UC Merced
Frontiers of Biogeography

Title
Variance partitioning and spatial eigenvector analyses with large macroecological datasets

Permalink
https://escholarship.org/uc/item/90d9g35b

Journal
Frontiers of Biogeography, 12(4)

Authors
Paula-Souza, Laura B. de
Diniz Filho, Jose Alexandre F.

Publication Date
2020

DOI
10.21425/F5FBG47295

Supplemental Material
https://escholarship.org/uc/item/90d9g35b#supplemental

Peer reviewed

RESEARCH ARTICLE

Variance partitioning and spatial eigenvector analyses with large macroecological datasets

Laura Barreto de Paula-Souza[1] (iD) & José Alexandre F. Diniz-Filho[1,2] (iD)

[1] *Programa de Pós-Graduação em Ecologia e Evolução, ICB, Universidade Federal de Goiás. Campus II (Samambaia), Goiânia, GO, Brasil. 74690-900;* [2] *Departamento de Ecologia, ICB, Universidade Federal de Goiás. Campus II (Samambaia), Goiânia, GO, Brasil. 74690-900.*
Corresponding author: José Alexandre F. Diniz-Filho: diniz@ufg.br

Abstract

Macroecological data are usually structured in space, so taking into account spatial autocorrelation in regression and correlation analyses is essential for a better understanding of patterns and processes. Many methods are available to deal with spatial autocorrelation, but there are some difficulties when one is dealing with huge geographical extents and fine-scale data. So, we propose a relatively simple and fast computer-intensive approach to deal with Principal Coordinate of Neighbor Matrices (PNCM)/ Moran's Eigenvector Mapping (MEM) analyses for large datasets, using global richness pattern of sharks as a model. We performed a variance partitioning approach by regressing species richness against environmental variables and spatial eigenvectors derived from PCNM. Due to the large number of ocean grid cells (> 9000), we ran the analyses 1000 timesby randomly subsampling each time 50 to 4500 cells and compared the distribution of the variance partitioning components, as well as the slopes of the environmental variables. We also estimated Moran's I coefficients for regression residuals to check if spatial eigenvectors took into account spatial autocorrelation. Comparing statistics of analyses with different sample sizes, we note that although the environmental component increases linearly, other components (unique space and shared) of the most important variables stabilize with about 1000 cells, whereas all other smaller effects tend to stabilize between 2500 and 3000 cells. Besides that, PCNM eigenvectors were able to control spatial autocorrelation very well. We showed that shark richness patterns are strongly and positively correlated with temperature range, according to the well-known pattern of distribution for the taxon, and strong negatively correlated with oxygen supplies, which are higher in polar zones where ice acts as a barrier to sharks. Our approach clearly shows that it is possible to perform a robust evaluation of global patterns of diversity using eigenvector approaches based on a resampling strategy and allows effective computation of the variance partitioning even when dealing with large datasets.

Highlights

- We provide the script for a simple, new and computationally feasible approach to deal with spatial autocorrelation using PCNM/MEM analyses for large datasets.

- We found the smallest necessary sample size to represent the total database of shark richness, simplifying eigenvector selection for spatial analyses.

- We performed variance partitioning analyses using a large dataset of shark diversity across more than 9000 grid cells worldwide, using a common Notebook computer and within a feasible running time.

- By running 1000 replicates of the analyses for different sample sizes (50 to 4000 cells) and comparing the resulting distribution of estimated parameters, we showed that the patterns stabilize at around 1000 cells in the system analysed, corresponding to about 10% of the total dataset.

- Our new approach allows researchers to perform spatial analyses of large datasets in feasible time and with reduced computational efforts, achieving an accurate approximation of effects, testing, and searching for the stabilization of parameters of interest.

Introduction

For some twenty years or so, macroecologists and biogeographers have increasingly recognized that it is important to take into account spatial autocorrelation when inferring ecological and evolutionary processes based on statistical analyses of geographical patterns in species richness and other similar patterns, such as those generated under ecogeographical rules (Lennon 2000, Diniz-Filho et al. 2003). In principle, spatial autocorrelation affects the Type I errors of regression and correlation analyses (Legendre 1993, Lichstein et al. 2002). Still, many discussions have focused on how coefficients shift when applying spatial statistics to macroecological data due to collinearity, scale, and hierarchy problems (Lennon 2000, Diniz-Filho et al. 2003, Beale et al. 2007, Kuhn 2007, Dormann 2007, Hawkins et al. 2007). Moreover, as processes are structured in space, quantifying and describing spatial autocorrelation in macroecology and ecology can be viewed as a new general approach to understanding such processes and not necessarily just a problem to be removed (Legendre 1993). Many methods are now available to deal with spatial autocorrelation (i.e., Dormann et al. 2007, Bini et al. 2009), which can be implemented in specialized software such as SAM (Rangel et al. 2006, 2010) and, more recently, in many R-packages (e.g., *spdep*, Bivand and Wong 2018; *gstat*, Pebesma 2004; *nlme,* Pinheiro et al. 2020; *adespatial*, Dray et al. 2020).

Yet, with the development of new GIS facilities and the possibility to quickly generate both fine grain species and environmental data at large geographical extents, spatial analyses may become a challenge. It is notoriously difficult to deal with sample sizes (n) of thousands of spatial units, especially if it involves (as it frequently does) eigenanalyses of spatial connectivity or distance matrices (**W** matrices with n x n dimensions;see Haining 2002, Griffith 2003). Although it is possible to use parallel processing and intensive computing techniques to deal with some of these problems, this solution is not always available, and, in any case, it may still take a long time to run (which is a constraint if, for example, distinct **W** matrices should be tested; Kissling and Carl 2008, Bauman et al. 2018b). With some techniques such as Principal Coordinate of Neighbor Matrices (PNCMs;Borcard and Legendre 2002, Dray et al. 2006) and subsequently developed but related approaches such as Moran's Eigenvector Mapping (MEMs; Dray et al. 2012), large matrices impose another additional challenge related to eigenvector selection to be used in the modeling (see Diniz-Filho and Bini 2005, Griffith and Peres-Neto 2006, Dray et al. 2012, Bauman et al. 2018a,b). Both techniques involve reconstructing spatial patterns using distance matrices (Euclidian distances for PCNMs or other types of weighting matrices for MEMs) by creating spatial eigenvectors that are newly composed variables that describe the spatial relationship among localities or sampling units. Thus, if these new variables are incorporated into the modeling process, they will take into account any spatial autocorrelation in the analyses and will be

sufficient, if appropriately fitted, to correct for inflated Type I errors in the other parameters of the model (see Borcard and Legendre 2002, Dray et al. 2006).

Here, we propose and evaluate a relatively fast and straightforward computer-intensive approach to deal with PCNM/MEM analyses for large datasets. The idea is simply to randomly sample the matrix thousands of times and evaluate the statistical distribution of parameters to be estimated. Although spatially-structured resampling techniques have been used to deal with spatial autocorrelation by generating samples a given distance apart (eliminating the pseudo-replication;e.g., Hawkins et al. 2007, Oliveira et al. 2014), our goal here is a bit different. The problem to be solved is the difficulty of analyzing, using PCNM or MEM, a global data set with thousands of grid cells to estimate spatial and environmental components of richness patterns, and not to eliminate neighbor cells to avoid autocorrelation. As these methods work with eigenvectors from pairwise distance or connectivity matrices, the time and capacity for computation will increase more than exponentially with sample size (both because matrices will increase in size in a square scale and because of complexity in extracting eigenvalues iteratively). We tested our approach by analyzing global richness patterns of sharks and correlate them with environmental data.

Methods

Data

Primary data for our paper consist of range maps obtained from IUCN for a total of 469 species of sharks distributed worldwide (IUCN 2019), representing about 78% of the species of Neoselachii. Taxonomy was based on the GBIF Backbone Taxonomy[1]. The ranges were polygons, which were overlaid using a global grid with a total of 9212 cells with oceanic centroids (coastal cells with land centroids were removed from the dataset, so the proportion of land area in each cell depends on the shape of the coastal line), with approximately 2° of latitude/longitude, using the R-package *LetsR* (Vilela and Villalobos 2015).

Eight environmental variables were obtained from the Bio-Oracle (Tyberghein et al. 2011) platform in the form of mean values and one variable in the form of range for the period between 2000 and 2014 and were also overlaid on the global grid. The variables used were mean temperature (°C), temperature range (°C), chlorophyll (mg.m^{-3}), dissolved molecular oxygen (mol.m^{-3}), sea ice concentration (fraction), pH, carbon phytoplankton biomass (umol.m^{-3}), primary productivity (g.m^{-3}.day^{-1}), and salinity (PSS).

Statistical analyses

We analyzed the data using a variance partitioning approach (Borcard et al. 1992, Legendre and Legendre 2013) by regressing species richness against environmental variables and spatial eigenvectors derived from PCNM. However, due to the large number of grid cells (> 9000), we performed the analyses 1000 times, by random sampling each time a given number of cells.

We tested sample sizes of 50, 100, 250, 500, 750, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 3000, 3500, 4000, and 4500 cells in order to find the smallest necessary sample size to represent the shark richness patterns, and thus simplifying analyses and time spent on it. Then we compared the distribution of the variance partitioning components, as well as the effect sizes of the environmental variables. We computed the median effect size (i.e., t-value for each regression slope) and the proportion of simulations in which a significant ($P < 0.05$) regression slope for each sample size was found.

As richness is strongly skewed, we obtained the R^2 for variance partitioning with the *glm* of R-package, with a Poisson link function, but the approach proposed here can be used with any modeling technique (indeed, a more standard partition using an OLS based on log-transformed richness provide qualitatively similar results, although with a higher fraction for the environmental component). The particular type of model that better fits the data, in general, can be tested with a few preliminary tests (as we did with our dataset) or some diagnosis statistics (such as AIC-based comparisons between classes of models or statistics for residuals distributions) incorporated into the replications for "a posteriori" decisions.

Spatial eigenvectors derived from PCNMs were obtained using the *pcnm* function in the *vegan* R package (Oksanen et al. 2007) and selected, for computational simplicity, based on the correlation with the response variable (which is roughly equivalent to stepwise functions, as eigenvectors are independent;see Bauman et al. 2018a,b). In any case, we estimated Moran's *I* coefficients (see Legendre and Legendre 2013) for regression residuals to check if spatial eigenvectors were enough to account for spatial autocorrelation and thus allow correct evaluation of Type I error of the slopes.

For the record, we analyzed the data in a Notebook computer with Intel I-5 processor and 8GB RAM.

Results

The richness pattern of sharks (Fig. 1) tracks the well-known pattern for the taxon, with high richness concentrated in the coastal zones and middle latitudes of both hemispheres. Besides that, the Indo-Pacific islands and tropical Oceania exhibit high richness as well (Tittensor et al. 2010, Lucifora et al. 2011, Worm and Tittensor 2018).

We provide detailed results for sample sizes equal to 100, 500, 1000, 2500, and 4500 cells, but detailed Tables with all results can be found in the Supplementary Material (Appendices S1 and S2). The variance partitioning for increasing sample sizes (Table 1) shows that the environmental effect (*a*), which starts with a relatively high median value of about 0.22 when only 100 cells are sampled, goes up to a median of 0.06 when 4500 cells were sampled. This reduction in the unique environmental component is not correlated with an increase in both shared (*b*) and unique geographical components (*c*) that tend to stabilize with smaller sample sizes (Fig. 2A).

However, although the unique environmental component in variance partitioning tends to increase linearly with sample size, the effect size of the most important environmental predictors in the full model (i.e., incorporating the selected spatial eigenvectors from PCNMs) was more or less stable across the sample sizes, above about 500 cells. As sample sizes increase, the median *t*-value of some variables tends to fluctuate, and the frequency of slopes with $P < 0.05$ tends to slightly increase, stabilizing when $n = 2500$ but not reaching values much higher than 80% of significant coefficients across simulations even with larger samples sizes. For the two variables with more important effects (i.e., temperature range and dissolved O_2), on the other hand, even with sample sizes smaller than 1000, more than 95% of the coefficients are significant (Fig. 2B). So, we understand that an increase in the unique environmental component is most likely due to an accumulation of minor effects that do not necessarily help to explain global patterns.

Just as a general reference, if we consider the stabilization of the main drivers and of the unique spatial component and its overlap with environmental variables, it is possible to see that, on average, the full model explains ca. 70% of the variation in species richness (CI 95% limits equal to 67.1% and 76.5%;Fig. 3A) and that environmental variables alone explains about 10% of this variation (CI 95% limits equal to 6.5% and 14.8%;Fig. 3B), without spatial
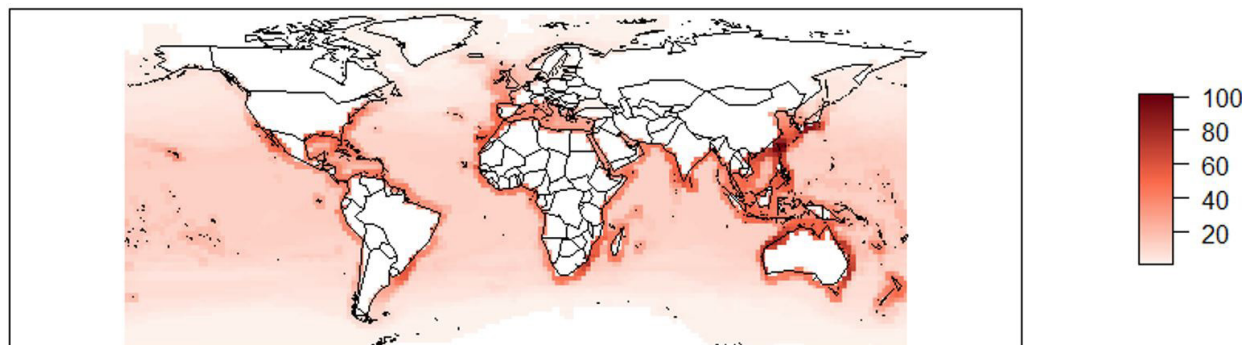

Fig 1. Richness patterns obtained by overlaying range maps of 469 shark species worldwide on a global grid with 9212 cells. Color scale indicates number of species.

Table 1. Median values of the components *a, b,* and *c* for variance partitioning, modeling shark richness with a GLM as a function of environmental variables and spatial eigenvectors, and their 95% lower and upper limits, based on 1000 samples with variable number of cells (n), ranging from 100 to 4500. The three components refer to environmental, shared, and spatial effects, respectively. The analyses were based on data for 469 species and a full dataset global grid comprising 9212 cells.

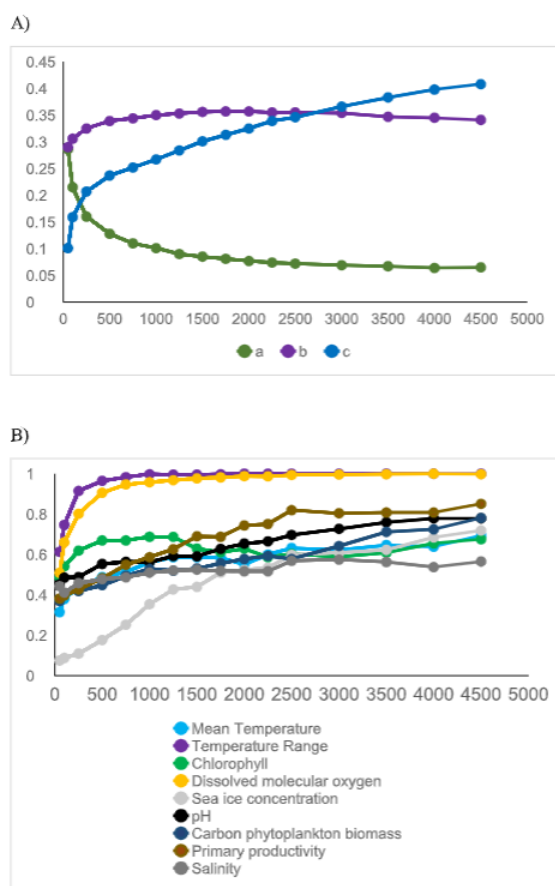| *n*	a	b	c
100	0.215	0.306	0.159
	(0.081 - 0.393)	(0.122 - 0.497)	(0.018 - 0.350)
500	0.128	0.339	0.237
	(0.074 - 0.188)	(0.260 - 0.423)	(0.157 - 0.320)
1000	0.101	0.35	0.267
	(0.066 - 0.148)	(0.288 - 0.413)	(0.206 - 0.330)
2500	0.072	0.355	0.346
	(0.048- 0.098)	(0.306- 0.405)	(0.293- 0.398)
4500	0.065	0.341	0.408
	(0.041- 0.089)	(0.295- 0.394)	(0.359- 0.453)

A)

B)

Fig 2. A) Tendency lines of variance partitioning components; median values for different samples sizes showing unique environmental (*a*), shared (*b*), and unique geographical (*c*) effects on shark richness worldwide. The analyses were based on data for 469 species and a full dataset global grid comprising 9212 cells. B) Tendency line of significance proportion of environmental variables on shark richness worldwide for different sample sizes, stabilizing between 2500 and 3000 cells. The analyses were based on data for 469 species and a full dataset global grid comprising 9212 cells.
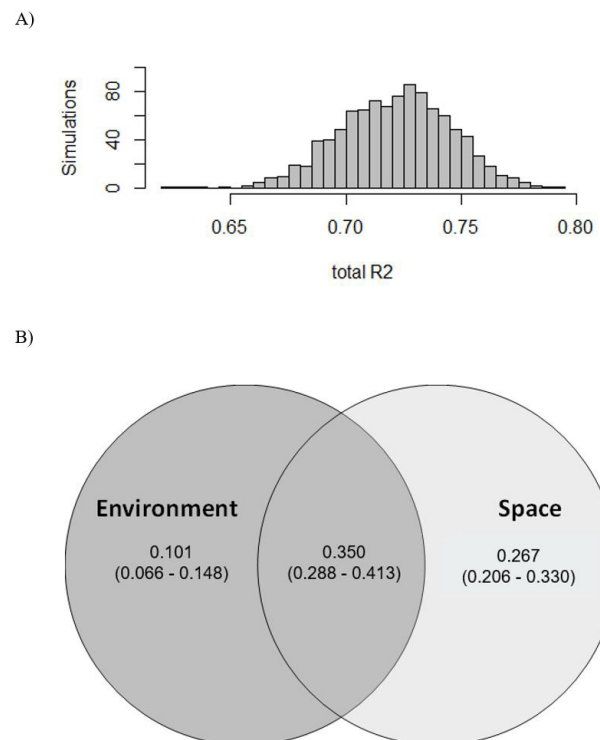
A)

B)

Fig 3. A) Distribution of total pseudo-R_2 of *glm* explaining shark species richness as a function of 9 environmental variables and spatial eigenvectors, for a total of 1000 random samples of 1000 grid cells, and, B) variance partitioning of the effects of the two sets of variables, with median values and 95% CI limits for the 1000 samples. The analyses were based on data for 469 species and a full dataset global grid comprising 9212 cells.

autocorrelation in model residuals. The effect sizes of GLMs reveal that temperature range followed by dissolved O_2 are clearly the most important drivers of shark richness (Table 2).

The number of spatial eigenvectors from PCNMs selected for analyses increased approximately

Table 2. Median of effect size (t-value) and frequency of significance for the GLM modeling shark richness as a function of nine environmental variables and spatial eigenvectors, for 5 different sample sizes (number of cells), based on 1000 replications. The analyses were based on data for 469 species and a full dataset global grid comprising 9212 cells.

N	100 t-value	freq (P<0.05)	500 t-value	freq (P<0.05)	1000 t-value	freq (P<0.05)	2500 t-value	freq (P<0.05)	4500 t-value	freq (P<0.05)
Mean Temperature	-0.266	0.38	-0.237	0.488	0.201	0.562	1.699669	0.632	2.468468	0.694
Temperature Range	3.523	0.747	7.037	0.965	8.914	0.998	12.46834	1	15.21962	1
Chlorophyll	1.924	0.541	2.875	0.67	2.966	0.688	0.780766	0.603	-1.911	0.676
Dissolved molecular oxygen	-2.691	0.66	-5.25	0.906	-6.689	0.958	-8.27014	0.994	-10.2487	0.998
Sea ice concentration	0.032	0.088	-0.617	0.176	-1.307	0.354	-2.37475	0.599	-3.02866	0.719
pH	-0.89	0.486	-1.324	0.553	-0.606	0.562	2.506806	0.697	3.467107	0.779
Carbon phytoplankton biomass	-0.599	0.407	-0.532	0.448	-0.455	0.529	1.541958	0.578	5.3355	0.781
Primary productivity	0.25	0.397	1.097	0.483	2.096	0.587	0.889	0.427	4.932928	0.851
Salinity	-0.371	0.411	0.682	0.478	0.842	0.51	0.339	0.458	1.701573	0.565

exponentially with increasing sample sizes, to account for more complex spatial patterns involved (Appendix S3). Even so, Moran's *I* values suggest that, in all cases, spatial eigenvectors were enough to take autocorrelation into account, with upper 95% limits of the confidence intervals never reaching values higher than 0.1 (and this upper limit also decreases with increasing sample size). The analyses took from a few seconds (when n = 50) to about three days (n = 4500) to run (Appendix S3) using the script available in the supplementary material (Appendix S4).

Discussion

Our approach clearly shows that it is possible to perform a consistent evaluation of global patterns of diversity using eigenvector approaches based on a resampling strategy. All statistics tend to shift with increasing sample sizes (i.e., number of cells), but in our example, the main effects associated with temperature range and dissolved O2 tend to stabilize between 500 and 1000 cells. Our simple approach allows a straightforward computation of the variance partitioning and would be easily repeated several times, if necessary (i.e., testing different **W**-matrices or performing richness deconstruction). The analyses with 2500 cells took approximately twelve hours in a notebook with Intel I-5 processor, which is quite feasible (although the time for processing all 9,000 cells would increase more than exponentially and would in fact not be possible on this same machine). Although using more powerful computers would result in faster computation and allow us to deal with large sample sizes (and even with the full dataset), showing that sampling the full dataset would provide similar results is still interesting. It can be very useful to researchers with limited computational resources at their disposal in many places across the world. In addition, it allows more complex analyses and many more tests, including the complex process of eigenvector selection for multiple **W**-matrices, in feasible computer times.

In practice, it is important to highlight in advance that we are discussing the stability of PCNM and variance partitioning for a particular dataset regarding patterns of shark diversity worldwide. We are not stating, for example, that results of a global analysis with about 9,000 cells will stabilize when sampling 5% of the dataset. This will actually depend on the relative magnitude of the drivers. Thus, when applying the approach proposed here to a new dataset, it is important to perform some iterative and quick tests searching for the stabilization of the results, which will also depend on the computer capabilities available to each researcher. For instance, a nice alternative would be to check for mean effect sizes using a small number of replications for increasingly samples sizes and when a stable effect size is achieved, increase the replications to better evaluate the frequency of significant results.

Also, contrary to other studies that undertook spatially-structured sampling (i.e., considering only cells that are situated a given geographic distance apart) to heuristically understand autocorrelation

effects on the analyses (i.e., Hawkins et al. 2007) or take them into account (in species distribution modeling; Oliveira et al. 2014), our goal is to reduce computational load. As in any variance partitioning approach, we also want to evaluate the magnitude of spatial effects on the response variable, both shared or independent of environmental variables.

In the context of the framework proposed here,, it is possible to assess the uncertainty of environmental variables due to spatial structure in data, so for example, with 500 cells, we can see that the 95% confidence interval ranges from 7.4% to 18.8%, clearly different from zero. Moreover, the effect sizes for individual predictors were also relatively stable, and the ones with the highest effects in our model are the temperature range and dissolved O2 (Table 2). The median of the effect size (i.e., t-value) and the proportion of significant slopes seem to be a suitable way to synthesize the results across the replications. We understand that the median of the parameters would estimate these effect sizes adequately.

Of course, other strategies to reduce sample size and allow a more straightforward computation of spatial and environmental effects would be possible, such as separating data for biogeographical region or any other characteristic (i.e., coastal versus deep-sea cells). Also, another simple idea is to increase cell size so analysis of the overall dataset is viable, but this involves a more complex discussion (so we're assuming that cell size was adequately defined "a priori" given data type and resolution both in diversity and environmental predictors). Anyway, assuming that processes are similar across regions, a pooled estimate of factors would allow a similar interpretation in terms of global diversity patterns. Even so, in this case, it would be necessary to decide first if a global set of eigenvectors would be extracted from a very large global **W**-matrix (a potential problem, especially if different **W** matrices should be considered) and then incorporated into "regional" variance partitioning. It would be possible to extract the eigenvectors for each region, but in this case, it may be difficult in principle to directly compare the spatial component of these partitions (although if autocorrelation is correctly taken into account by PCNM, as evaluated by residuals Moran's I, coefficients of the explanatory variables would be comparable). This seems to be in some way analogous to the problem of selecting eigenvectors in PCNMs and MEMs independently or not of the environmental variables (i.e., Safi and Pettorelli 2010).

We recognize that if patterns in different regions are driven by different processes, regional analyses may not be entirely informative about global spatial patterns. In this case, rather than doing separate PCNMs for each region it would be more interesting to use Geographically Weighted Regression (GWR; Fotheringham et al. 2002) or the new related approach of Geographically Weighted Path Analysis (GWpath; Barreto et al. 2019) to model local effects and evaluate non-stationarity patterns in data (and GWR is computationally much simpler than PCNM for handling large datasets). GWR and related approaches

Frontiers of Biogeography 2020, I2.4, e47295   6

solve the problem of regionalization by exploring the effects of environmental variation in a continuous way, mapping the coefficients of regressions or path analysis for all cells in the grid, but for each one there is a vector of weights based on the spatial distances from the focal cell to all other cells. Although these analyses are undoubtedly interesting and reveal more complex patterns, the goals of such approaches would be different, focused on local or regional effects, and not in a global analysis of richness patterns. Also, such analyses would not allow the evaluation of the relative impact of spatial structure and its overlap with environmental effects, which may potentially reveal other spatially-structured unknown variables driving patterns in species richness (although in principle the coefficients of the environmental variables would be, on average, accurate in respect to spatial autocorrelation, despite that they will vary in space).

Although our goal in this paper is just to show the application of a simple approach to undertake variance partitioning in a large dataset, it is opportune to realize that it reveals that richness patterns in sharks at a global scale are explained by variables with well-known effects, such as temperature range and dissolved O_2. Higher temperature ranges are found mostly in temperate zones of oceans associated with convergence richer areas in resources attracting predators like sharks (Hyrenbach et al. 2000). Although top predators with big body size such as sharks need high levels of oxygen for metabolism and have been associated with areas of higher oxygen supplies (as pointed out by Blank et al. 2002), our results showed a strong negative relation between richness and dissolved oxygen because the biggest supply of dissolved oxygen in oceans are found in polar zones, where the ice acts as a physical and even physiological barrier.

In summary, our approach allows feasible spatial analyses with large datasets without taking into account biases due to autocorrelation but rather to incorporate spatial structure in the data (as in the variance partitioning approach), with the advantage of allowing a better evaluation of the distribution of parameters, their correlation structure, and uncertainty. In our study case, as sample sizes around 1000 cells provide an accurate approximation of effects, it is feasible to use the approach proposed here to explore more complex spatial structures or different deconstruction of richness patterns.

Acknowledgments

Supplementary Materials

The following materials are available as part of the online article from https://escholarship.org/uc/fb

Appendix S1. Median values of the components a, b and c for variance partitioning, modeling shark richness with a GLM as a function of environmental variables and spatial eigenvectors.

Appendix S2. Median of effect size (t-value) and frequency of significance for the GLM modeling shark richness as a function of nine environmental variables and spatial eigenvectors.

Appendix S3. Number of filters selected by PCNM analyses and time taken to perform 1000 simulations for each sample size in a notebook Intel I-5 processor with 8GB RAM.

Appendix S4. A simple R script for performing the approach proposed.

References

Bauman, D., Drouet, T., Dray, S. & Vleminckx, J. (2018a) Disentangling good from bad practices in the selection of spatial or phylogenetic eigenvectors. Ecography, 41, 1638-1649.

Bauman, D., Drouet, T., Fortin, M.J. & Dray, S. (2018b) Optimizing the choice of a spatial weighting matrix in eigenvector-based methods. Ecology, 99, 2159-2166.

Beale, C.M., Lennon, J.J., Elston, D.A., Brewer, M.J. & Yearsley, J.M. (2007) Red herrings remain in geographical ecology: a reply to Hawkins et al. (2007). Ecography, 30, 845-847.

Bini, M.L., Diniz-Filho, J.A.F., Rangel, T.F., et al. (2009) Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. Ecography, 32, 193-204.

Bivand, R. & Wong, D.W.S. (2018) Comparing implementations of global and local indicators of spatial association. Test, 27, 716–748.

Blank, J.M., Morrissette, J.M., Davie, P.S. & Block, B. A. (2002) Effects of temperature, epinephrine and Ca2+ on the hearts of yellowfin tuna (*Thunnus albacares*). Journal of Experimental Biology, 205, 1881-1888.

Borcard, D. & Legendre, P. (2002) All-scale spatial analyses of ecological data by means of principal coordinates of neighbour matrices. Ecological Modelling, 153, 51–68.

Borcard, D., Legendre, P. & Drapeau, P. (1992) Partialling out the spatial component of ecological variation. Ecology, 73, 1045–1055.

Diniz-Filho, J.A.F. & Bini, L.M. (2005) Modelling geographical patterns in species richness

using eigenvector-based spatial filters. Global Ecology and Biogeography, 14, 177-185.

Diniz-Filho, J.A.F., Bini, L.M. & Hawkins, B.A. (2003) Spatial autocorrelation and red herrings in geographical ecology. Global Ecology and Biogeography, 12, 53-64.

Dormann, C.F. (2007) Effects of incorporating spatial autocorrelation into the analyses of species distribution data. Global Ecology and Biogeography, 16, 129-138.

Dormann, C.F., McPherson, J., Araújo, M.B., et al. (2007) Methods to account for spatial autocorrelation in the analyses of distributional species data: a review. Ecography, 30, 609-628.

Dray, S., Bauman, D., Blanchet, G. et al. R Core Team (2020) adespatial: Multivariate Multiscale Spatial Analysis. R package version 0.3-8.

Dray, S., Legendre, P. & Peres-Neto, P.R. (2006) Spatial modeling: a comprehensive framework for principal coordinate analyses of neighbor matrices (PCNM). Ecological Modelling, 196, 483–493.

Dray, S., Pélissier, R., Couteron, P., et al. (2012) Community ecology in the age of multivariate multiscale spatial analyses. Ecological Monographs, 82, 257–75.

Fotheringham, A.S., Brunsdon, C. & Charlton, M. (2002) Geographically weighted regression: the analyses of spatially varying relationships. John Wiley & Sons Ltd., 284 pp.

Griffith, D.A. & Peres-Neto, P. (2006) Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. Ecology, 87, 2603-2613.

Griffith, D.A. (2003) Spatial autocorrelation and spatial filtering: gaining understanding through theory and visualization. Springer-Verlag, 250 pp.

Haining, R. (2003) Spatial data analysis. Cambridge University Press.

Hawkins, B.A., Diniz-Filho, J.A.F., Bini, L.M., et al. (2007) Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology. Ecography, 30, 375-284.

Hyrenbach, K.D., Forney, K.A. & Dayton, P.K. (2000) Marine protected areas and ocean basin management. Aquatic conservation: marine and freshwater ecosystems, 10, 437-458.

IUCN (2019) The IUCN Red List of Threatened Species. Version 6.2. https://www.iucnredlist.org. Downloaded on 04 June 2019.

Kissling, W.D. & Carl, G. (2008) Spatial autocorrelation and the selection of simultaneous autoregressive models. Global Ecology and Biogeography, 17, 59-71.

Kühn, I. (2007) Incorporating spatial autocorrelation may invert observed patterns. Diversity and Distributions, 13, 66-69.

Legendre, P. & Legendre, L. (2012) Numerical ecology, 3rd edn. Elsevier, Oxford, UK.

Legendre, P. (1993) Spatial autocorrelation: trouble or new paradigm? Ecology, 74, 1659-1673.

Lennon, J.J. (2000) Red-shifts and red herrings in geographical ecology. Ecography, 23, 101-113.

Lichstein, J.W., Simons, T.R., Shriner, S.A. & Franzreb, K.E. (2002) Spatial autocorrelation and autoregressive models in ecology. Ecological Monographs, 72, 445-463.

Lucifora, L.O., García, V.B. & Worm, B. (2011) Global diversity hotspots and conservation priorities for sharks. PloS One, 6, e19356, https://doi.org/10.1371/journal.pone.0019356

Oksanen, J., Blanchet, F.G., Friendly, M., et al. (2007) Vegan: community ecology package. R package version 2.5-6.

Oliveira, G., Rangel, T.F., Lima-ribeiro, M.S., Terribile, L.C. & Diniz-Filho, J. A. F. (2014) Evaluating, partitioning, and mapping the spatial autocorrelation component in ecological niche modeling: a new approach based on environmentally equidistant records. Ecography, 37, 637-647.

Pebesma, E.J. (2004) Multivariable geostatistics in S: the gstat package. Computers & Geosciences, 30, 683-691.

Pinheiro, J., Bates, D., DebRoy, S. & Sarkar, D. (2020). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-144.

Rangel, T.F., Diniz-Filho, J.A.F. & Bini, L.M. (2006) Towards an integrated computational tool for spatial analyses in macroecology and biogeography. Global Ecology and Biogeography, 15, 321–327.

Rangel, T.F., Diniz-Filho, J.A.F. & Bini, L.M. (2010) SAM: a comprehensive application for spatial analyses in macroecology. Ecography, 33, 46-50.

Safi, K. & Pettorelli, N. (2010) Phylogenetic, spatial and environmental components of extinction risk in carnivores. Global Ecology and Biogeography, 19, 352-362.

Tittensor, D.P., Mora, C., Jetz, W., Lotze, H.K., Ricard, D., Berghe, E.V. & Worm, B. (2010) Global

patterns and predictors of marine biodiversity across taxa. Nature, 466, 1098-1101.

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F. & De Clerck, O. (2012) Bio-ORACLE: a global environmental dataset for marine species distribution modelling. Global Ecology and Biogeography, 21, 272–281.

Vilela, B. & Villalobos, F. (2015) letsR: a new R package for data handling and analyses in macroecology. Methods in Ecology and Evolution, 6, 1229-1234.

Worm, B. & Tittensor, D.P. (2018) A theory of global biodiversity (MPB-60). Princeton University Press. 232 pp.