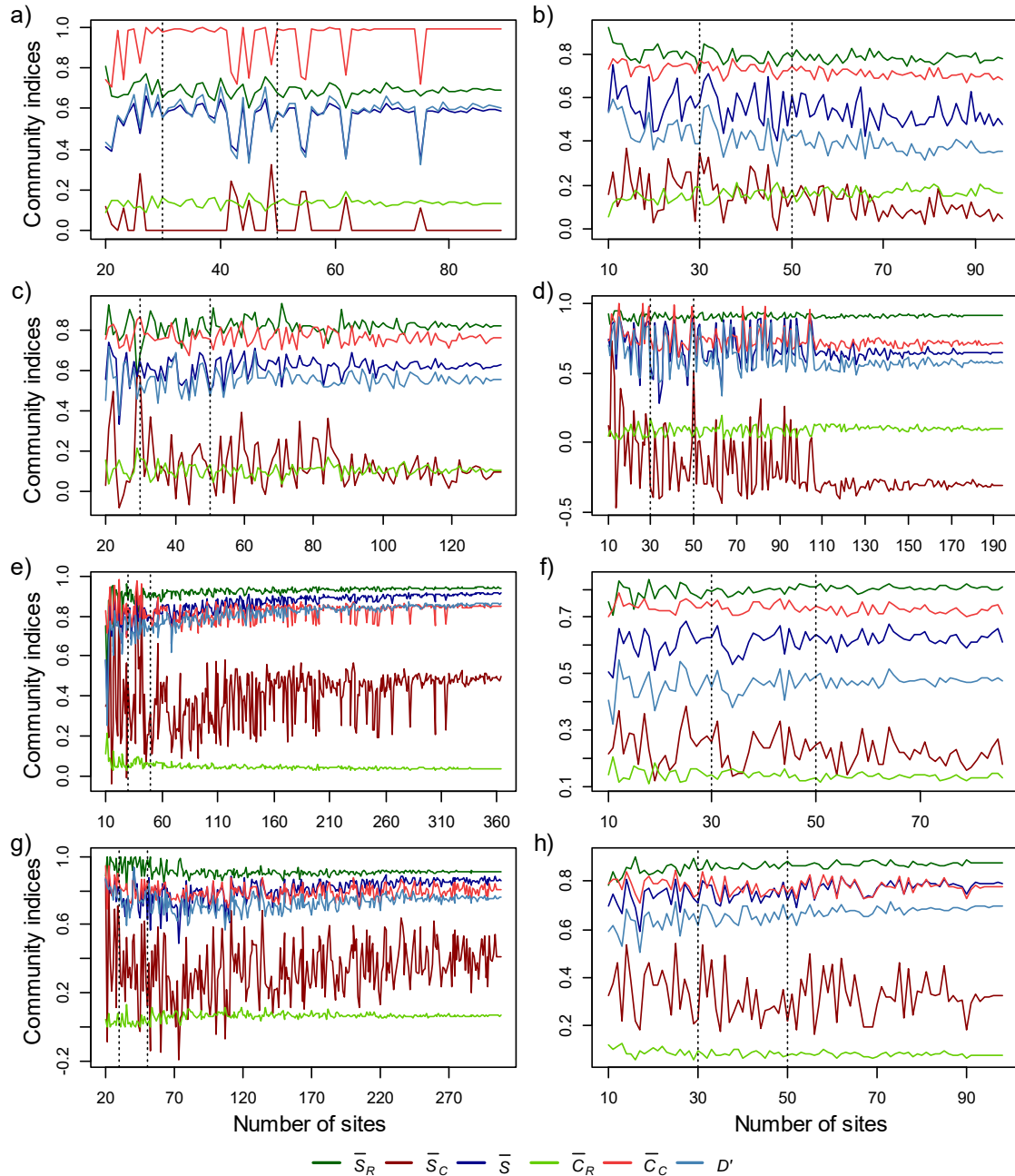
241 estimation of global metrics.

FIGURE 5 Variation of community-level metrics with number of sites in eight, 87+-site, databases

from Jeliazkov et al. (2020) (a-g) and Calatayud et al. (2019) (h): (a) BrindAmour2011a; (b)

Pavoine2011; (c) Jeliazkov2014; (d) Barbaro2009a; (e) Chmura2016; (f) Ribera2001; (g)

Goncalves2014a; (h) ants_data_Xavi_Darwin_A. Abbreviations: $\bar{S}_R$, average silhouette rare species;

$\bar{S}_C$, idem common species; $\bar{S}$, idem all species; $\bar{C}_R$, Commonness coefficient rare species; $\bar{C}_C$, idem

common species; *D'*, Normalized Dunn's coefficient. Details of these datasets are given in Table S1, Supporting Information.

242   Fig. 5 shows the variation of the global metrics with the number of sites in eight

243   of the 20 datasets. (Results for the remaining 12 datasets are given in Fig. S3,

244   Supporting Information.) No clear trend of variation with the number of sites was

245   apparent (Fig. 5 and Fig. S3, Supporting Information). Although in some datasets large

246   fluctuations occurred ($\bar{S}_C$ in particular was quite labile in some examples), variation in

247   global parameters did not seem related to sample size. The results suggest that 30 to 50

248   sites are sufficient to yield reliable estimates although bootstrapping should be used to

249   capture their variability.

250   **Final remarks**

251   FuzzyQ provides a new quantitative framework to study the distribution of common and

252   rare species in ecological communities. The approach supplies simple and intuitive

253   ecological indicators that can give both clear, actionable insights into the nature of

254   ecological communities and a powerful way to monitor quantitatively environmental

255   change on ecosystems. We show that the approach works satisfactorily with a wide

256   range of communities varying in species richness, dispersion and abundance currencies.

257   The only obvious limitation in its application is that fuzzy clustering requires that $k \geq$

258    $n/2$ -1 (Kaufman & Rousseeuw, 1990). As in our case $k = 2$, FuzzyQ cannot be applied

259   to communities composed of ≤ 5 species. In addition, the application of `fuzzyqBoot`

260   in communities with low number of species can lead to a number of null replicates

261   because of this limitation.

262   Comparison among communities is greatly facilitated by the fact that the method

263   is relatively independent of the number of sites or sampling units considered. However,

264   the use of FuzzyQ in comparative settings comes with an important caveat. Since

265   FuzzyQ is based on AORs and occupancy is known to vary with spatial scale (Hui,

17

266 Veldtman, & McGeoch, 2010; Steenweg, Hebblewhite, Whittington, Lukacs, &

267 McKelvey, 2018), differences in scaling can compromise comparison among

268 communities. In our second working example, helminth communities of individual fish

269 were evaluated as sites. Therefore, we consider that the comparison makes biological

270 sense. However, we cannot completely rule out that potential differences between the

271 native and introduced areas in fish mobility could introduce a hidden bias (Steenweg et

272 al., 2018).

273   Likewise, it has been shown that rarity at coarse scales can be substantially

274 biased because species of similar occupancies at that level may have very different

275 occupancies at finer scales (He & Condit, 2007). Thus, assessment and monitoring of

276 rarity should be performed at the appropriate scale for suitable conservation and

277 management plans. Nevertheless, for samples taken at nested spatial or temporal scales,

278 FuzzyQ provides a convenient tool to assess how scale affects patterns of commonness

279 and rarity. In addition, the approach is versatile as it can be readily adapted to other

280 categorizations (by considering more clusters) or to other criteria of rarity (by

281 introducing additional/different traits when computing the dissimilarity matrix).

282 Therefore, we consider that FuzzyQ is a potentially valuable analytical tool in

283 community ecology and conservation biology.

284 **Data Availability**

285 Table 2 provides information of the availability of the datasets used herein. R scripts

286 and R markdown files to run the illustrative examples, and the FuzzyQ R package are

287 available at https://ligophorus.github.io/FuzzyQ/ (**DOI no. pending**).

288 **Acknowledgements**

294   **TABLE 2** Availability of the datasets used with FuzzyQ.

Datasets	Source	Reference
Mammals Powdermill	https://doi.org/10.6073/pasta/83c888854e239a79597999895bb61cfe	Merritt (2011)
Helminths so-iuy mullet	https://doi.org/10.7910/DVN/IWIKOL	Llopis-Belenguer et al. (2020)
Large communities	https://idata.idiv.de/ddm/Data/ShowData/286	Jeliazkov et al. (2020)
Large communities	https://doi.org/10.6084/m9.figshare.9906092	Calatayud et al. (2019)

295   **Authors' contributions**

296   JAB conceived the idea. JAB and SM developed the theory and outlined the study. CM, CLB

297   and JAB wrote the scripts and developed the package. CLB, IBC, VLS and SM set and verified

298   the analytical methods. All authors contributed decisively to shape the research, provided

299   critical feedback on the drafts and gave final approval for submission.

300   **References**

301   Alroy, J. (2015). The shape of terrestrial abundance distributions. *Science Advances*, *1*(8),

302   e1500082. doi:10.1126/sciadv.1500082

303   Antão, L. H., Connolly, S. R., Magurran, A. E., Soares, A., & Dornelas, M. (2017). Prevalence

304   of multimodal species abundance distributions is linked to spatial and taxonomic breadth.

305   *Global Ecology and Biogeography*, *26*(2), 203–215. doi:10.1111/geb.12532

306   Arnan, X., Gaucherel, C., & Andersen, A. N. (2011). Dominance and species co-occurrence in

307   highly diverse ant communities: a test of the interstitial hypothesis and discovery of a three-

308   tiered competition cascade. *Oecologia*, *166*(3), 783–794. doi:10.1007/s00442-011-1919-y

309   Blasco-Moreno, A., Pérez-Casany, M., Puig, P., Morante, M., & Castells, E. (2019). What does

310   a zero mean? Understanding false, random and structural zeros in ecology. *Methods in*

311   *Ecology and Evolution*, *10*(7), 949–959. doi:10.1111/2041-210X.13185

312   Calatayud, J., Andivia, E., Escudero, A., Melián, C. J., Bernardo-Madrid, R., Stoffel, M., …

313   Madrigal-González, J. (2019). Positive associations among rare species and their persistence

19

314   in ecological assemblages. *Nature Ecology & Evolution*, 1–6. doi:10.1038/s41559-019-1053-

315   5

316   Davies, K. F., Margules, C. R., & Lawrence, J. F. (2000). Which traits of species predict

317   population declines in experimental forest fragments? *Ecology*, *81*(5), 1450–1461.

318   doi:10.1890/0012-9658(2000)081[1450:WTOSPP]2.0.CO;2

319   Dee, L. E., Cowles, J., Isbell, F., Pau, S., Gaines, S. D., & Reich, P. B. (2019). When do

320   ecosystem services depend on rare species? *Trends in Ecology & Evolution*, *34*(8), 746–758.

321   doi:10.1016/j.tree.2019.03.010

322   Efron, B., & Tibshirani, R. J. (1994). An introduction to the bootstrap. New York, NY:

323   Chapman and Hall. doi:10.1201/9780429246593

324   Enquist, B. J., Feng, X., Boyle, B., Maitner, B., Newman, E. A., Jørgensen, P. M., … McGill,

325   B. J. (2019). The commonness of rarity: Global and future distribution of rarity across land

326   plants. *Science Advances*, *5*(11), eaaz0414. doi:10.1126/sciadv.aaz0414

327   Flather, C. H., & Sieg, C. H. (2007). Species rarity: definition, causes, and classification. In

328   M.G. Raphael & R. Molina (Eds.), *Conservation of rare or little-known species: biological,*

329   *social, and economic considerations* (pp. 40–66). Washington D.C.: Island Press.

330   Gaston, K. J. (1994). Rarity (1st ed., Vol. 13). Dordrecht, Netherlands: Springer.

331   doi:10.1007/978-94-011-0701-3

332   Gaston, K. J. (1997). What is rarity? In W.E. Kunin & K.J. Gaston (Eds.), *The biology of rarity.*

333   *Causes and consequences of rare-common differences* (pp. 30-47). Dordrecht, Netherlands:

334   Springer. doi:10.1007/978-94-011-5874-9_3

335   Gaston, K. J., Blackburn, T. M., Greenwood, J. J. D., Gregory, R. D., Quinn, R. M., & Lawton,

336   J. H. (2000). Abundance–occupancy relationships. *Journal of Applied Ecology*, *37*(s1), 39–

337   59. doi:10.1046/j.1365-2664.2000.00485.x

338   Gaston, K. J., & Fuller, R. A. (2008). Commonness, population depletion and conservation

339   biology. *Trends in Ecology & Evolution*, *23*(1), 14–19. doi:10.1016/j.tree.2007.11.001

20

340   Gaston, K. J., & He, F. (2011). Species occurrence and occupancy. In A.E. Magurran & B.J.

341   McGill (Eds.), *Biological diversity: frontiers in measurement and assessment* (pp. 141–151).

342   Oxford, U.K.: Oxford University Press.

343   Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*,

344   *27*(4), 857–871. doi:10.2307/2528823

345   Gray, J. S., Bjørgesæter, A., & Ugland, K. I. (2005). The impact of rare species on natural

346   assemblages. *Journal of Animal Ecology*, *74*(6), 1131–1139. doi:10.1111/j.1365-

347   2656.2005.01011.x

348   Hanski, I. (1991). Single-species metapopulation dynamics: concepts, models and observations.

349   *Biological Journal of the Linnean Society*, *42*(1–2), 17–38. doi:10.1111/j.1095-

350   8312.1991.tb00549.x

351   He, F., & Condit, R. (2007, July). The distribution of species: occupancy, scale, and rarity.

352   doi:10.1017/CBO9780511814938.005

353   Hui, C., Veldtman, R., & McGeoch, M. A. (2010). Measures, perceptions and scaling patterns

354   of aggregated species distributions. *Ecography*, *33*(1), 95–102. doi:10.1111/j.1600-

355   0587.2009.05997.x

356   Jeliazkov, A., Mijatovic, D., Chantepie, S., Andrew, N., Arlettaz, R., Barbaro, L., … Chase, J.

357   M. (2020). A global database for metacommunity ecology, integrating species, traits,

358   environment and space. *Scientific Data*, *7*, 6. doi: 10.1038/s41597-019-0344-7

359   Kaufman, L., & Rousseeuw, P. J. (1990). Finding groups in data: An introduction to cluster

360   analysis. New York, N.Y.: John Wiley & Sons. doi:10.1002/9780470316801

361   Kostadinova, A. (2008). A checklist of macroparasites of *Liza haematocheila* (Temminck &

362   Schlegel) (Teleostei: Mugilidae). *Parasites & Vectors*, *1*(1), 48. doi:10.1186/1756-3305-1-48

363   Leitão, R. P., Zuanon, J., Villéger, S., Williams, S. E., Baraloto, C., Fortunel, C., … Mouillot,

364   D. (2016). Rare species contribute disproportionately to the functional structure of species

365   assemblages. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1828),

366   20160084. doi:10.1098/rspb.2016.0084

21

367   Llopis-Belenguer, C., Blasco-Costa, I., Balbuena, J. A., Sarabeev, V., & Stouffer, D. B. (2020).

368   Native and invasive hosts play different roles in host–parasite networks. *Ecography*, *43*(4).

369   doi:10.1111/ecog.04963

370   Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2019). cluster: Cluster

371   analysis basics and extensions. R package version 2.1.0. Retrieved 29 July 2020, from

372   https://CRAN.R-project.org/package=cluster

373   Magurran, A. E., & Henderson, P. A. (2011). Commonness and rarity. In A.E. Magurran & B.J.

374   McGill (Eds.), *Biological diversity: frontiers in measurement and assessment* (pp. 97–104).

375   Oxford, U.K.: Oxford University Press.

376   McGill, B. (2011). Species abundance distributions. In A.E. Magurran & B.J. McGill (Eds.),

377   *Biological diversity: frontiers in measurement and assessment* (pp. 105–122). Oxford, U.K.:

378   Oxford University Press.

379   McGill, B. J., Rampal S. Etienne, John S. Gray, David Alonso, Marti J. Anderson, Habtamu

380   Kassa Benecha, … Ethan P. White. (2007). Species abundance distributions: moving beyond

381   single prediction theories to integration within an ecological framework. *Ecology Letters*,

382   *10*(10), 995–1015. doi:10.1111/j.1461-0248.2007.01094.x

383   Merritt, J. (2018). Long term mammal data from Powdermill biological station 1979-1999 [Data

384   set]. Environmental Data Initiative.

385   doi:10.6073/PASTA/101D5D3DEC9C688A7FECA3AB2E969369

386   R Core Team (2020). R: A language and environment for statistical computing. R Foundation

387   for Statistical Computing, Vienna, Austria. https://www.R-project.org/. Accessed 7 August

388   2020.

389   Sarabeev, V., Balbuena, J. A., & Morand, S. (2017). Testing the enemy release hypothesis:

390   Abundance and distribution patterns of helminth communities in grey mullets (Teleostei:

391   Mugilidae) reveal the success of invasive species. *International Journal for Parasitology*,

392   *47*, 687–696. doi:10.1016/j.ijpara.2017.05.006

393   Sarabeev, Volodimir, Balbuena, J. A., & Morand, S. (2018). *International Journal for*

394   *Parasitology*, *48*(9–10), 793–803. doi:10.1016/j.ijpara.2018.04.001

395   Steenweg, R., Hebblewhite, M., Whittington, J., Lukacs, P., & McKelvey, K. (2018). Sampling

396   scales define occupancy and underlying occupancy–abundance relationships in animals.

397   *Ecology*, *99*(1), 172–183. doi:10.1002/ecy.2054

398   Violle, C., Thuiller, W., Mouquet, N., Munoz, F., Kraft, N. J. B., Cadotte, M. W., … Mouillot,

399   D. (2017). Functional rarity: The ecology of outliers. *Trends in Ecology & Evolution*, *32*(5),

400   356–367. doi:10.1016/j.tree.2017.02.002

401   Werner, E. E., Davis, C. J., Skelly, D. K., Relyea, R. A., Benard, M. F., & McCauley, S. J.

402   (2014). Cross-scale interactions and the distribution-abundance relationship. *PLOS ONE*,

403   *9*(5), e97387. doi:10.1371/journal.pone.0097387

404   Williamson, M., & Gaston, K. J. (2005). The lognormal distribution is not an appropriate null

405   hypothesis for the species–abundance distribution. *Journal of Animal Ecology*, *74*(3), 409–

406   422. doi:10.1111/j.1365-2656.2005.00936.x

407   Yu, J., & Dobson, F. S. (2000). Seven forms of rarity in mammals. *Journal of Biogeography*,

408   *27*(1), 131–139. doi:10.1046/j.1365-2699.2000.00366.x