Mapping diversity indices: not a trivial issue

Granger Victoria 1, *, Bez Nicolas 2, Fromentin Jean-Marc 3, Meynard Christine 4, Jadaud Angelique 3, Mérigot Bastien 1

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
Introduction

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

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

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

The direct approach, called "assemble first, and predict later" (Ferrier et al. 2002; Ferrier & Guisan 2006; Mokany et al. 2011), consists in computing directly a diversity index at sampled locations and then in interpolating those values at unsampled locations in each grid point of
the studied area. Although scientific literature provides a plethora of interpolation
techniques (e.g., James & McCulloch 1990), their use needs particular cautious when dealing
with diversity indices. Unlike other quantitative variables, diversity indices are not spatially
additive, i.e., the diversity of a studied area (e.g., region) is not the sum of the local
diversities. Note that, even though they are connected, the (spatial) additivity to which we
refer here is not the additive partitioning of regional γ-diversity into the mean local α-
diversities and β diversity as described by Lande (1996). Additivity of indices has been
discussed from a theoretical point of view (Keylock 2005; Hoffmann 2006), but considering
this property in a mapping context is lacking. For instance, let us consider the species
richness at two locations A and B being equal to 5 and 2, respectively, while 2 species are
shared between the two locations. If the species richness would be additive, its value for the
pooled area of locations A and B would be equal to 7 (Carrasco et al. 2008). However, since
these two locations have two species in common, the actual species richness is equal to 5.
This simplistic example shows that the species richness of an area that includes several
locations is different than the sum of the species richness in all locations if some locations
share similar species. This index would be additive only if all the locations have no species in
common (e.g., Keylock 2005; Hoffmann 2006), which is a very restrictive situation in natural
communities. This problem is thus related to the similarity in species composition between
locations, i.e., β-diversity (Magurran 2004; Anderson et al. 2011; Pavoine 2012).
Spatial additivity is particularly critical for interpolation techniques (and thus mapping), as
they rely on linear combinations of values of diversity indices (Michalakopoulos &
Panagiotou 1997; Rivoirard et al. 2000). When applied on additive variables, like absolute
abundance, traditional spatial interpolation methods (such as kriging, distance weighting
etc.) are consistent with the fact that the index value of an area composed of several pooled
locations is equal to the mean value of the index in these locations. Thus, considering
arithmetic mean of interpolated diversity indices would be accurate only if the index is
spatially additive, regardless of the interpolation method being applied. To circumvent this
problem, we proposed, in the frame of the direct approach, to combine geostatistical
techniques and β-diversity concept to interpolate local α-diversity indices over a given area
(Couteron & Pelissier 2004). This goal is not to estimate the “total species richness of an
area” (γ-diversity, e.g., Ugland et al. 2003).

Note that the lack of spatial additivity does not only affect the number of species, but also
the relative abundance (proportion) that are used in other facets of species diversity.
Appendix A summarizes results of a simple test of additivity conducted on other
complementary widely used diversity indices. None of them strictly respect this property.
Therefore, we applied the direct and the indirect approaches using two common diversity
indices (the species richness and the Rao’s quadratic entropy) and four datasets of different
faunistic and floristic groups collected from scientific surveys. We finally discuss lines of
research that may resolve the problems raised.

Material and methods

Data

We considered four different datasets.

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

121 by the fishing gear were considered (Gaertner et al. 2010, 2013). Abundance was
122 standardized to 1 km$^2$, 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 Diversity indices

138 Generally, more than one index is necessary to describe species diversity (Pavoine &
139 Bonsall 2011). Different indices indeed allow to quantify different facets, mainly species
140 number, evenness, or more complex variations considering taxonomic, phylogenetic and/or
141 functional differences between species (Devictor et al. 2010; Meynard et al. 2011; Pavoine
142 2012; Stuart-Smith et al. 2013). Here, we considered two diversity indices widely used in
143 ecology of communities and in diversity mapping studies (e.g. Devictor et al. 2010; Stuart-
Smith et al. 2013), knowing that the spatial additivity issue is true for other indices as well as those dealing with other biodiversity levels, such as genetic diversity (see end of the Introduction section and Appendix A). First, we computed species richness, the most intuitive and popular index in both marine and terrestrial diversity studies. This index was applied on all four above datasets.

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

\[ Q = \sum_{i=1}^{S} \sum_{j=1}^{S} p_i p_j d_{ij} \]

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

Statistical analysis

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

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

(1)

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

However, the variogram, i.e., arithmetic spatial variance of index value between locations ($\alpha$-diversity), does not quantify ecologically diversity variations (see example described in the introduction). Thus, replacing it by a $\beta$-diversity (i.e., an adequate measure of species replacement among locations) should ensure a more accurate quantification of diversity variation among locations. We thus propose an alternative methodological framework for interpolating diversity indices, called $\beta$-kriging. It consists in replacing the weighting function usually expressed as the spatial variance above (i.e., theoretical
variogram) by a spatial β-diversity model fitting the empirical β-diversity model previously proposed (Couteron & Pelissier 2004). We call it β-gram, which is defined as:

$$\beta(h) = \frac{1}{N(h)} \sum_{i \neq j} \beta(s_i - s_j)$$ (2)

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

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

The indirect approach:

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

The performance of each interpolation technique, in terms of the accuracy in estimating diversity index value, was assessed by comparing the deviations of estimates from the observed data through the use of the leave-one-out cross-validation (Stone 1974). In such procedure, a given sampled location is deleted from the data set and is estimated by performing the method, using the remaining locations. The operation is then repeated for all sampled locations. The estimated values are finally compared to the observed field values by mean of scatter plots, deviations from the first bisector (i.e., y=x, the case where observed and predicted values are equal), slopes of the linear regression and coefficients of determination R².

Results

The direct approach

Patterns between β-grams and variograms computed for the direct approach based on species richness on the four datasets were different (Fig. 1). Species replacement (i.e., β-diversity) was relatively high at even very short distances (strong nugget effects in the β-grams), while species richness was less contrasted at the same scale (see variograms in Fig.
The results of the leave-one-out cross-validation procedure are presented in Fig. 2. For all datasets, regression slopes between observed and estimated values ranged between 0.89 and 1.05 for kriging, between 1.13 and 1.65 for β-kriging according to the dataset considered. R² values remained rather low (0.22 < R² < 0.41) for both procedures. The scatter plots of observed values versus predicted values were highly dispersed around the first bisector, showing that both classical kriging and β-kriging had poor prediction performances. The range of estimated values by β-kriging was different, and generally more restricted, than by classical kriging. For instance about the Forest India dataset, while observed values ranged between 1 and 59 species, the estimated values by kriging ranged between 17.32 and 43.88 species and between 7.75 and 34.41 species by β-kriging. The differences in estimated values between classical kriging and β-kriging directly came from the differences between the theoretical β-gram and variogram (red dotted lines in Fig. 1).

For Rao’s quadratic entropy, the direct approach was applied only to the Mediterranean fish data, due to availability of species taxonomic differences data (see materials and methods section). The variogram and β-gram were also different (see Fig B.1). Both interpolation methods provided again poor prediction performances (Fig 3.a). Regression lines for both kriging and β-kriging procedures presented a slope inferior to 1 (0.77 for kriging and 0.6 for β-kriging) and both intercepts for both regressions were equal to 0.13 and 0.24 respectively, again far from the first bisector. Furthermore, R² values were very low, i.e., equal to 0.15 and 0.08 for kriging and β-kriging scatterplots, respectively. The estimated values ranged between 0.36 and 0.72 for β-kriging and between 0.39 and 0.71 for classical kriging, while the observed values were much wider, i.e., between 0.03 and 0.75 (Fig. 3.a).
The indirect approach

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

Discussion

In this study we emphasized that interpolating and mapping diversity indices (i.e., estimating values at all locations to map the studied area from some sampled locations) is problematic, and we illustrated this on several datasets collected from scientific surveys. First, we have seen that the traditional direct approach cannot provide accurate mapping because of the lack of spatial additivity of diversity indices. We thus proposed an alternative procedure, called the β-kriging, by combining geostatistical tools and β-diversity concept to model the spatial variations in diversity index. However, even if β-kriging is more ecologically founded, it does not really improve the predictions of species richness or quadratic entropy indices made by classical kriging, using a variogram. Although β-kriging fails to predict accurately diversity index, β-gram can be considered as an interesting tool to study diversity variations between spatially distant locations of a given area (Couteron & Pelissier 2004, Pavoine 2005, Shen et al. 2013, Parmentier et al. 2014,
Pélissier and Goreaud accepted). Notably β-gram can be implemented to study the spatial structure of functional or phylogenetic diversity in the framework of the spatial point processes (Shen et al. 2013), as proposed by Pélissier and Goreaud (accepted). For instance, the null hypothesis of species equivalence (i.e., absence of spatial structure in species relatedness) can be tested by using a Monte Carlo randomization procedure shuffling the between-species distances (i.e., permuting simultaneously the rows and columns in the dij matrix). Then the observed β-gram (i.e., diversity index computed on each pairwise sampled locations in function of spatial distances between these locations) is compared to the confidence envelopes generated by the Monte Carlo randomization to determine if the null hypothesis can be, or not, accepted (see for more details Shen et al. 2013; Pélissier and Goreaud accepted).

Second, regarding the indirect approach, most species of a given assemblage and/or community are known to present low to very low levels of abundance and/or occurrence (Gaston 1994; Martin et al. 2005). Modelling the spatial structure (e.g., the variogram) and the spatial distributions (for instance through kriging) of those rare species could hardly be performed with traditional statistical tools (see examples of experimental variograms for several species in Appendix D). For instance, for the Medits dataset that include 186 fish species, the probability of presence for each species shows that the vast majority of species are rare or extremely rare (65% of the species distributions get more than 95% of 0), or present high punctual abundance (see Appendix E). In this case, kriging based on species spatial autocorrelation is no longer operational for spatial interpolation for most species, as already stressed by Morfin et al. (2012). Note that the issue of zero-inflated data is actually a
common feature in ecological study and it is restricted to marine assemblages (Martin et al. 2005). The use of the indirect approach can further create a bias in predicted index values relative to the observed ones (see for instance the application on quadratic entropy). It can be attributed to the fact that the indirect approach smoothes the presence or abundance of the species and their distribution range. In other words, it creates presence in locations where species were not observed. Furthermore, this smoothing can hardly capture some discontinuities in the spatial distribution (e.g., highly fragmented and/or disturbed area). In such situation, a k-nearest neighbors algorithm’s method could be applied (Altman, 1992), knowing that the capacity of the method to deal with discontinuities decreases with the increasing number of neighbors considered. Consequently, the indirect approach could only be applied on the most abundant (common) species in communities, which seriously restraints the objectives of any diversity study by shedding the light on a few species, and that may not be the ones of conservation concern.

Perspectives

Following the above statements, we suggest two directions of possible improvements. First, the bias identified in the indirect approach comes from interpolation method and more certainly from the fact that diversity indices are non linear with regards to the individual layers. For instance, in the case of the Rao's quadratic entropy index, there is a quadratic link between species proportion and the index. A way of avoiding bias, is to simulate each species distribution conditionally on the observed data (Chilès and Delfiner,
and to use these simulations rather than the interpolations. In the same way that the mean of log-transformed data is not the log-transformed mean, the diversity index will be estimated by the mean of the transformed simulations and not by the transformed mean. It is worth remaining here that the aim of a conditional simulation is to create a distribution for each species that mimic the true spatial heterogeneity of the variable. This contrasts with interpolation (e.g., kriging) which estimates the expected species distributions (i.e., a smoothed version of the study variable). Conditional simulations preserve the variance of the observed data without smoothing and represent different equally possible spatial distribution of the studied variable. It would be a viable alternative when the spatial structure of each species is known. However, this method is also challenged by zero-inflated data to map rare species in the same ways as kriging.

Second, an alternative strategy to map diversity indices is to use models including abiotic and/or biotic explanatory variables (e.g., generalized linear or additive models GLM/GAM, machine learning methods, co-kriging methods, Olden et al. 2008, Ballesteros-Mejia et al. 2013, Hernández-Stefanoni et al. 2011). It is acknowledged that three main drivers act on species distributions and diversity at different spatial scales, i.e., (i) abiotic constraints, (ii) dispersal and (iii) biotic interactions (e.g., predation, competition and facilitation, see Loreau & Mouquet 1999; Soberon 2007). Ignoring in models a combination of these explicative variables may lead to a certain part of unexplained variability (Boulangeat et al. 2012; Cavieres et al. 2014). However, some of these variable values are not always known for every species in natural communities (e.g., biotic interactions or dispersal limitations). When biotic information are not available, it is usual to only deal with abiotic predictors. For instance, Leathwick et al. (2006) mapped species richness of demersal
fish considering only environmental variables in GAMs and Boosted Regression Trees (BRTs), for which the explained deviances varied between 45% and 60%. Bhattarai and Vetaas (2003) applied GLMs to study variation in species richness of different groups of herbaceous in function of environmental variables for which explained deviance of models highly varied according to the group (between 14% and 62%).

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

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

In conclusion, we showed that mapping index by interpolation methods used in the frame of direct or indirect approach may not be accurate because diversity indices are not spatially additive and many species in natural communities are rare. The use of the indirect
approach comes with the large burden of having to ignore or at least downplay the rarest species for which individual species distribution model is hardly feasible. Unfortunately, it differs from the crucial aim to consider all species of communities, and these rare species are usually of particular interest, notably from a conservation perspective, but also for ecosystem functioning as recently demonstrated (Mouillot et al. 2013). In the frame of the direct approach, the β-gram can be an interesting tool to study diversity variations between spatially distant locations of a given area, but the β-kriging procedure failed to predict accurately diversity index, as other traditional interpolation methods. Thus, considerable progress has still to be made and we highlight that conditional simulations and models taking into account biotic and abiotic explanatory variables could provide a solution for an accurate diversity indices mapping.

Aknowledgements

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

Data accessibility

Demersal fish: contact Angelique.jadaud@ifremer.fr

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

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


Fig. 1. Spatial structure of species richness measured by variogram and β-gram, for each dataset. Y-axis: green continuous curves represent the empirical variogram and the empirical β-diversity model computed from the additive partitioning for each pair of locations. The red dotted lines represent the theoretical continuous model (spherical or linear on the left and right panels, respectively) fitted to the empirical variogram or β-gram. x-axis: distance between locations in degree.
Fig. 2. Results of leave-one-out cross validation procedure for species richness. Procedure used to assess predictive performance of the direct approach by classical kriging (in blue) and additive β-kriging (in red) for species richness. Species richness computed on four datasets of different faunistic/floristic groups. The gray line represents the first bisector (i.e., y=x), the case where observed and predicted index values are equal.
Fig. 3. Results of leave-one-out cross validation procedure for Rao’s quadratic entropy. Procedure used to assess predictive performance of the direct and the indirect approaches for Rao’s quadratic entropy. Rao’s quadratic entropy computed only on Mediterranean demersal fish data due to availability in species taxonomic differences.

a) the comparison between classical kriging (in blue) and additive β-kriging (in red) procedure on Mediterranean fish species b) the comparison between the direct approach by classical kriging (in blue) and the indirect approach (purple) by inverse distance weighting.

The gray line represents the first bisector (i.e., y=x), the case where observed and predicted index values are equal.

Supplementary material online

**Appendix S1. Additivity test.**

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

<table>
<thead>
<tr>
<th>Diveristy component</th>
<th>Indices</th>
<th>Formula</th>
<th>St A</th>
<th>St B</th>
<th>$StA + StB$</th>
<th>$I(StA+StB)/2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species number</td>
<td>$S$</td>
<td>Number of species per location</td>
<td>5</td>
<td>2</td>
<td>3.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Evenness and species number</td>
<td>$H'$</td>
<td>$-\sum_{i=1}^{S} \ln(p_i) p_i$</td>
<td>1.25</td>
<td>0.69</td>
<td>0.96</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>$E''$</td>
<td>$-\sum_{i=1}^{S} \ln(p_i) p_i$</td>
<td>3.5</td>
<td>1.95</td>
<td>2.73</td>
<td>1.93</td>
</tr>
<tr>
<td></td>
<td>Evenness</td>
<td>Divergence</td>
<td>Berger-Parker</td>
<td>Q (quadratic entropy)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>----------</td>
<td>------------</td>
<td>---------------</td>
<td>-----------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-D</td>
<td>$1 - \sum_{i=1}^{S} p_i^2$</td>
<td>0.62 0.47 0.55 0.36</td>
<td>2.65 1.91 2.28 1.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simpson’s PIE</td>
<td>$\frac{N}{N-1} (1 - D)$</td>
<td>0.67 0.50 0.59 0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/D</td>
<td>$\frac{1}{\sum_{i=1}^{S} p_i^2}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heip’s</td>
<td>$\frac{e^H - 1}{S - 1}$</td>
<td>0.63 0.95 0.79 0.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simpson’s evenness</td>
<td>$\frac{1 - D}{(1 - \frac{1}{S})}$</td>
<td>0.78 0.95 0.84 0.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Berger-Parker</td>
<td>$\frac{N_{max}}{N}$</td>
<td>0.57 0.61 0.59 0.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference between</td>
<td>$\frac{\sum_{i=1}^{S} \sum_{j=1}^{S} p_i p_j d_{ij}}{S \sum_{i=1}^{S} p_i^2}$</td>
<td>0.42 0.29 0.35 0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>species</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix S2. β-gram model.

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

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

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

Kriging results are known to be sensitive to the shape of the model but not to its level. Only the kriging estimation variance is affected by the level of the model. Whilst we were only interested in interpolation and not in estimation variance, we used standardised β-grams.
Fig S2.1. Empirical and theoretical variogram and empirical and theoretical β-gram computed for Rao’s quadratic entropy for Mediterranean fish dataset. Y-axis: green continuous curves represent the empirical β-diversity model computed from the additive partitioning for each pair of locations and the empirical variogram. The red dotted lines represent the continuous model (linear of spherical) fitted to the empirical β-gram or variogram. X-axis: distance between locations in degree.
Appendix S3. The partitioning of β-diversity.

Methodology for partitioning γ-diversity into α and β components has been long debated (Loreau 2000; Anderson et al. 2011) and remains a sensitive ecological issue for which no agreement has been achieved. In the multiplicative partitioning (Whittaker 1972), the regional diversity γ, can be written as: \( \gamma = \alpha^\star \beta \). In the additive one (Lande 1996), \( \gamma = \alpha + \beta \) (Lande 1996). The partitioning depends on the mathematical framework of the used indices, for instance some diversity indices do not support an additive partitioning because they do not fulfil the concavity property (Jost 2007; Jost et al. 2010). While both partitioning frameworks have a clear definition and interpretation for species richness, and they can be linked mathematically, no arguments are available to advise to choose one between the diversity partitioning frameworks. Using additive or multiplicative partitioning allows considering unified approach to compute β-diversity for all the main indices available to describe complementary diversity facets (e.g. species richness, Simpson’s and quadratic entropy) while using pair-wise dissimilarity indices (e.g. Jaccard) would be restricted to presence/absence data. Both partitioning provided similar results (Fig S3.1), and we used an additive partitioning framework to quantify the β-diversity in our article.
Fig. S3.1. Results of leave-one-out cross-validation analysis used to assess predictive performance of $\beta$-kriging interpolation procedure using a multiplicative (red) and additive (red) partitioning of $\beta$-diversity for species richness computed on Mediterranean fish data, Gulf of Lions. The continuous blue line represents the first bisector.

Rao’s quadratic entropy partitioning

Some methodological issues regarding the partitioning of Rao's quadratic entropy have been recorded recently: the index could lead to negative values for $\beta$-diversity, or the measured values for $\beta$ are extremely low even for complete species replacement between communities. For the additive framework, according to (Ricotta 2005; De Bello et al. 2010)
mean local diversity $Q_\alpha$ could be computed as $Q_\alpha = \sum_{c=1}^{N} \omega_c Q_{\alpha c}$ where $\omega_c$ are the weights given to each location $c$ (among $N$ locations) and are the same as those used to compute total relative abundances $P_i$, representing the weight given to each species, $P_i = \omega_c p_{ic}$ (Villeger et al. 2012). In this case, $\beta$-diversity is finally written as follow, and it is always positive: $\beta = \gamma - \sum_{c=1}^{N} \omega_c \alpha_c$. The total relative abundances $P_i$ are used to compute $\gamma$-diversity.

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

Secondly, to avoid $\beta$ to be low regardless of the real turnover between locations within multiplicative and additive framework, Jost’s correction is recommended, where $Q_\alpha$ is written as $Q_j^\alpha = \frac{1}{1-Q_\alpha}$ and $Q_j^\gamma = \frac{1}{1-Q_\gamma}$. Briefly, mean local and total diversities computed on each pair of locations are transformed to their equivalent number of species.

In case of additive frame, $\beta$ is defined as $\beta_{add}^j = Q_j^\gamma - Q_j^\alpha$ and in the case of multiplicative framework $\beta_{mult}^j = \frac{Q_j^\gamma}{Q_j^\alpha}$. Note that in case of Rao’s quadratic entropy, (De Bello et al. 2010) showed that there is a direct link between additive and multiplicative $\beta$-diversity: $\beta_{mult}^j = \frac{\beta_{add}^j}{Q_j^\gamma}$. 
Appendix S4. Individual species experimental variogram

Fig. S4.1. Experimental variogram of 5 fish Mediterranean demersal fish (MEDITS data, Gulf of Lions, France).
Appendix S5. Distribution of species’ occurrence for the four datasets.

**Fig. S5.1.** Distribution of species’ occurrence for the four datasets of different faunistic/floristic groups considered.
References


