

Mapping diversity indices: not a trivial issue

Granger Victoria^{1, *}, Bez Nicolas², Fromentin Jean-Marc³, Meynard Christine⁴, Jadaud Angelique³,
Mérigot Bastien¹

¹ Université Montpellier; UMR 9190 MARBEC "Marine biodiversity, exploitation & conservation" (CNRS, IFREMER, IRD, UM); Sète France

² Institut de Recherche pour le Développement; UMR 9190 MARBEC "Marine biodiversity, exploitation & conservation" (CNRS, IFREMER, IRD, UM); Sète France

³ Institut Français de Recherche pour l'Exploitation de la Mer; UMR 9190 MARBEC "Marine biodiversity, exploitation & conservation" (CNRS, IFREMER, IRD, UM); Sète France

⁴ Virginia Institute of Marine Science; College of William and Mary; P.O. Box 1346 Gloucester Point VA 23062 USA

* Corresponding author : Victoria Granger, email address : granger.vica@gmail.com

Abstract :

Mapping diversity indices, that is estimating values in all locations of a given area from some sampled locations, is central to numerous research and applied fields in ecology. Two approaches are used to map diversity indices without including abiotic or biotic variables: (i) the indirect approach, which consists in estimating each individual species distribution over the area, then stacking the distributions of all species to estimate and map a posteriori the diversity index, (ii) the direct approach, which relies on computing a diversity index in each sampled locations and then to interpolate these values to all locations of the studied area for mapping. For both approaches, we document drawbacks from theoretical and practical viewpoints and argue about the need for adequate interpolation methods. First, we point out that the indirect approach is problematic because of the high proportion of rare species in natural communities. This leads to zero-inflated distributions, which cannot be interpolated using standard statistical approaches. Secondly, the direct approach is inaccurate because diversity indices are not spatially additive, that is the diversity of a studied area (e.g. region) is not the sum of the local diversities. Therefore, the arithmetic variance and some of its derivatives, such as the variogram, are not appropriate to ecologically measure variation in diversity indices. For the direct approach, we propose to consider the β -diversity, which quantifies diversity variations between locations, by the mean of a β -gram within the interpolation procedure. We applied this method, as well as the traditional interpolation methods for comparison purposes on different faunistic and floristic data sets collected from scientific surveys. We considered two common diversity indices, the species richness and the Rao's quadratic entropy, knowing that the above issues are true for complementary species diversity indices as well as those dealing with other biodiversity levels such as genetic diversity. We conclude that none of the approaches provided an accurate mapping of diversity indices and that further methodological developments are still needed. We finally discuss lines of research that may resolve this key issue, dealing with conditional simulations and models taking into account biotic and abiotic explanatory variables.

Keywords : interpolation methods, map, quadratic entropy, spatial statistics, species diversity, species richness, β -diversity

49 **Introduction**

50 Given the increasing rate of change in biological diversity, mediated by ever
51 increasing direct human pressures and global environmental change, species diversity is of
52 major interest both in theoretical and applied studies (Lavergne *et al.* 2010; Sterling *et al.*
53 2010; Dawson *et al.* 2011; Thuiller *et al.* 2011; Cardinale *et al.* 2012). In this context, accurate
54 mapping of diversity indices is a key tool to study spatio-temporal variations in natural
55 communities, to identify priority areas of protection and to support effective conservation
56 planning (Devictor *et al.* 2010; Merckx *et al.* 2010; Thuiller *et al.* 2011, Stuart-Smith *et al.*
57 2014).

58 Mapping a diversity index consists in estimating values of the index at all locations
59 of a given area in which only some locations have been sampled. Ecologists used two main
60 approaches for spatial interpolation of diversity index and its mapping without including
61 abiotic or biotic variables: the indirect and direct approaches. However, both approaches
62 have some drawbacks from theoretical and practical viewpoints.

63 The indirect approach, called "predict first, assemble later" (Ferrier & Guisan 2006), consists
64 in layering presence or abundance of each individual species (which have been modelled)
65 and then computing *a posteriori* a diversity index by combining all layers. However, the
66 scarcity of many species in natural communities leads to a high proportion of zero-inflated
67 distributions, which can hardly be interpolated using standard interpolation techniques, such
68 as kriging (Heilbron 1994; Morfin *et al.* 2012) and more generally, all regression techniques.
69 This clearly makes the indirect approach difficult to apply in practice.

70 The direct approach, called "assemble first, and predict later" (Ferrier *et al.* 2002; Ferrier &
71 Guisan 2006; Mokany *et al.* 2011), consists in computing directly a diversity index at sampled
72 locations and then in interpolating those values at unsampled locations in each grid point of

73 the studied area. Although scientific literature provides a plethora of interpolation
74 techniques (e.g., James & McCulloch 1990), their use needs particular cautious when dealing
75 with diversity indices. Unlike other quantitative variables, diversity indices are not spatially
76 additive, i.e., the diversity of a studied area (e.g., region) is not the sum of the local
77 diversities. Note that, eventhough they are connected, the (spatial) additivity to which we
78 refer here is not the additive partitioning of regional γ -diversity into the mean local α -
79 diversities and β diversity as described by Lande (1996). Additivity of indices has been
80 discussed from a theoretical point of view (Keylock 2005; Hoffmann 2006), but considering
81 this property in a mapping context is lacking. For instance, let us consider the species
82 richness at two locations A and B being equal to 5 and 2, respectively, while 2 species are
83 shared between the two locations. If the species richness would be additive, its value for the
84 pooled area of locations A and B would be equal to 7 (Carrasco *et al.* 2008). However, since
85 these two locations have two species in common, the actual species richness is equal to 5.
86 This simplistic example shows that the species richness of an area that includes several
87 locations is different than the sum of the species richness in all locations if some locations
88 share similar species. This index would be additive only if all the locations have no species in
89 common (e.g., Keylock 2005; Hoffmann 2006), which is a very restrictive situation in natural
90 communities. This problem is thus related to the similarity in species composition between
91 locations, i.e., β -diversity (Magurran 2004; Anderson *et al.* 2011; Pavoine 2012).
92 Spatial additivity is particularly critical for interpolation techniques (and thus mapping), as
93 they rely on linear combinations of values of diversity indices (Michalakopoulos &
94 Panagiotou 1997; Rivoirard *et al.* 2000). When applied on additive variables, like absolute
95 abundance, traditional spatial interpolation methods (such as kriging, distance weighting
96 etc.) are consistent with the fact that the index value of an area composed of several pooled

97 locations is equal to the mean value of the index in these locations. Thus, considering
98 arithmetic mean of interpolated diversity indices would be accurate only if the index is
99 spatially additive, regardless of the interpolation method being applied. To circumvent this
100 problem, we proposed, in the frame of the direct approach, to combine geostatistical
101 techniques and β -diversity concept to interpolate local α -diversity indices over a given area
102 (Couteron & Pelissier 2004). This goal is not to estimate the “total species richness of an
103 area” (γ -diversity, e.g., Ugland et al. 2003).

104

105 Note that the lack of spatial additivity does not only affect the number of species, but also
106 the relative abundance (proportion) that are used in other facets of species diversity.

107 Appendix A summarizes results of a simple test of additivity conducted on other
108 complementary widely used diversity indices. None of them strictly respect this property.
109 Therefore, we applied the direct and the indirect approaches using two common diversity
110 indices (the species richness and the Rao's quadratic entropy) and four datasets of different
111 faunistic and floristic groups collected from scientific surveys. We finally discuss lines of
112 research that may resolve the problems raised.

113

114 **Material and methods**

115 **Data**

116 We considered four different datasets.

- 117 A) The first dataset reports demersal fish abundance in the Gulf of Lions (France)
118 located in North-western Mediterranean Sea (3°W to 5.2°E; 42.5 to 43.8°N). The 66
119 scientific bottom trawls analysed have been carried out in 2012, in the frame of the
120 international MEDITS program (Bertrand *et al.* 2002). 186 species properly sampled

- 121 by the fishing gear were considered (Gaertner *et al.* 2010, 2013). Abundance was
122 standardized to 1 km², for each species caught (Morfin *et al.* 2012; Gaertner *et al.*
123 2013).
- 124 B) The second one reports woody plant species abundance in the central Western Ghats
125 region, Karnataka, India (74.25°–75.5° E; 15.25°–13.5° N) in a network of 96 sampling
126 sites. This data provides abundance on 334 tree species collected in 96 sampling sites
127 during 1996–1997 (merged for this study) (Ramesh *et al.* 2010).
- 128 C) The third dataset reports butterfly diversity and abundance in Boulder County Open
129 Space, Colorado, USA (105.1°–105.3° W; 39.9°–40.1° N) collected over 66 sites in the
130 years 1999 and 2000 (merged for this study). The data contain butterfly species
131 diversity and individual species' abundance of 58 species from five butterfly families
132 (Oliver *et al.* 2006).
- 133 D) The fourth dataset consists of vascular plant and bryophyte species composition and
134 plant and soil biogeochemical data in Great Britain (6.3° W to 1.25° E; 50.5°N to
135 60.2°N) collected over 56 acid grasslands in 2002. These data provide abundance on
136 391 vascular species plants (Stevens *et al.* 2011) .

137

138 **Diversity indices**

139 Generally, more than one index is necessary to describe species diversity (Pavoine &
140 Bonsall 2011). Different indices indeed allow to quantify different facets, mainly species
141 number, evenness, or more complex variations considering taxonomic, phylogenetic and/or
142 functional differences between species (Devictor *et al.* 2010; Meynard *et al.* 2011; Pavoine
143 2012; Stuart-Smith *et al.* 2013). Here, we considered two diversity indices widely used in
144 ecology of communities and in diversity mapping studies (e.g. Devictor *et al.* 2010; Stuart-

145 Smith et al. 2013), knowing that the spatial additivity issue is true for other indices as well as
146 those dealing with other biodiversity levels, such as genetic diversity (see end of the
147 Introduction section and Appendix A). First, we computed species richness, the most
148 intuitive and popular index in both marine and terrestrial diversity studies. This index was
149 applied on all four above datasets.

150 The second application dealt with Rao's quadratic entropy index (Rao 1982), which gained
151 popularity because of its mathematical proprieties and its wide range of applications
152 (Pavoine 2012). This index is defined as:

$$153 \quad Q = \sum_{i=1}^S \sum_{j=1}^S p_i p_j d_{ij}$$

154 where p_i and p_j are the relative abundance of the i^{th} and j^{th} species, d_{ij} the difference (e.g.,
155 taxonomic, phylogenetic or functional dissimilarity/distance) between two species i and j
156 stored in a distance matrix. In our study, distances between species were constructed using
157 the Linnaean taxonomic classification. The distance between two species from the same
158 genus was set to 1, two species from the same family but different genus was 2, and so on.
159 We considered a taxonomy including 5 levels (species, genus, family, order and class).
160 Taxonomic distances were normalized between 0 and 1, providing an index's range between
161 these values. This index was applied only on the first data set of demersal fish abundance in
162 the Gulf of Lions (dataset A), due to availability of taxonomic data to compute quadratic
163 entropy.

164

165 **Statistical analysis**

166 **The direct approach:**

167 The direct approach aims thus at modelling directly the diversity indices. In other
 168 words, the local α -diversity values at all locations of an area are mapped through an explicit
 169 spatial linear interpolation method. Spatial autocorrelation of the index (the statistical
 170 relationship among points) is the main element for producing maps in geostatistical
 171 interpolation by a self-sufficient method (without explanatory variables). Among spatial
 172 interpolation methods, kriging is the best linear estimator (Matheron 1963), i.e., the one of
 173 minimum variance. It is based on the spatial structure of the α -diversity which is quantified
 174 by the empirical semi-variogram (i.e., computed on sampled data, Matheron 1963, Wagner
 175 2003):

$$176 \quad V(h) = \frac{1}{2|N(h)|} \sum_{N(h)} (\alpha_{s_i} - \alpha_{s_j})^2 \quad (1)$$

177 where $N(h)$ is the number of pairs of locations separated by a distance h , α_{s_i} and α_{s_j} are the
 178 values of the α -diversity in locations i and j . Then a theoretical variogram (e.g., linear,
 179 spherical or Gaussian variogram model) fitting the empirical variogram is used as the
 180 interpolation function, i.e., to estimate values between locations (Matheron 1963, Wagner
 181 2003).

182 However, the variogram, i.e., arithmetic spatial variance of index value between
 183 locations (α -diversity), does not quantify ecologically diversity variations (see example
 184 described in the introduction). Thus, replacing it by a β -diversity (i.e., an adequate measure
 185 of species replacement among locations) should ensure a more accurate quantification of
 186 diversity variation among locations. We thus propose an alternative methodological
 187 framework for interpolating diversity indices, called β -kriging. It consists in replacing the
 188 weighting function usually expressed as the spatial variance above (i.e., theoretical

189 variogram) by a spatial β -diversity model fitting the empirical β -diversity model previously
 190 proposed (Couteron & Pelissier 2004). We call it β -gram, which is defined as:

$$191 \quad \beta(h) = \frac{1}{|N(h)|} \sum_{N(h)} \beta(s_i - s_j) \quad (2)$$

192

193 Equation 2 can be viewed as an empirical variogram, but representing the average pairwise
 194 diversity variation between locations separated by a distance h , with $\beta_i(s_i - s_j)$ being the
 195 variation (β -diversity) between each pair of locations (Appendix B provides details on the β -
 196 kriging procedure). Independently of the index used to measure the diversity, γ -diversity
 197 (here considered as the total diversity of two locations) can be partitioned into local α -
 198 diversity (i.e., mean of diversity of the two locations) and β -diversity reflecting the variation
 199 in diversity between the two locations (Magurran 2004; Anderson *et al.* 2011; Pavoine
 200 2012). Two partitions are commonly considered to compute β -diversity: the additive (Lande
 201 1996) and the multiplicative partitioning (Whittaker 1972) (Appendix C). The advantage of
 202 such partitioning is that they can be applied to a wide range of indices. Because both led to
 203 the same results for the direct/indirect approach, we focused on the additive partitioning
 204 where $\gamma = \bar{\alpha} + \beta$ (Lande 1996, for the related results see Appendix C for more details).

205 We applied kriging and β -kriging methods on species richness and Rao's quadratic
 206 entropy indices.

207

208 **The indirect approach:**

209 This approach consists in modelling each species distribution and then computing *a*
 210 *posteriori* a diversity index by combining all species distributions of the community. We
 211 interpolated species distributions by inverse distance weighting. Estimates were obtained as

212 a weighted average of the density values from the neighbouring values, their contribution
213 being weighted as an inverse function of the distance to the kernel. We applied inverse
214 distance weighting which allowed modelling distribution of all 186 species in Mediterranean
215 fish dataset without modelling their spatial autocorrelation, in contrast to kriging. We thus
216 made the assumption of a unique weighting function for all species distributions (including
217 the rare ones).

218

219 **Methods performance:**

220 The performance of each interpolation technique, in terms of the accuracy in
221 estimating diversity index value, was assessed by comparing the deviations of estimates
222 from the observed data through the use of the leave-one-out cross-validation (Stone 1974).
223 In such procedure, a given sampled location is deleted from the data set and is estimated by
224 performing the method, using the remaining locations. The operation is then repeated for all
225 sampled locations. The estimated values are finally compared to the observed field values by
226 mean of scatter plots, deviations from the first bisector (i.e., $y=x$, the case where observed
227 and predicted values are equal), slopes of the linear regression and coefficients of
228 determination R^2 .

229

230 **Results**

231 **The direct approach**

232 Patterns between β -grams and variograms computed for the direct approach based
233 on species richness on the four datasets were different (Fig. 1). Species replacement (i.e., β -
234 diversity) was relatively high at even very short distances (strong nugget effects in the β -
235 grams), while species richness was less contrasted at the same scale (see variograms in Fig.

236 1). The results of the leave-one-out cross-validation procedure are presented in Fig. 2. For
237 all datasets, regression slopes between observed and estimated values ranged between 0.89
238 and 1.05 for kriging, between 1.13 and 1.65 for β -kriging according to the dataset
239 considered. R^2 values remained rather low ($0.22 < R^2 < 0.41$) for both procedures. The scatter
240 plots of observed values *versus* predicted values were highly dispersed around the first
241 bisector, showing that both classical kriging and β -kriging had poor prediction performances.
242 The range of estimated values by β -kriging was different, and generally more restricted, than
243 by classical kriging. For instance about the Forest India dataset, while observed values
244 ranged between 1 and 59 species, the estimated values by kriging ranged between 17.32
245 and 43.88 species and between 7.75 and 34.41 species by β -kriging. The differences in
246 estimated values between classical kriging and β -kriging directly came from the differences
247 between the theoretical β -gram and variogram (red dotted lines in Fig. 1).

248

249 For Rao's quadratic entropy, the direct approach was applied only to the
250 Mediterranean fish data, due to availability of species taxonomic differences data (see
251 materials and methods section). The variogram and β -gram were also different (see Fig B.1).
252 Both interpolation methods provided again poor prediction performances (Fig 3.a).
253 Regression lines for both kriging and β -kriging procedures presented a slope inferior to 1
254 (0.77 for kriging and 0.6 for β -kriging) and both intercepts for both regressions were equal to
255 0.13 and 0.24 respectively, again far from the first bisector. Furthermore, R^2 values were
256 very low, i.e., equal to 0.15 and 0.08 for kriging and β -kriging scatterplots, respectively. The
257 estimated values ranged between 0.36 and 0.72 for β -kriging and between 0.39 and 0.71 for
258 classical kriging, while the observed values were much wider, i.e., between 0.03 and 0.75
259 (Fig. 3.a).

260

261 **The indirect approach**

262 For Rao's quadratic entropy, the indirect approach was applied only to
263 Mediterranean fish dataset (see above). The results are presented in Fig 3.b. The linear
264 regression between predicted and estimated Rao's quadratic entropy by indirect approach
265 presented a slope of 0.6, and the same range of regression values that those obtained by
266 direct approach (Fig 2.b.). The intercept for the regression was equal to 0.21. Furthermore,
267 R^2 value was equal to 0.04. The distribution of observed quadratic entropy values ranged
268 from 0.03 to 0.75, while the predicted values only ranged between 0.5 and 0.72. In addition,
269 there is a bias close to 10% of the observed mean.

270

271 **Discussion**

272 In this study we emphasized that interpolating and mapping diversity indices (i.e.,
273 estimating values at all locations to map the studied area from some sampled locations) is
274 problematic, and we illustrated this on several datasets collected from scientific surveys.

275 First, we have seen that the traditional direct approach cannot provide accurate
276 mapping because of the lack of spatial additivity of diversity indices. We thus proposed an
277 alternative procedure, called the β -kriging, by combining geostatistical tools and β -diversity
278 concept to model the spatial variations in diversity index. However, even if β -kriging is more
279 ecologically founded, it does not really improve the predictions of species richness or
280 quadratic entropy indices made by classical kriging, using a variogram.

281 Although β -kriging fails to predict accurately diversity index, β -gram can be considered as an
282 interesting tool to study diversity variations between spatially distant locations of a given
283 area (Couteron & Pelissier 2004, Pavoine 2005, Shen *et al.* 2013, Parmentier *et al.* 2014,

284 Pélissier and Goreaud accepted). Notably β -gram can be implemented to study the spatial
285 structure of functional or phylogenetic diversity in the framework of the spatial point
286 processes (Shen *et al.* 2013), as proposed by Pélissier and Goreaud (accepted). For instance,
287 the null hypothesis of species equivalence (i.e., absence of spatial structure in species
288 relatedness) can be tested by using a Monte Carlo randomization procedure shuffling the
289 between-species distances (i.e., permuting simultaneously the rows and columns in the d_{ij}
290 matrix). Then the observed β -gram (i.e., diversity index computed on each pairwise sampled
291 locations in function of spatial distances between these locations) is compared to the
292 confidence envelopes generated by the Monte Carlo randomization to determine if the null
293 hypothesis can be, or not, accepted (see for more details Shen *et al.* 2013; Pélissier and
294 Goreaud accepted).

295

296 Second, regarding the indirect approach, most species of a given assemblage and/or
297 community are known to present low to very low levels of abundance and/or occurrence
298 (Gaston 1994; Martin *et al.* 2005). Modelling the spatial structure (e.g., the variogram) and
299 the spatial distributions (for instance through kriging) of those rare species could hardly be
300 performed with traditional statistical tools (see examples of experimental variograms for
301 several species in Appendix D). For instance, for the Medits dataset that include 186 fish
302 species, the probability of presence for each species shows that the vast majority of species
303 are rare or extremely rare (65% of the species distributions get more than 95% of 0), or
304 present high punctual abundance (see Appendix E). In this case, kriging based on species
305 spatial autocorrelation is no longer operational for spatial interpolation for most species, as
306 already stressed by Morfin *et al.* (2012). Note that the issue of zero-inflated data is actually a

307 common feature in ecological study and it is restricted to marine assemblages (Martin *et al.*
308 2005).
309 The use of the indirect approach can further create a bias in predicted index values relative
310 to the observed ones (see for instance the application on quadratic entropy). It can be
311 attributed to the fact that the indirect approach smoothes the presence or abundance of the
312 species and their distribution range. In other words, it creates presence in locations where
313 species were not observed. Furthermore, this smoothing can hardly capture some
314 discontinuities in the spatial distribution (e.g., highly fragmented and/or disturbed area). In
315 such situation, a k-nearest neighbors algorithm's method could be applied (Altman, 1992),
316 knowing that the capacity of the method to deal with discontinuities decreases with the
317 increasing number of neighbors considered.

318 Consequently, the indirect approach could only be applied on the most abundant
319 (common) species in communities, which seriously restraints the objectives of any diversity
320 study by shedding the light on a few species, and that may not be the ones of conservation
321 concern.

322

323 **Perspectives**

324

325 Following the above statements, we suggest two directions of possible improvements.

326 First, the bias identified in the indirect approach comes from interpolation method
327 and more certainly from the fact that diversity indices are non linear with regards to the
328 individual layers. For instance, in the case of the Rao's quadratic entropy index, there is a
329 quadratic link between species proportion and the index. A way of avoiding bias, is to
330 simulate each species distribution conditionally on the observed data (Chilès and Delfiner,

331 2012), and to use these simulations rather than the interpolations. In the same way that the
332 mean of log-transformed data is not the log-transformed mean, the diversity index will be
333 estimated by the mean of the transformed simulations and not by the transformed mean. It
334 is worth remaining here that the aim of a conditional simulation is to create a distribution for
335 each species that mimic the true spatial heterogeneity of the variable. This contrasts with
336 interpolation (e.g., kriging) which estimates the expected species distributions (i.e., a
337 smoothed version of the study variable). Conditional simulations preserve the variance of
338 the observed data without smoothing and represent different equally possible spatial
339 distribution of the studied variable. It would be a viable alternative when the spatial
340 structure of each species is known. However, this method is also challenged by zero-inflated
341 data to map rare species in the same ways as kriging.

342

343 Second, an alternative strategy to map diversity indices is to use models including
344 abiotic and/or biotic explanatory variables (e.g., generalized linear or additive models
345 GLM/GAM, machine learning methods, co-kriging methods, Olden *et al.* 2008, Ballesteros-
346 Mejia *et al.* 2013, Hernández-Stefanoni *et al.* 2011). It is acknowledged that three main
347 drivers act on species distributions and diversity at different spatial scales, i.e., (i) abiotic
348 constraints, (ii) dispersal and (iii) biotic interactions (e.g., predation, competition and
349 facilitation, see Loreau & Mouquet 1999; Soberon 2007). Ignoring in models a combination
350 of these explicative variables may lead to a certain part of unexplained variability
351 (Boulangéat *et al.* 2012; Cavieres *et al.* 2014). However, some of these variable values are
352 not always known for every species in natural communities (e.g., biotic interactions or
353 dispersal limitations). When biotic information are not available, it is usual to only deal with
354 abiotic predictors. For instance, Leathwick *et al.* (2006) mapped species richness of demersal

355 fish considering only environmental variables in GAMs and Boosted Regression Trees (BRTs)
356 for which the explained deviances varied between 45% and 60%. Bhattarai and Vetaas (2003)
357 applied GLMs to study variation in species richness of different groups of herbaceous in
358 function of environmental variables for which explained deviance of models highly varied
359 according to the group (between 14% and 62%).

360 When biotic information are available, the indirect approach could benefit from the
361 development of species interaction distributions models, using multispecies interactions
362 matrix (Kissling *et al.* 2012). Pellissier *et al.* 2013 proposed a combined approach including
363 both biotic and abiotic predictors. They implemented food web models that can infer the
364 potential interaction links between species as a constraint in species distribution models that
365 include environmental predictors. More broadly, Thuiller *et al.* 2013 proposed a promising
366 framework for species distribution modelling, derived from metapopulation theory, which
367 accounts for abiotic constraints, dispersal, biotic interactions as well as local adaptation
368 under changing environmental conditions.

369
370 The difficulty to accurately map indices by the direct or indirect approach is directly
371 transposable to other levels of diversity than species diversity, such as genetic diversity, for
372 which indices have different names and input data but identical mathematical formula. For
373 instance, in genetic diversity, allelic richness, Nei and Π indices are the equivalent of species
374 richness, Simpson diversity $1-D$ and quadratic entropy respectively (Nei 1973; Nei & Li 1979).

375
376 In conclusion, we showed that mapping index by interpolation methods used in the
377 frame of direct or indirect approach may not be accurate because diversity indices are not
378 spatially additive and many species in natural communities are rare. The use of the indirect

379 approach comes with the large burden of having to ignore or at least downplay the rarest
380 species for which individual species distribution model is hardly feasible. Unfortunately, it
381 differs from the crucial aim to consider all species of communities, and these rare species
382 are usually of particular interest, notably from a conservation perspective, but also for
383 ecosystem functioning as recently demonstrated (Mouillot *et al.* 2013). In the frame of the
384 direct approach, the β -gram can be an interesting tool to study diversity variations between
385 spatially distant locations of a given area, but the β -kriging procedure failed to predict
386 accurately diversity index, as other traditional interpolation methods. Thus, considerable
387 progress has still to be made and we highlight that conditional simulations and models
388 taking into account biotic and abiotic explanatory variables could provide a solution for an
389 accurate diversity indices mapping.

390

391 **Acknowledgements**

392 We thank the Editor Pedro Peres-Neto, two anonymous reviewers and S. Pavoine for their
393 comments on an earlier version of the manuscript. V.G. is supported by a Presidency PhD
394 grant of University Montpellier 2. This work is part of the Groupe de Recherche (GDR) 3645
395 Statistical Ecology of CNRS.

396

397 **Data accessibility**

398 Demersal fish: contact Angelique.jadaud@ifremer.fr

399 Woody plant: <http://www.esapubs.org/archive/ecol/E087/061/metadata.htm>

400 Butterfly: <http://esapubs.org/archive/ecol/E091/216/default.htm>

401 Vascular plant: <http://esapubs.org/archive/ecol/E092/128/default.htm>

402 **References**

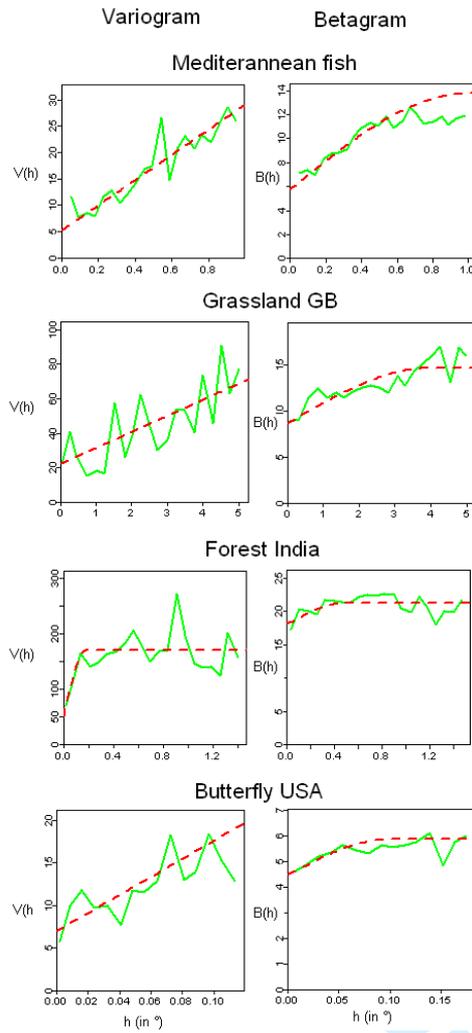
- 403 Altman, N. (1992). An introduction to kernel and nearest-neighbor nonparametric
404 regression. *The American Statistician*, **46**, 175–185.
- 405 Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., Freestone, A.L., Sanders,
406 N.J., Cornell, H. V, Comita, L.S., Davies, K.F., Harrison, S.P., Kraft, N.J.B., Stegen, J.C. &
407 Swenson, N.G. (2011). Navigating the multiple meanings of β diversity: a roadmap for
408 the practicing ecologist. *Ecology letters*, **14**, 19–28.
- 409 Ballesteros-Mejia, L., Kitching, I.J., Jetz, W., Nagel, P. & Beck, J. (2013). Mapping the
410 biodiversity of tropical insects: Species richness and inventory completeness of African
411 sphingid moths. *Global Ecology and Biogeography*, **22**, 586–595.
- 412 Bhattarai, K. & Vetaas, O. (2003). Variation in plant species richness of different life forms
413 along a subtropical elevation gradient in the Himalayas, east Nepal. *Global Ecology and*
414 *Biogeography*, **12**, 327–340.
- 415 Bertrand, J., De Sola, L., Papaconstantinou, C., Relini, G. & Souplet, A. (2002). The general
416 specifications of the MEDITS surveys. *Scientia Marina*, **66**, 9–17.
- 417 Boulangeat, I., Gravel, D. & Thuiller, W. (2012). Accounting for dispersal and biotic
418 interactions to disentangle the drivers of species distributions and their abundances.
419 *Ecology letters*, **15**, 584–93.
- 420 Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A.,
421 Mace, G.M., Tilman, D., Wardle, D. a, Kinzig, A.P., Daily, G.C., Loreau, M., Grace, J.B.,
422 Larigauderie, A., Srivastava, D.S. & Naeem, S. (2012). Biodiversity loss and its impact on
423 humanity. *Nature*, **486**, 59–67.
- 424 Cavieres, L.A., Brooker, R.W., Butterfield, B.J., Cook, B.J., Kikvidze, Z., Lortie, C.J., Michalet,
425 R., Pugnaire, F.I., Schöb, C., Xiao, S., Anthelme, F., Björk, R.G., Dickinson, K.J.M.,
426 Cranston, B.H., Gavilan, R., Gutiérrez-Giron, A., Kanka, R., Maalouf, J.P., Mark, A.F.,
427 Noroozi, J., Parajuli, R., Phoenix, G.K., Reid, A.M., Ridenour, W.M., Rixen, C., Wipf, S.,
428 Zhao, L., Escudero, A., Zaitchik, B.F., Lingua, E., Aschehoug, E.T. & Callaway, R.M. (2014).
429 Facilitative plant interactions and climate simultaneously drive alpine plant diversity.
430 *Ecology Letters*, **17**, 193–202.
- 431 Couteron, P. & Pelissier, R. (2004). Additive apportioning of species diversity: towards more
432 sophisticated models and analyses. *Oikos*, **107**, 215–221.
- 433 Dawson, T.P., Jackson, S.T., House, J.I., Prentice, I.C. & Mace, G.M. (2011). Beyond
434 Predictions: Biodiversity Conservation in a Changing Climate. *Science*, **332**, 53–58.
- 435 Devictor, V., Mouillot, D., Meynard, C., Jiguet, F., Thuiller, W. & Mouquet, N. (2010). Spatial
436 mismatch and congruence between taxonomic, phylogenetic and functional diversity:

- 437 the need for integrative conservation strategies in a changing world. *Ecology Letters*,
438 **13**, 1030–1040.
- 439 Ferrier, S., G. Watson, and J. Pearce (2002). Extended statistical approaches to modelling
440 spatial pattern in biodiversity in northeast New SouthWales. I. Species-level modelling.
441 *Biodiversity and Conservation* (11), 2275–2307
- 442 Ferrier, S. & Guisan, A. (2006). Spatial modelling of biodiversity at the community level.
443 *Journal of Applied Ecology*, **43**, 393–404.
- 444 Gaertner, J.-C., Maiorano, P., Mérigot, B., Colloca, F., Politou, C.-Y., Gil De Sola, L., Bertrand,
445 J. a, Murenu, M., Durbec, J.-P., Kallianiotis, A. & Mannini, A. (2013). Large-Scale
446 Diversity of Slope Fishes: Pattern Inconsistency between Multiple Diversity Indices. *PLoS*
447 *one*, **8**, e66753.
- 448 Gaertner, J.-C., Mérigot, B., Rélini, G., Bertrand, J.A., Mazouni, N., Politou, C.-Y., Gil De Sola,
449 L., Kallianiotis, A., Carpentieri, P., Murenu, M., Durbec, J.-P., Vrgoc, N. & Ungaro, N.
450 (2010). Reproducibility of the multi-component aspect of species diversity across
451 different areas and scales: towards the constitution of a shortlist of complementary
452 indices for monitoring fish diversity? *Ecography*, **33**, 1123–1135.
- 453 Gaston, K.J. (1994). *Rarity*.
- 454 Heilbron, D. (1994). Zero-altered and other regression models for count data with added
455 zeros. *Biometrical Journal*, **36**, 531–547.
- 456 Hoffmann, S. (2006). *Concavity and additivity in diversity measurement: re-discovery of an*
457 *unknown concept*. Univ., FEMM.
- 458 James, F.C. & McCulloch, C.E. (1990). Multivariate Analysis in Ecology and Systematics:
459 Panacea or Pandora's Box? (C.E. McCulloch, Ed.). *Annual Review of Ecology and*
460 *Systematics*, **21**, 129–166.
- 461 Journel, A.G. (1974). Geostatistics for Conditional Simulation of Ore Bodies. *Economic*
462 *Geology*, **69**, 673–687.
- 463 Keylock, C. (2005). Simpson diversity and the Shannon Wiener index as special cases of a
464 generalized entropy. *Oikos*, **109**, 203–207.
- 465 Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., Mcinerly, G.J., Montoya,
466 J.M., Römermann, C., Schiffers, K., Schurr, F.M., Singer, A., Svenning, J.C., Zimmermann,
467 N.E. & O'Hara, R.B. (2012). Towards novel approaches to modelling biotic interactions in
468 multispecies assemblages at large spatial extents. *Journal of Biogeography*, **39**, 2163–
469 2178.
- 470 Lande, R. (1996). Statistics and partitioning of species diversity, and similarity among
471 multiple communities. *Oikos*, **76**, 5–13.

- 472 Lavergne, S., Mouquet, N., Thuiller, W. & Ronce, O. (2010). Biodiversity and Climate Change:
473 Integrating Evolutionary and Ecological Responses of Species and Communities. *Annual*
474 *Review of Ecology Evolution and Systematics*, **41**, 321–350.
- 475 Leathwick, J., Elith, J., Francis, M., Hastie, T. & Taylor, P. (2006). Variation in demersal fish
476 species richness in the oceans surrounding New Zealand: an analysis using boosted
477 regression trees. *Marine Ecology Progress Series*, **321**, 267–281.
- 478 Loreau, M. & Mouquet, N. (1999). Immigration and the Maintenance of Local Species
479 Diversity. *The American naturalist*, **154**, 427–440.
- 480 Magurran, A.E. (2004). *Measuring Biological Diversity*. Blackwell Science, Oxford.
- 481 Martin, T.G., Wintle, B. a, Rhodes, J.R., Kuhnert, P.M., Field, S. a, Low-Choy, S.J., Tyre, A.J. &
482 Possingham, H.P. (2005). Zero tolerance ecology: improving ecological inference by
483 modelling the source of zero observations. *Ecology letters*, **8**, 1235–46.
- 484 Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, **58**, 1246–1266.
- 485 Merckx, B., Van Meirvenne, M., Steyaert, M., Vanreusel, A., Vincx, M. & Vanaverbeke, J.
486 (2010). Mapping nematode diversity in the Southern Bight of the North Sea. *Marine*
487 *Ecology Progress Series*, **406**, 135–145.
- 488 Meynard, C.N., Devictor, V., Mouillot, D., Thuiller, W., Jiguet, F. & Mouquet, N. (2011).
489 Beyond taxonomic diversity patterns: how do α , β and γ components of bird functional
490 and phylogenetic diversity respond to environmental gradients across France? *Global*
491 *Ecology and Biogeography*, **20**, 893–903.
- 492 Michalakopoulos, T. & Panagiotou, G. (1997). Information Technologies in the Minerals
493 Industry: Proceedings of the First International Conference on Information Technologies
494 in the Minerals Industry Via the Internet
- 495 Mokany, K., Harwood, T.D., Overton, J.M., Barker, G.M. & Ferrier, S. (2011). Combining α and
496 β diversity models to fill gaps in our knowledge of biodiversity. *Ecology letters*, **14**,
497 1043–51.
- 498 Morfin, M., Fromentin, J.-M., Jadaud, A. & Bez, N. (2012). Spatio-temporal patterns of key
499 exploited marine species in the northwestern mediterranean sea. *PLoS ONE*, **7**.
- 500 Mouillot, D., Bellwood, D.R., Baraloto, C., Chave, J., Galzin, R., Harmelin-Vivien, M., Kulbicki,
501 M., Lavergne, S., Lavorel, S., Mouquet, N., Paine, C.E.T., Renaud, J. & Thuiller, W. (2013).
502 Rare Species Support Vulnerable Functions in High-Diversity Ecosystems. *PLoS Biology*,
503 **11**.
- 504 Nei, M. (1973). Analysis of Gene Diversity in Subdivided Populations. *Proceedings of the*
505 *National Academy of Sciences of the United States of America*, **70**, 3321–3323.

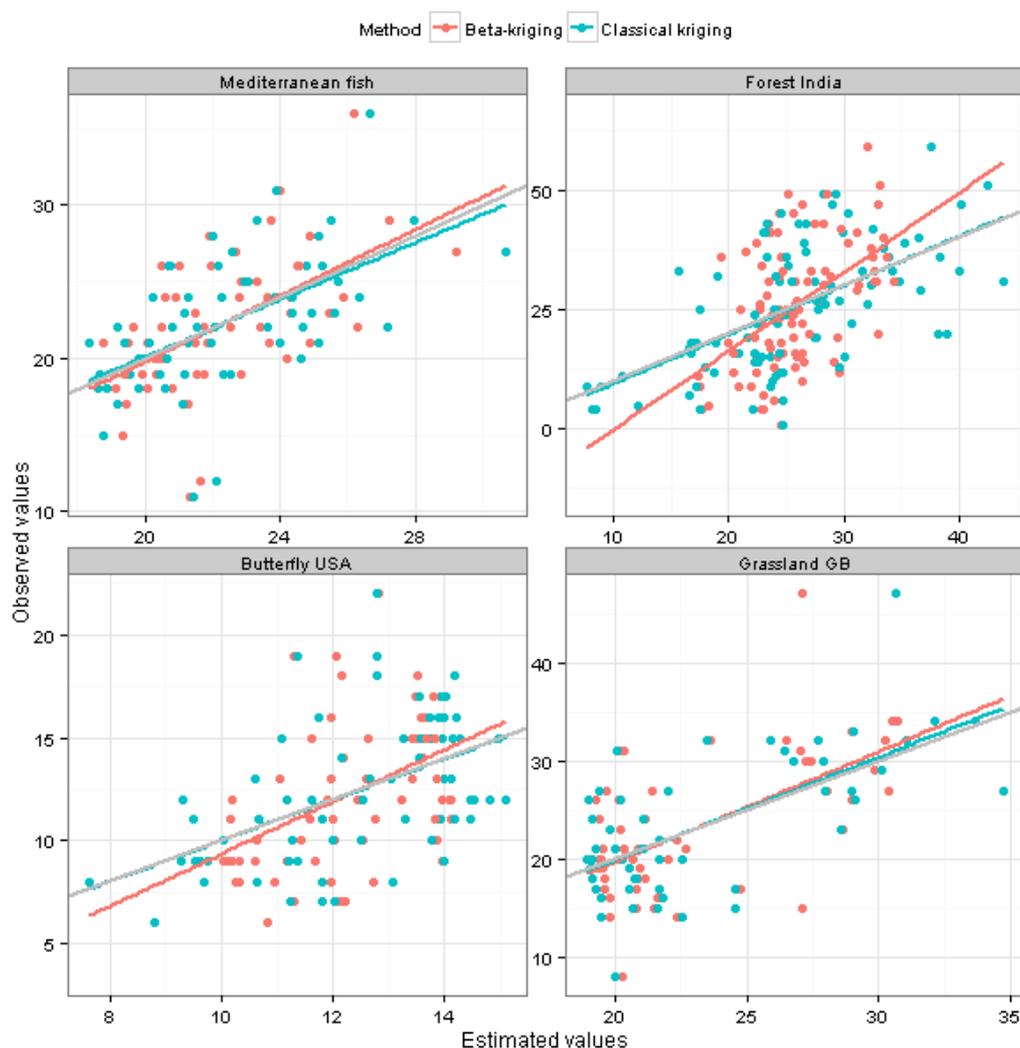
- 506 Nei, M. & Li, W.H. (1979). Mathematical model for studying genetic variation in terms of
507 restriction endonucleases. *Proceedings of the National Academy of Sciences of the*
508 *United States of America*, **76**, 5269–5273.
- 509 Oliver, J., Prudic, K. & Collinge, S. (2006). Boulder County Open Space butterfly diversity and
510 abundance. *Ecology*, **87**, 85721.
- 511 Parmentier, I., Réjou-Méchain, M., Chave, J., Vleminckx, J., Thomas, D.W., Kenfack, D.,
512 Chuyong, G.B. & Hardy, O.J. (2014). Prevalence of phylogenetic clustering at multiple
513 scales in an African rain forest tree community. *Journal of Ecology*, **102**, 1008–1016.
- 514 Pavoine, S. (2012). Clarifying and developing analyses of biodiversity: towards a
515 generalisation of current approaches. *Methods in Ecology and Evolution*, **3**, 509–518.
- 516 Pavoine, S. & Bonsall, M.B. (2011). Measuring biodiversity to explain community assembly: a
517 unified approach. *Biological Reviews of the Cambridge Philosophical Society*, **2**, 93–111.
- 518 Pavoine, S. (2005). Méthodes statistiques pour la mesure de la biodiversité. Ph. D. thesis.
519 Université Lyon 1, 391p.
- 520 Pellissier, L., Rohr, R.P., Ndiribe, C., Pradervand, J.N., Salamin, N., Guisan, A. & Wisz, M.
521 (2013). Combining food web and species distribution models for improved community
522 projections. *Ecology and Evolution*, **3**, 4572–4583.
- 523 Pélissier, R. & Goreaud, F. (accepted). ads Package for R: A Fast Unbiased Implementation of
524 the K-function Family for Studying Spatial Point Patterns in Irregular-shaped Sampling
525 Windows. *Journal of Statistical Software*.
- 526 Ramesh, B.R., Swaminath, M.H., Patil, S. V., Pélissier, R., Venugopal, P.D., Aravajy, S.,
527 Elouard, C. & Ramalingam, S. (2010). Forest stand structure and composition in 96 sites
528 along environmental gradients in the central Western Ghats of India. *Ecology*, **91**, 3118–
529 3118.
- 530 Rao, C. (1982). Diversity and dissimilarity coefficients: a unified approach. *Theoretical*
531 *Population Biology*, **43**, 24–43.
- 532 Rivoirard, J., Simmonds, J., Foote, K., Fernandes, P. & Bez, N. (2000). *Geostatistics for*
533 *estimating fish abundance*. Blackwell Science Ltd.
- 534 Shen, G., Wiegand, T., Mi, X. & He, F. (2013). Quantifying spatial phylogenetic structures of
535 fully stem-mapped plant communities. *Methods in Ecology and Evolution*, **4**, 1132–
536 1141.
- 537 Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species.
538 *Ecology Letters*, **10**, 1115–1123.

- 539 Sterling, E.J., Gomez, A. & Porzecanski, A.L. (2010). A systemic view of biodiversity and its
540 conservation: Processes, interrelationships, and human culture. *BioEssays news and*
541 *reviews in molecular cellular and developmental biology*, **32**, 1090–1098.
- 542 Stevens, C.J., Dupré, C., Dorland, E., Gaudnik, C., Gowing, D.J.G., Diekmann, M., Alard, D.,
543 Bobbink, R., Corcket, E., Mountford, J.O., Vandvik, V., Aarrestad, P.A., Muller, S. & Dise,
544 N.B. (2011). Grassland species composition and biogeochemistry in 153 sites along
545 environmental gradients in Europe. *Ecology*, **92**, 1544.
- 546 Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal*
547 *of the Royal Statistical Society. Series B (Methodological)*, **36**, 111–147.
- 548 Stuart-Smith, R.D., Bates, A.E., Lefcheck, J.S., Duffy, J.E., Baker, S.C., Thomson, R.J., Stuart-
549 Smith, J.F., Hill, N. a., Kininmonth, S.J., Airoidi, L., Becerro, M. a., Campbell, S.J., Dawson,
550 T.P., Navarrete, S. a., Soler, G. a., Strain, E.M. a., Willis, T.J. & Edgar, G.J. (2013).
551 Integrating abundance and functional traits reveals new global hotspots of fish
552 diversity. *Nature*, **501**, 539–542.
- 553 Thuiller, W., Lavergne, S., Roquet, C., Boulangeat, I., Lafourcade, B. & Araujo, M.B. (2011).
554 Consequences of climate change on the tree of life in Europe. *Nature*, **470**, 531–534.
- 555 Thuiller, W., Münkemüller, T., Lavergne, S., Mouillot, D., Mouquet, N., Schiffrers, K. & Gravel,
556 D. (2013). A road map for integrating eco-evolutionary processes into biodiversity
557 models. *Ecology Letters*, **16**, 94–105.
- 558 Wagner, H.H. (2003). Spatial covariance in plant communities: integrating ordination,
559 geostatistics, and variance testing. *Ecology* **84**, 1045-1057.
- 560 Whittaker, R.H. (1972). Evolution and measurement of species diversity (A.E. Magurran &
561 B.J. McGill, Eds.). *Taxon*, **21**, 213–251.
- 562
- 563



564

565 **Fig. 1.** Spatial structure of species richness measured by variogram and β -gram, for each
 566 dataset. Y-axis: green continuous curves represent the empirical variogram and the
 567 empirical β -diversity model computed from the additive partitioning for each pair of
 568 locations. The red dotted lines represent the theoretical continuous model (spherical or
 569 linear on the left and right panels, respectively) fitted to the empirical variogram or β -
 570 gram. x-axis: distance between locations in degree.

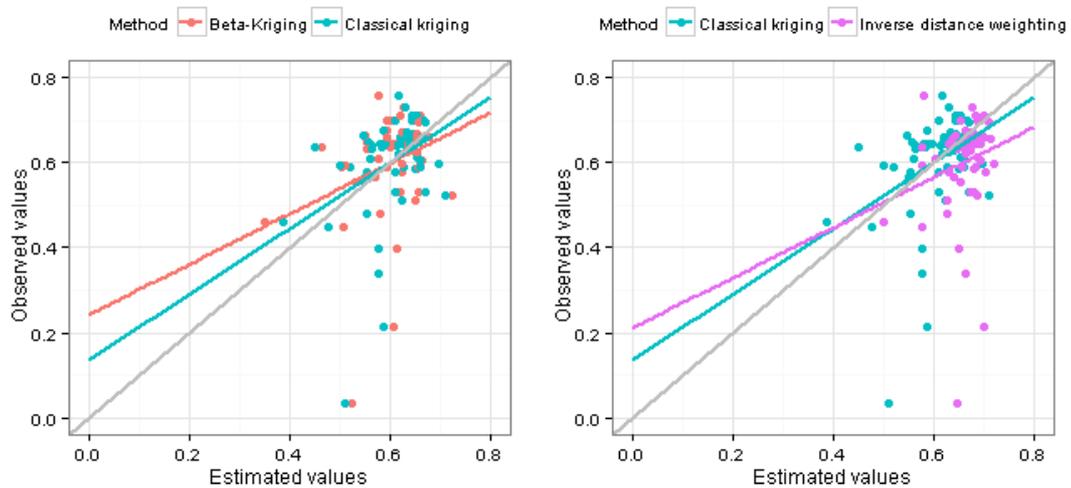


571

572 **Fig. 2.** Results of leave-one-out cross validation procedure for species richness. Procedure
 573 used to assess predictive performance of the direct approach by classical kriging (in blue)
 574 and additive β -kriging (in red) for species richness. Species richness computed on four
 575 datasets of different faunistic/floristic groups. The gray line represents the first bisector (i.e.,
 576 $y=x$), the case where observed and predicted index values are equal.

577

578



579

580 **Fig. 3.** Results of leave-one-out cross validation procedure for Rao's quadratic entropy.

581 Procedure used to assess predictive performance of the direct and the indirect approaches

582 for Rao's quadratic entropy. Rao's quadratic entropy computed only on Mediterranean

583 demersal fish data due to availability in species taxonomic differences.

584 a) the comparison between classical kriging (in blue) and additive β -kriging (in red)

585 procedure on Mediterranean fish species b) the comparison between the direct approach by

586 classical kriging (in blue) and the indirect approach (purple) by inverse distance weighting.

587 The gray line represents the first bisector (i.e., $y=x$), the case where observed and predicted

588 index values are equal.

589

1 **Granger V., Bez N., Fromentin J.M., Meynard C., Jadaud A., Mériqot B. (2015). Mapping**
 2 **diversity indices: not a trivial issue. *Methods in Ecology and Evolution*.**

3 **Supplementary material online**

4 **Appendix S1. Additivity test.**

5 **Table 1:** Usual names, formula of studied diversity indices. S is the number of species in the
 6 community, p_i is the relative abundance of the i^{th} species, N is the total number of
 7 individuals in a location, N_{max} is the number of individuals of the most abundant species, d_{ij}
 8 the difference (phylogenetic, functional or taxonomic) between two species i and j stored in
 9 a dissimilarity distance matrix. Column St A et St B represent the computed indices for
 10 locations A et B. Indices were computed for the aggregated area, by summing individuals of
 11 species and the mean value of this measure is presented, column $I(StB+StA)/2$. It can be
 12 compared to the mean value of two observations, column $mean(StA, StB)$. In this example,
 13 locations A and B have two species in common. A species distance matrix for Rao's quadratic
 14 entropy was simulated by a normal standardised distribution (not shown). Note that in case
 15 of all species equidistant, *i.e.* $d_{ij}=1$ for every species pairs, this index reduces to the Simpson
 16 diversity index, $1-D$. Abundance in location A are (1; 2; 8; 2; 1) and in location B (11; 7; 0; 0;
 17 0)

Diveristy component	Indices	Formula	St A	St B	$\frac{StA + StB}{2}$	$I(StA+StB)/2$
Species number	S	Number of species per location	5	2	3.5	2.5
Evenness and species number	H'	$-\sum_{i=1}^S \ln(p_i)p_i$	1.25	0.69	0.96	0.68
	E ^{H'}	$e^{-\sum_{i=1}^S \ln(p_i)p_i}$	3.5	1.95	2.73	1.93

	1-D	$1 - \sum_{i=1}^S p_i^2$	0.62	0.47	0.55	0.36
	Simpson's PIE	$\frac{N}{N-1}(1-D)$	0.67	0.50	0.59	0.37
	1/D	$\frac{1}{\sum_{i=1}^S p_i^2}$	2.65	1.91	2.28	1.74
Evenness	Heip's	$\frac{e^{H'} - 1}{S - 1}$	0.63	0.95	0.79	0.96
	Simpson's evenness	$\frac{1-D}{(1-\frac{1}{S})}$	0.78	0.95	0.84	0.60
	Berger-Parker	$\frac{N_{\max}}{N}$	0.57	0.61	0.59	0.19
Difference between species	Q (quadratic entropy)	$\sum_{i=1}^S \sum_{j=1}^S p_i p_j d_{ij}$	0.42	0.29	0.35	0.25

18

19

20 **Appendix S2. β -gram model.**

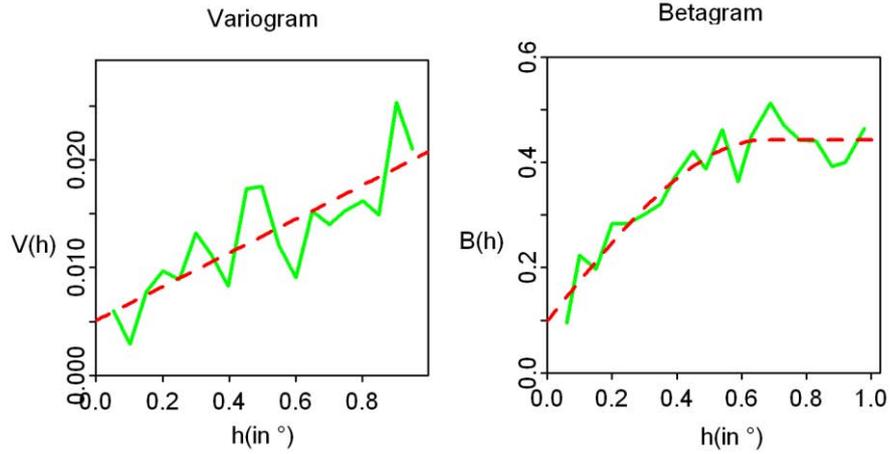
21 Variogram models must be chosen amongst the family of conditionally negative functions so
22 that the variance of any linear combination of the study variable is never negative. The
23 kriging weights obtained with such variograms correspond to the optimal weights, *i.e.* the
24 weights making the estimation variance minimum. The β -kriging implemented here uses the
25 same principle with a plugged-in model corresponding to the β -gram. In essence, β -kriging
26 weights correspond to the weights making the estimation variance minimum even though
27 such a variance gets no straightforward meaning. However, one needs to make sure that
28 these pseudo-variances are non negative. Moreover a β -gram is always positive, like any
29 variogram. Therefore existing variogram models are advised to be used.

30 The β -models were made of two components, the variation due to spatial dependence and
31 the random or "nugget" variation. The nugget of β -gram, *i.e.* the intercept in the β -gram
32 model, reflects both the spatial variation at smaller lags than the minimum sample
33 separation and the unexplained variation. If it exists, the range (of influence) is the
34 maximum distance at which diversity values are correlated. The coefficient of determination
35 (R^2) computed on adjustment of models to experimental β -gram was used to select a model.

36 One last thing to consider, in case of multiplicative partitioning of diversity, the value of β -
37 diversity ranges between 1 and the maximum number of locations used to compute β . In
38 this case standardization such as $\beta-1$ is required to avoid a nugget effect of 1 in the β -gram.

39 Kriging results are known to be sensitive to the shape of the model but not to its level. Only
40 the kriging estimation variance is affected by the level of the model. Whilst we were only
41 interested in interpolation and not in estimation variance, we used standardised β -grams.

42



43

44 **Fig S2.1.** Empirical and theoretical variogram and empirical and theoretical β -gram

45 computed for Rao's quadratic entropy for Mediterranean fish dataset. Y-axis : green

46 continuous curves represent the empirical β -diversity model computed from the additive

47 partitioning for each pair of locations and the empirical variogram. The red dotted lines

48 represent the continuous model (linear or spherical) fitted to the empirical β -gram or

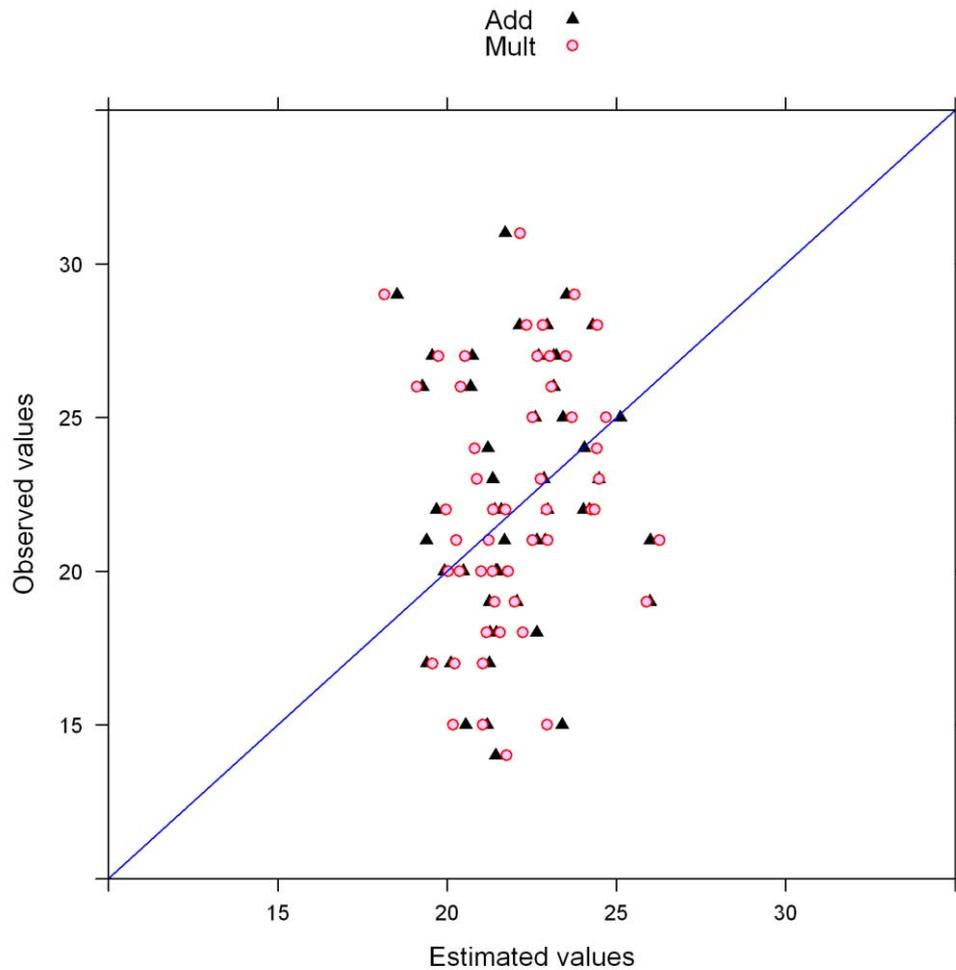
49 variogram. X-axis: distance between locations in degree.

50

51 **Appendix S3. The partitioning of β -diversity.**

52 Methodology for partitioning γ -diversity into α and β components has been long
53 debated (Loreau 2000; Anderson *et al.* 2011) and remains a sensitive ecological issue for
54 which no agreement has been achieved. In the multiplicative partitioning (Whittaker 1972),
55 the regional diversity γ , can be written as : $\gamma = \bar{\alpha} * \beta$. In the additive one (Lande 1996), $\gamma = \bar{\alpha} + \beta$
56 (Lande 1996). The partitioning depends on the mathematical framework of the used indices,
57 for instance some diversity indices do not support an additive partitioning because they do
58 not fulfil the concavity property (Jost 2007; Jost *et al.* 2010). While both partitioning
59 frameworks have a clear definition and interpretation for species richness, and they can be
60 linked mathematically, no arguments are available to advise to choose one between the
61 diversity partitioning frameworks. Using additive or multiplicative partitioning allows
62 considering unified approach to compute β -diversity for all the main indices available to
63 describe complementary diversity facets (*e.g.* species richness, Simpson's and quadratic
64 entropy) while using pair-wise dissimilarity indices (*e.g.* Jaccard) would be restricted to
65 presence/absence data. Both partitioning provided similar results (**Fig S3.1**), and we used an
66 additive partitioning framework to quantify the β -diversity in our article.

67



68

69 **Fig. S3.1.** Results of leave-one-out cross-validation analysis used to assess predictive
 70 performance of β -kriging interpolation procedure using a multiplicative (red) and additive
 71 (red) partitioning of β -diversity for species richness computed on Mediterranean fish data,
 72 Gulf of Lions. The continuous blue line represents the first bisector.

73

74 **Rao's quadratic entropy partitioning**

75 Some methodological issues regarding the partitioning of Rao's quadratic entropy
 76 have been recorded recently: the index could lead to negative values for β -diversity, or the
 77 measured values for β are extremely low even for complete species replacement between
 78 communities. For the additive framework, according to (Ricotta 2005; De Bello *et al.* 2010)

79 mean local diversity Q_{α}^{-} could be computed as $Q_{\alpha}^{-} = \sum_{i=1}^N \omega_c Q_{\alpha c}$ where ω_c are the
 80 weights given to each location c (among N locations) and are the same as those used to
 81 compute total relative abundances P_i representing the weight given to each species, $P_i = \omega_c p_{ic}$
 82 (Villegger *et al.* 2012). In this case, β -diversity is finally written as follow, and it is always

83 positive: $\beta = \gamma - \sum_{c=1}^N \omega_c \alpha_c$ The total relative abundances P_i are used to compute γ -diversity.

84 Two approaches are possible: 1) the weights can be equal to $1/n$ and therefore all α
 85 measures will receive the same weight 2) the weights can be defined as the number of
 86 individuals per location divided the total number of individuals in a region (for more details
 87 see (De Bello *et al.* 2010)).

88

89 Secondly, to avoid β to be low regardless of the real turnover between locations
 90 within multiplicative and additive framework, Jost's correction is recommended, where Q_{α}^{-}

91 is written as $Q_{\alpha}^J = \frac{1}{1 - Q_{\alpha}^{-}}$ and $Q_{\gamma}^J = \frac{1}{1 - Q_{\gamma}}$. Briefly, mean local and total diversities

92 computed on each pair of locations are transformed to their equivalent number of species.

93 In case of additive frame, β is defined as $\beta_{add}^J = Q_{\gamma}^J - Q_{\alpha}^J$ and in the case of

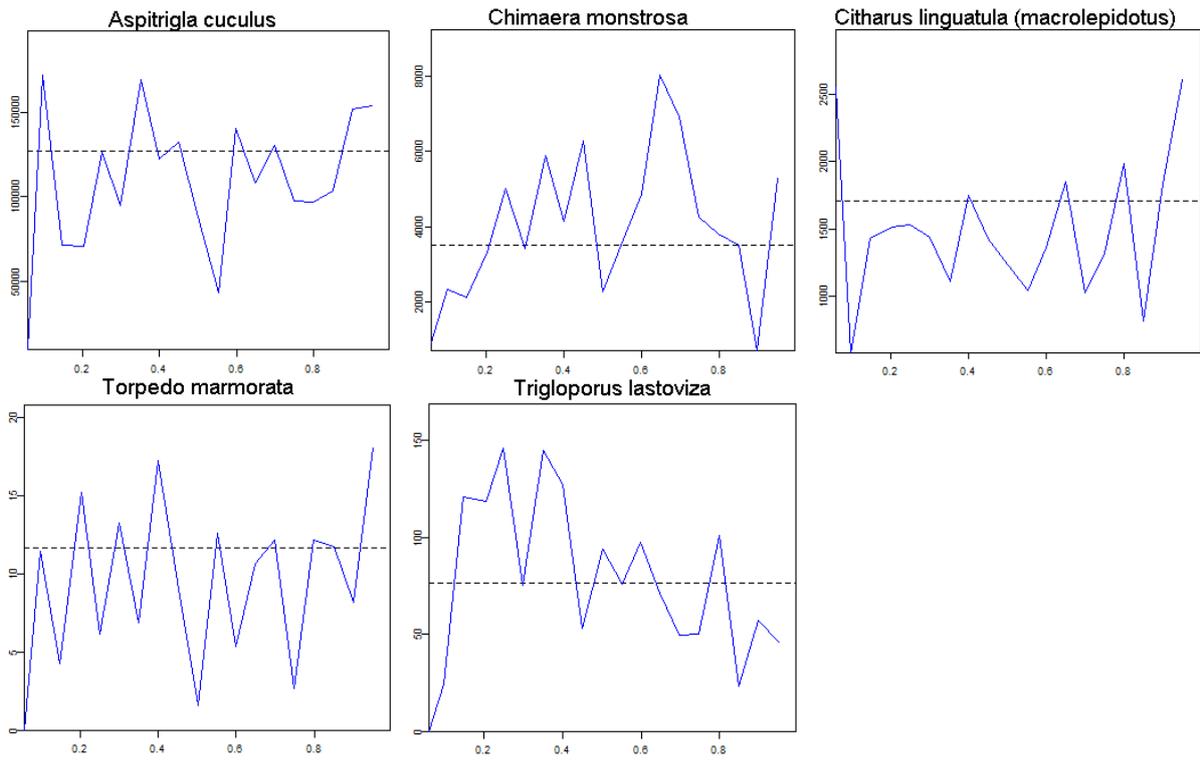
94 multiplicative framework $\beta_{mult}^J = \frac{Q_{\gamma}^J}{Q_{\alpha}^J}$. Note that in case of Rao's quadratic entropy, (De

95 Bello *et al.* 2010) showed that there is a direct link between additive and multiplicative β -

96 diversity: $\beta_{mult}^J = \frac{\beta_{add}^J}{Q_{\gamma}^J}$.

97

98 **Appendix S4. Individual species experimental variogram**

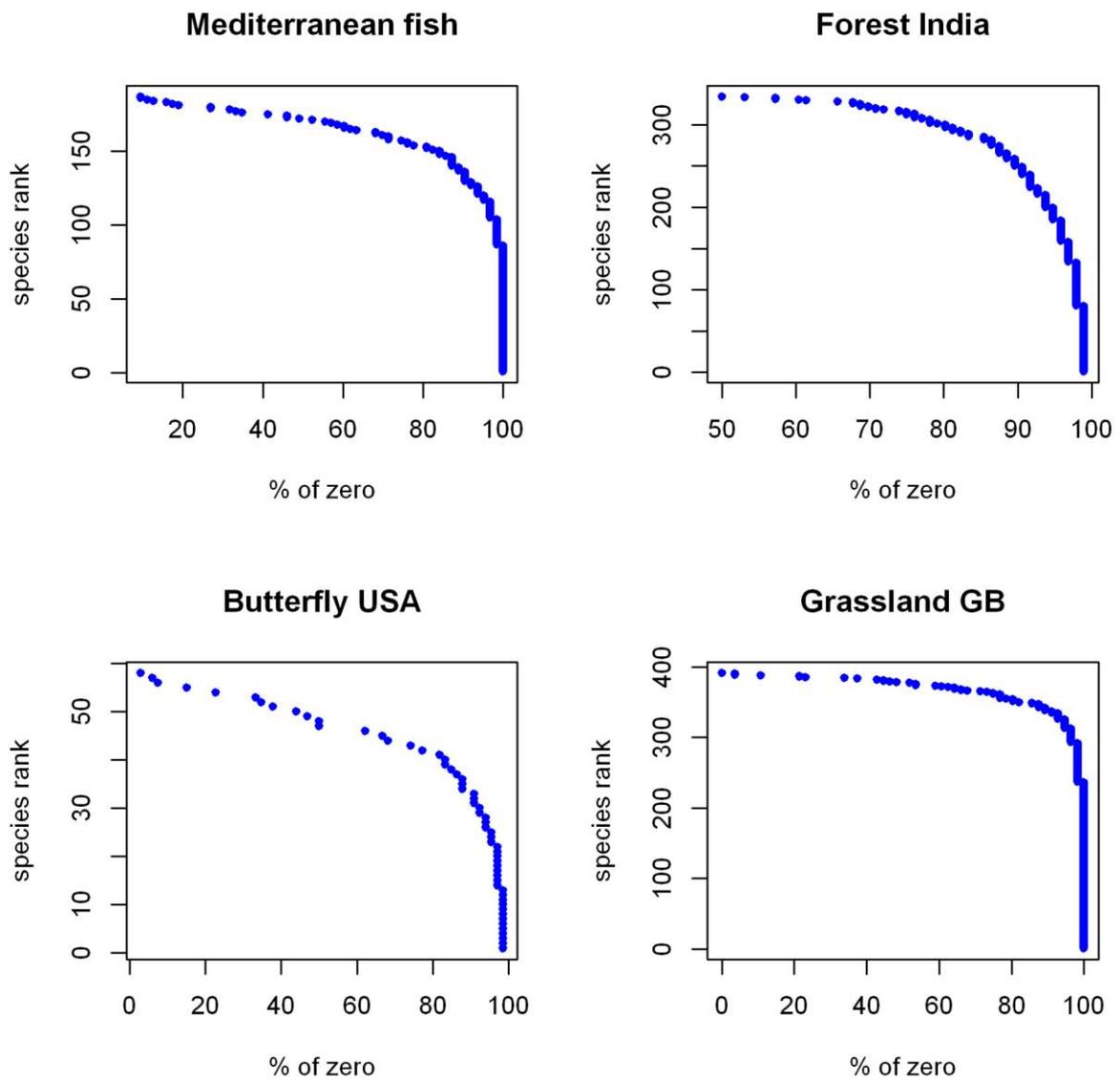


99

100 **Fig. S4.1.** Experimental variogram of 5 fish Mediterranean demersal fish (MEDITS data, Gulf
101 of Lions, France).

102

103



105
106 **Fig. S5.1.** Distribution of species' occurrence for the four datasets of different
107 faunistic/floristic groups considered.

108

109

110

111

112 **References**

- 113 Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., Freestone, A.L. et al. (2011)
114 Navigating the multiple meanings of B diversity: a roadmap for the practicing ecologist.
115 *Ecology Letters*, 14, 19–28.
- 116 de Bello, F., Lavergne, S., Meynard, C.N., Lepš, J. & Thuiller, W. (2010) The partitioning of
117 diversity: showing Theseus a way out of the labyrinth. *Journal of Vegetation Science*, 21,
118 992–1000.
- 119 Jost, L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology*,
120 88, 2427–2439.
- 121 Jost, L., DeVries, P., Walla, T., Greeney, H., Chao, A., Ricotta, C. (2010) Partitioning diversity
122 for conservation analyses. *Divers. Distrib.*, 16, 65–76.
- 123 Lande, R. (1996) Statistics and partitioning of species diversity, and similarity among multiple
124 communities. *Oikos*, 76, 5–13.
- 125 Loreau, M. (2000). Biodiversity and ecosystem functioning: recent theoretical advances.
126 *Oikos*, 91, 3–17.
- 127 Ricotta, C. (2005). Through the Jungle of Biological Diversity. *Acta Biotheoretica*, 53, 29–38.
- 128 Villéger, S., Ramos Miranda, J., Flores Hernandez, D. & Mouillot D. (2012). Low functional β -
129 diversity despite high taxonomic β -diversity among tropical estuarine fish communities. *PLoS*
130 *One*. 7, e40679
- 131 Whittaker, R.H. (1972) Evolution and measurement of species diversity. *Taxon*, 21, 213–251