Mapping diversity indices: not a trivial issue

V Granger

N Bez

JM Fromentin

C Meynard
Virginia Institute of Marine Science

A Jadaud

Methods in Ecology and Evolution

Methods in Ecology and Evolution 2015, 6, 688–696 doi: 10.1111/2041-210X.12357

FORUM

Mapping diversity indices: not a trivial issue

Victoria Granger[1]*, Nicolas Bez[2], Jean-Marc Fromentin[3], Christine Meynard[4], Angélique Jadaud[3] and Bastien Mérigot[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; and* [4]*Virginia Institute of Marine Science, College of William and Mary, P.O. Box 1346, Gloucester Point, VA 23062, USA*

Summary

1. 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.

2. 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.

3. 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.

4. 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.

Key-words: 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, Gomez & Porzecanski 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.* 2013).

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

*Correspondence author. E-mail: granger.vica@gmail.com

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, Watson & Pearce 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, that is 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, that is 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, such as absolute abundance, traditional spatial interpolation methods (such as kriging, distance weighting) 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 S1 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 data sets of different faunistic and floristic groups collected from scientific surveys. We finally discuss lines of research that may resolve the problems raised.

Materials and methods

DATA

We considered four different data sets.

- The first data set reports demersal fish abundance in the Gulf of Lions (France) located in north-western Mediterranean Sea (3°W to 5·2°E; 42·5–43·8°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). A total of 186 species properly sampled by the fishing gear were considered during this program (Gaertner *et al.* 2010, 2013). Abundance was standardized to 1 km^2, for each species caught (Morfin *et al.* 2012; Gaertner *et al.* 2013).
- The second one reports woody plant species abundance in the central Western Ghats region, Karnataka, India (74·25°–75·5° E; 15·25°–13·5° N) in a network of 96 sampling sites. These data provide abundance on 334 tree species collected in 96 sampling sites during 1996–1997 (merged for this study) (Ramesh *et al.* 2010).
- The third data set reports butterfly diversity and abundance in Boulder County Open Space, Colorado, USA (105·1°–105·3° W; 39·9°–40·1° N) collected over 66 sites in the years 1999 and 2000 (merged for this study). The data contain butterfly species diversity and individual species' abundance of 58 species from five butterfly families (Oliver, Prudic & Collinge 2006).
- The fourth data set consists of vascular plant and bryophyte species composition and plant and soil biogeochemical data in Great Britain (6·3° W to 1·25° E; 50·5°N to 60·2°N) collected over 56 acid grasslands in 2002. These data provide abundance on 391 vascular species plants (Stevens *et al.* 2011).

DIVERSITY INDICES

Generally, more than one index is necessary to describe species diversity (Pavoine & Bonsall 2011). Different indices indeed allow to quantify different facets, mainly species number, evenness, or more complex variations considering taxonomic, phylogenetic and/or functional differences between species (Devictor *et al.* 2010; Meynard *et al.* 2011; Pavoine 2012; Stuart-Smith *et al.* 2013). Here, we considered two diversity indices widely used in 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 S1). 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 data sets.

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 five 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 (data set 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 α-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), that is the one of minimum variance. It is based on the spatial structure of the α-diversity which is quantified by the empirical semivariogram (i.e. computed on sampled data, Matheron 1963; Wagner 2003):

$$V(h) = \frac{1}{2|N(h)|} \sum\nolimits_{N(h)} (\alpha_{S_i} - \alpha_{S_j})^2 \qquad \text{eqn 1}$$

where $N(h)$ is the number of pairs of locations separated by a distance h, α_{Si} and α_{Sj} are the values of the α-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, that is to estimate values between locations (Matheron 1963; Wagner 2003).

However, the variogram, that is arithmetic spatial variance of index value between locations (α-diversity), does not quantify ecologically diversity variations (see example described in the introduction). Thus, replacing it by a β-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 β-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:

$$\beta(h) = \frac{1}{|N(h)|} \sum\nolimits_{N(h)} \beta(S_i - S_j) \qquad \text{eqn 2}$$

Equation 2 can be viewed as an empirical variogram, but representing the average pairwise diversity variation between locations separated by a distance h, with $\beta(s_i\text{–}s_j)$ being the variation (β-diversity) between each pair of locations (Appendix S2 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). Two partitions are commonly considered to compute β-diversity: the additive (Lande 1996) and the multiplicative partitioning (Whittaker 1972) (Appendix S3). 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 S3 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 a weighted average of the density values from the neighbouring values, their contribution being weighted as an inverse function of the distance to the kernel. We applied inverse distance weighting which allowed modelling distribution of all species in Mediterranean fish data set without modelling their spatial autocorrelation, in contrast to kriging. We thus made the assumption of a unique weighting function for all species distributions (including the rare ones).

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 R^2.

Results

THE DIRECT APPROACH

Patterns between β-grams and variograms computed for the direct approach based on species richness on the four data sets 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. 1). The results of the leave-one-out cross-validation procedure are presented in Fig. 2. For all data sets, regression slopes between observed

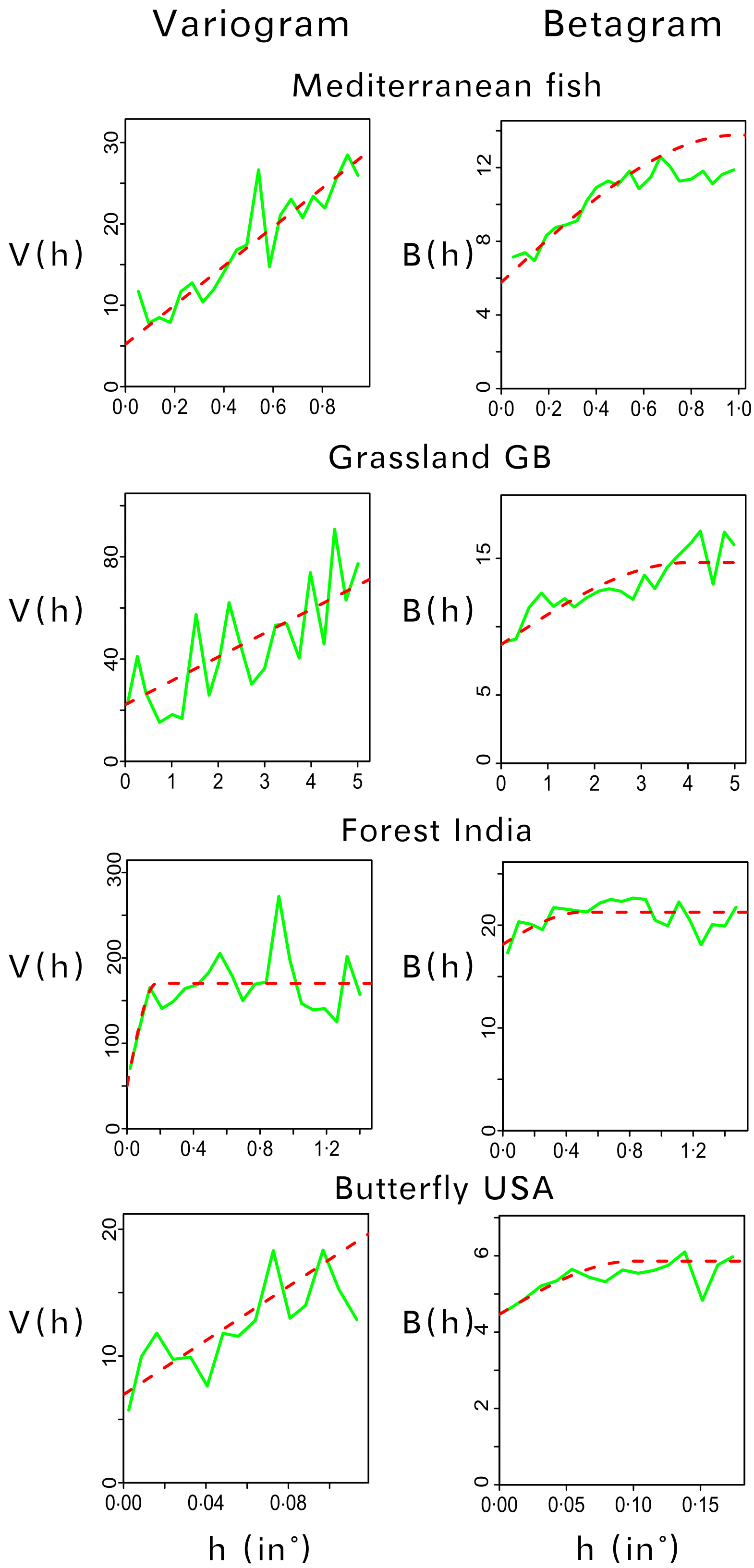

Fig. 1. Spatial structure of species richness measured by variogram and β-gram, for each data set. *Y*-axis: green continuous curves represent the empirical variogram and the empirical β-diversity model computed from the additive partitioning for each pair of locations. The red dotted lines represent the theoretical continuous model (spherical or linear) fitted to the empirical variogram or β-gram. *X*-axis: distance between locations in degree.

and estimated values ranged between 0·89 and 1·05 for kriging, between 1·13 and 1·65 for β-kriging according to the data set considered. R^2 values remained rather low ($0{\cdot}22 < R^2 < 0{\cdot}41$) 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 data set, 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. S2.1 in Appendix S2). Both interpolation methods provided again poor prediction performances (Fig. 3a). Regression lines for both kriging and β-kriging procedures presented a slope inferior to 1 (0·77 for kriging and 0·6 for β-kri-

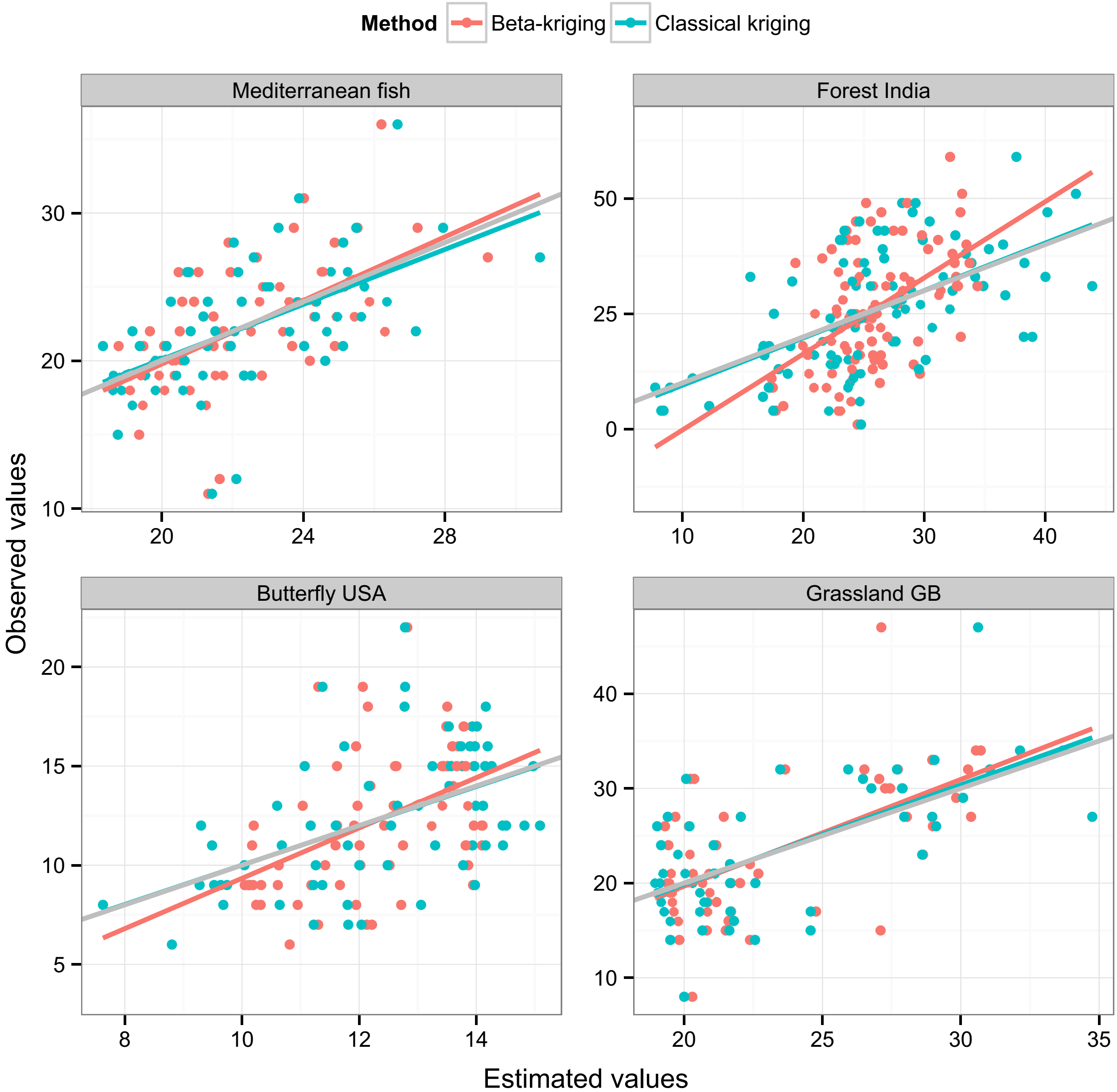

Fig. 2. Results of leave-one-out cross-validation procedure for species richness. Procedure used to assess predictive performance of the direct approach by classical kriging (in blue) and additive β-kriging (in red) for species richness. Species richness computed on four data sets of different faunistic/floristic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

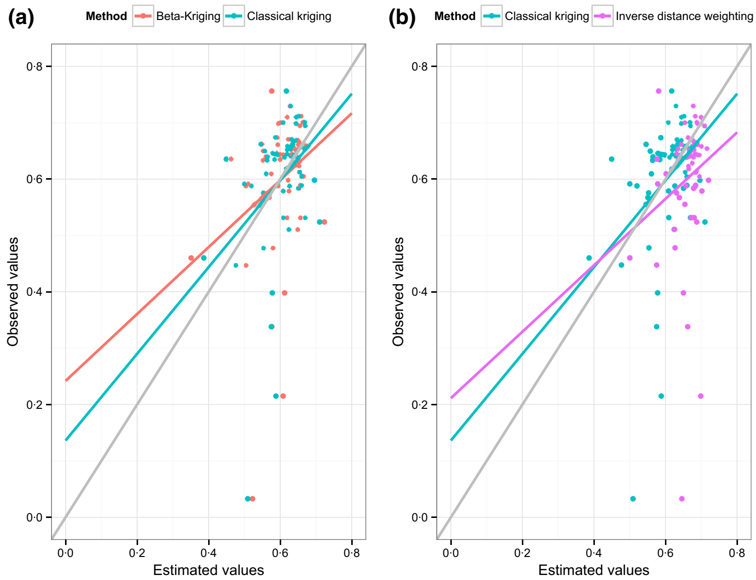

Fig. 3. Results of leave-one-out cross-validation procedure for Rao's quadratic entropy. Procedure used to assess predictive performance of the direct and the indirect approaches for Rao's quadratic entropy. Rao's quadratic entropy computed only on Mediterranean demersal fish data due to availability 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 grey line represents the first bisector (i.e. $y = x$), the case where observed and predicted index values are equal.

ging), and both intercepts for both regressions were equal to 0·13 and 0·24, respectively, again far from the first bisector. Furthermore, R^2 values were very low, that is 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, that is between 0·03 and 0·75 (Fig. 3a).

THE INDIRECT APPROACH

For Rao's quadratic entropy, the indirect approach was applied only to Mediterranean fish data set (see above). The results are presented in Fig. 3b. 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. 2b). 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 data sets 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 & Goreaud 2015). 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 & Goreaud (2015). 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 whether the null hypothesis can be, or not, accepted (see for more details Shen *et al.* 2013; Pélissier & Goreaud 2015).

Secondly, 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 S4). For instance, for the MEDITS data set 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 S5). 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 not 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 neighbours 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 neighbours 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 nonlinear with regard 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, 2012; Journel, 1974) 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.

Secondly, 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, that is (i) abiotic constraints, (ii) dispersal and (iii) biotic interactions (e.g. predation, competition and facilitation, see Loreau & Mouquet 1999; Soberón 2007). Ignoring in models a combination of these explicative variables may lead to a certain part of unexplained variability (Boulangeat, Gravel & Thuiller 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 is 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 & 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.

Acknowledgements

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. This paper is Contribution No. 3448 of the Virginia Institute of Marine Science, College of William & Mary.

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.

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

References

Altman, N. (1992) An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, **46**, 175–185.

Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., Freestone, A.L. *et al.* (2011) Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecology Letters*, **14**, 19–28.

Ballesteros-Mejia, L., Kitching, I.J., Jetz, W., Nagel, P. & Beck, J. (2013) Mapping the biodiversity of tropical insects: species richness and inventory completeness of African sphingid moths. *Global Ecology and Biogeography*, **22**, 586–595.

Bertrand, J., De Sola, L., Papaconstantinou, C., Relini, G. & Souplet, A. (2002) The general specifications of the MEDITS surveys. *Scientia Marina*, **66**, 9–17.

Bhattarai, K. & Vetaas, O. (2003) Variation in plant species richness of different life forms along a subtropical elevation gradient in the Himalayas, east Nepal. *Global Ecology and Biogeography*, **12**, 327–340.

Boulangeat, I., Gravel, D. & Thuiller, W. (2012) Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. *Ecology Letters*, **15**, 584–593.

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P. *et al.* (2012) Biodiversity loss and its impact on humanity. *Nature*, **486**, 59–67.

Carrasco, P., Chiles, J.-P. & Seguret, S. (2008) Additivity, Metallurgical Recovery, and Grade. 8th International Geostatistics Congress, 1–5 December, Santiago, Chile.

Cavieres, L.A., Brooker, R.W., Butterfield, B.J., Cook, B.J., Kikvidze, Z., Lortie, C.J. *et al.* (2014) Facilitative plant interactions and climate simultaneously drive alpine plant diversity. *Ecology Letters*, **17**, 193–202.

Chilès, J.P. & Delfiner, P. (2012) *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York, NY.

Couteron, P. & Pelissier, R. (2004) Additive apportioning of species diversity: towards more sophisticated models and analyses. *Oikos*, **107**, 215–221.

Dawson, T.P., Jackson, S.T., House, J.I., Prentice, I.C. & Mace, G.M. (2011) Beyond predictions: biodiversity conservation in a changing climate. *Science*, **332**, 53–58.

Devictor, V., Mouillot, D., Meynard, C., Jiguet, F., Thuiller, W. & Mouquet, N. (2010) Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. *Ecology Letters*, **13**, 1030–1040.

Ferrier, S. & Guisan, A. (2006) Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, **43**, 393–404.

Ferrier, S., Watson, G. & Pearce, J. (2002) Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodiversity and Conservation*, **11**, 2275–2307.

Gaertner, J.-C., Mérigot, B., Rélini, G., Bertrand, J.A., Mazouni, N., Politou, C.-Y. *et al.* (2010) Reproducibility of the multi-component aspect of species diversity across different areas and scales: towards the constitution of a shortlist of complementary indices for monitoring fish diversity? *Ecography*, **33**, 1123–1135.

Gaertner, J.-C., Maiorano, P., Mérigot, B., Colloca, F., Politou, C.-Y., Gil De Sola, L. *et al.* (2013) Large-scale diversity of slope fishes: pattern inconsistency between multiple diversity indices. *PLoS ONE*, **8**, e66753.

Gaston, K.J. (1994) *Rarity*. Chapman et Hall, London.

Heilbron, D. (1994) Zero-altered and other regression models for count data with added zeros. *Biometrical Journal*, **36**, 531–547.

Hernández-Stefanoni, J.L., Alberto Gallardo-Cruz, J., Meave, J.A., Dupuy, J.M. & Gallardo-Cruz, J.A. (2011) Combining geostatistical models and remotely sensed data to improve tropical tree richness mapping. *Ecological Indicators*, **11**, 1046–1056.

Hoffmann, S. (2006) *Concavity and additivity in diversity measurement: re-discovery of an unknown concept*. Univ., FEMM.

James, F.C. & McCulloch, C.E. (1990) Multivariate analysis in ecology and systematics: panacea or pandora's box? (C.E. McCulloch, Ed.). *Annual Review of Ecology and Systematics*, **21**, 129–166.

Journel, A.G. (1974) Geostatistics for conditional simulation of ore bodies. *Economic Geology*, **69**, 673–687.

Keylock, C. (2005) Simpson diversity and the Shannon Wiener index as special cases of a generalized entropy. *Oikos*, **109**, 203–207.

Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., Mcinerny, G.J. *et al.* (2012) Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*, **39**, 2163–2178.

Lande, R. (1996) Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos*, **76**, 5–13.

Lavergne, S., Mouquet, N., Thuiller, W. & Ronce, O. (2010) Biodiversity and climate change: integrating evolutionary and ecological responses of species and communities. *Annual Review of Ecology Evolution and Systematics*, **41**, 321–350.

Leathwick, J., Elith, J., Francis, M., Hastie, T. & Taylor, P. (2006) Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series*, **321**, 267–281.

Loreau, M. & Mouquet, N. (1999) Immigration and the maintenance of local species diversity. *The American Naturalist*, **154**, 427–440.

Magurran, A.E. (2004) *Measuring Biological Diversity*. Blackwell Science, Oxford.

Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J. & Possingham, H.P. (2005) Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters*, **8**, 1235–1246.

Matheron, G. (1963) Principles of geostatistics. *Economic Geology*, **58**, 1246–1266.

Merckx, B., Van Meirvenne, M., Steyaert, M., Vanreusel, A., Vincx, M. & Vanaverbeke, J. (2010) Mapping nematode diversity in the Southern Bight of the North Sea. *Marine Ecology Progress Series*, **406**, 135–145.

Meynard, C.N., Devictor, V., Mouillot, D., Thuiller, W., Jiguet, F. & Mouquet, N. (2011) Beyond taxonomic diversity patterns: how do α, β and γ components of bird functional and phylogenetic diversity respond to environmental gradients across France? *Global Ecology and Biogeography*, **20**, 893–903.

Michalakopoulos, T. & Panagiotou, G. (1997) Information Technologies in the Minerals Industry: Proceedings of the First International Conference on Information Technologies in the Minerals Industry Via the Internet

Mokany, K., Harwood, T.D., Overton, J.M., Barker, G.M. & Ferrier, S. (2011) Combining α and β diversity models to fill gaps in our knowledge of biodiversity. *Ecology Letters*, **14**, 1043–1051.

Morfin, M., Fromentin, J.-M., Jadaud, A. & Bez, N. (2012) Spatio-temporal patterns of key exploited marine species in the northwestern mediterranean sea. *PLoS ONE*, **7**, e37907.
Mouillot, D., Bellwood, D.R., Baraloto, C., Chave, J., Galzin, R., Harmelin-Vivien, M. *et al.* (2013) Rare species support vulnerable functions in high-diversity ecosystems. *PLoS Biology*, **11**, e1001569.
Nei, M. (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 3321–3323.
Nei, M. & Li, W.H. (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, **76**, 5269–5273.
Olden, J.D., Lawler, J.J. & Poff, N.L. (2008) Machine learning methods without tears: a primer for ecologists. *The Quarterly Review of Biology*, **83**, 171–193.
Oliver, J., Prudic, K. & Collinge, S. (2006) Boulder County Open Space butterfly diversity and abundance. *Ecology*, **87**, 85721.
Parmentier, I., Réjou-Méchain, M., Chave, J., Vleminckx, J., Thomas, D.W., Kenfack, D., Chuyong, G.B. & Hardy, O.J. (2014) Prevalence of phylogenetic clustering at multiple scales in an African rain forest tree community. *Journal of Ecology*, **102**, 1008–1016.
Pavoine, S. (2005) Méthodes statistiques pour la mesure de la biodiversité. Ph. D. thesis. Université Lyon 1, 391p.
Pavoine, S. (2012) Clarifying and developing analyses of biodiversity: towards a generalisation of current approaches. *Methods in Ecology and Evolution*, **3**, 509–518.
Pavoine, S. & Bonsall, M.B. (2011) Measuring biodiversity to explain community assembly: a unified approach. *Biological Reviews of the Cambridge Philosophical Society*, **2**, 93–111.
Pélissier, R. & Goreaud, F. (2015) Ads package for R: a fast unbiased implementation of the K-function family for studying spatial point patterns in irregular-shaped sampling windows. *Journal of Statistical Software*, **63**, 1–18.
Pellissier, L., Rohr, R.P., Ndiribe, C., Pradervand, J.N., Salamin, N., Guisan, A. & Wisz, M. (2013) Combining food web and species distribution models for improved community projections. *Ecology and Evolution*, **3**, 4572–4583.
Ramesh, B.R., Swaminath, M.H., Patil, S.V., Pélissier, R., Venugopal, P.D., Aravajy, S., Elouard, C. & Ramalingam, S. (2010) Forest stand structure and composition in 96 sites along environmental gradients in the central Western Ghats of India. *Ecology*, **91**, 3118.
Rao, C. (1982) Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology*, **43**, 24–43.
Rivoirard, J., Simmonds, J., Foote, K., Fernandes, P. & Bez, N. (2000) *Geostatistics for Estimating Fish Abundance*. Blackwell Science Ltd, Oxford.
Shen, G., Wiegand, T., Mi, X. & He, F. (2013) Quantifying spatial phylogenetic structures of fully stem-mapped plant communities. *Methods in Ecology and Evolution*, **4**, 1132–1141.
Soberón, J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, **10**, 1115–1123.
Sterling, E.J., Gomez, A. & Porzecanski, A.L. (2010) A systemic view of biodiversity and its conservation: processes, interrelationships, and human culture. *BioEssays News and Reviews in Molecular Cellular and Developmental Biology*, **32**, 1090–1098.
Stevens, C.J., Dupré, C., Dorland, E., Gaudnik, C., Gowing, D.J.G., Diekmann, M. *et al.* (2011) Grassland species composition and biogeochemistry in 153 sites along environmental gradients in Europe. *Ecology*, **92**, 1544.
Stone, M. (1974) Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, **36**, 111–147.
Stuart-Smith, R.D., Bates, A.E., Lefcheck, J.S., Duffy, J.E., Baker, S.C., Thomson, R.J. *et al.* (2013) Integrating abundance and functional traits reveals new global hotspots of fish diversity. *Nature*, **501**, 539–542.
Thuiller, W., Lavergne, S., Roquet, C., Boulangeat, I., Lafourcade, B. & Araujo, M.B. (2011) Consequences of climate change on the tree of life in Europe. *Nature*, **470**, 531–534.
Thuiller, W., Münkemüller, T., Lavergne, S., Mouillot, D., Mouquet, N., Schiffers, K. & Gravel, D. (2013) A road map for integrating eco-evolutionary processes into biodiversity models. *Ecology Letters*, **16**, 94–105.
Ugland, K.I., Gray, J.S. & Ellingsen, K.E. (2003) The species-accumulation curve and estimation of species richness. *Journal of Animal Ecology*, **72**, 888–897.
Wagner, H.H. (2003) Spatial covariance in plant communities: integrating ordination, geostatistics, and variance testing. *Ecology*, **84**, 1045–1057.
Whittaker, R.H. (1972) Evolution and measurement of species diversity. *Taxon*, **21**, 213–251.

Received 22 December 2014; accepted 23 February 2015
Handling Editor: Pedro Peres-Neto

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Additivity test.

Appendix S2. β-gram model.

Appendix S3. The partitioning of β-diversity.

Appendix S4. Individual species experimental variogram.

Appendix S5. Distribution of species' occurrence for the four datasets.