ECOGRAPHY

Research

Disentangling good from bad practices in the selection of spatial or phylogenetic eigenvectors

David Bauman, Thomas Drouet, Stéphane Dray and Jason Vleminckx

D. Bauman (http://orcid.org/0000-0001-9115-6518) (dbauman@ulb.ac.be)(davbauman@gmail.com) and T. Drouet, Univ. Libre de Bruxelles, Laboratoire d'Écologie Végétale et Biogéochimie, Brussels, Belgium. – S. Dray, Univ. Lyon, Univ. Claude Bernard Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive, Villeurbanne, France. – J. Vleminckx (http://orcid.org/0000-0001-7557-0866), Dept of Integrative Biology, Univ. of California, Berkeley, CA, USA, and Univ. Libre de Bruxelles, Service d'Évolution Biologique et Écologie, Brussels, Belgium.

Ecography
41: 1638–1649, 2018
doi: 10.1111/ecog.03380

Subject Editor: Luis Mauricio Bini
Editor-in-Chief: Miguel Araújo
Accepted 18 December 2017

Eigenvector mapping techniques are widely used by ecologists and evolutionary biologists to describe and control for spatial and/or phylogenetic patterns in their data. The selection of an appropriate subset of eigenvectors is a critical step (misspecification can lead to highly biased results and interpretations), and there is no consensus yet on how to proceed. We conducted a ten-year review of the practices of eigenvector selection and highlighted three main procedures: selecting the subset of descriptors minimising the Akaike information criterion (AIC), using a forward selection with double stopping criterion after testing the global model significance (FWD), and selecting the subset minimising the autocorrelation in the model residuals (MIR). We compared the type I error rates, statistical power, and R^2 estimation accuracy of these methods using simulated data. Finally, a real dataset was analysed using variation partitioning analysis to illustrate to what extent the different selection approaches affected the ecological interpretation of the results. We show that, while the FWD and MIR approaches presented a correct type I error rate and were accurate, the AIC approach displayed extreme type I error rates (100%), and strongly overestimated the R^2. Moreover, the AIC approach resulted in wrong ecological interpretations, as it overestimated the pure spatial fraction (and the joint spatial-environmental fraction to a lesser extent) of the variation partitioning. Both the FWD and MIR methods performed well at broad and medium scales but had a very low power to detect fine-scale patterns. The FWD approach selected more eigenvectors than the MIR approach but also returned more accurate R^2 estimates. Hence, we discourage any future use of the AIC approach, and advocate choosing between the MIR and FWD approaches depending on the objective of the study: controlling for spatial or phylogenetic autocorrelation (MIR) or describing the patterns as accurately as possible (FWD).

Keywords: Akaike information criterion (AIC), ecological processes, eigenvector selection, forward selection, Moran's eigenvector maps (MEM), phylogenetic eigenvector regression (PVR), principal coordinates of neighbour matrices (PCNM), spatial eigenvector mapping (SEVM), spatial patterns, variation partitioning

NORDIC SOCIETY OIKOS

www.ecography.org

Introduction

Identifying and explaining spatial structures in the distribution of organisms in natural communities are two longstanding challenges in ecology (Cormack and Ord 1979, Legendre and Fortin 1989, Legendre and Legendre 2012). Clustered distributions may result from spatially heterogeneous habitat filters, community intrinsic processes (e.g. dispersal limitation), historical processes or biotic interactions (Legendre and Fortin 1989, Legendre 1993, Barbier et al. 2008, Pinto and MacDougall 2010). Disentangling the relative influence of these processes is a challenging task. Indeed, even if only community-intrinsic processes occur, spatially clustered distributions may overlap by chance with spatially structured environmental variables ('spatial nuisance', Peres-Neto and Legendre 2010). On the other hand, spatial patterns of species distribution may give crucial clues about possible underlying ecological processes (McIntire and Fajardo 2009), a view in which space is not considered an issue but a surrogate of the underlying ecological processes ('spatial legacy', Peres-Neto and Legendre 2010, Legendre and Legendre 2012). These paradigms of 'nuisance' and 'legacy' also apply for temporal and phylogenetic patterns (Diniz-Filho and Bini 2005, Griffith and Peres-Neto 2006, Diniz-Filho et al. 2012, Baho et al. 2015).

Several methods have been developed to overcome the 'nuisance' issue or properly detect the 'legacy', depending on the objectives of the study. Among others, eigenvector-based methods produce predictors that can be used either to 1) control for autocorrelation in model residuals, or 2) highlight and interpret the patterns of autocorrelation. In the spatial context, Griffith (1996) provided pioneering work in the geographic literature showing how spatial eigenvectors can be used to describe map patterns or control for spatial autocorrelation in residuals. In the same way, evolutionary biologists have been using phylogenetic eigenvector regression (PVR; Diniz-Filho et al. 1998, 2012) to control for phylogenetic autocorrelation in comparative analysis or to study patterns of trait evolution along a phylogeny (see Diniz-Filho et al. 2012 for more details). This paper focuses on spatial eigenvectors, but the general conclusions probably hold for the phylogenetic context.

Eigenvector-based methods have been receiving increasing attention in the past decade in community ecology, especially with the development of Moran's eigenvector maps (MEM, Dray et al. 2006). This method generalises the ad hoc principal coordinates of neighbour matrices (PCNM, Borcard and Legendre 2002, Griffith and Peres-Neto 2006) that consist of the principal coordinate analysis of a truncated geographic distance. Whereas the initial definition of PCNM suffers from a lack of statistical formalism (e.g. the empirical basis of the definition of the truncated Euclidean distance matrix, or the scaling of the PCNM base functions), MEM provides a well-defined theoretical framework based on the notion of spatial autocorrelation, which encompasses PCNM, and will thus be used in the rest of the paper (see Dray et al. 2006 for further details).

The first step to use MEM is to define the spatial relationships between sampling sites using a spatial weighting matrix (**W**). The latter is built by computing the Hadamard product of a connectivity (**B**) and weighting (**A**) matrix (Dray et al. 2006). As various candidates can be considered for the **B** and **A** matrices, a broad range of **W** matrices can be generated (see Dray et al. 2006 for details). Even if the choice of a spatial weighting matrix is a crucial issue, this work does not deal with this question, and the following developments consider that this step has been performed properly (although there currently is no consensus method for the choice of a **W** matrix, see Discussion). The **W** matrix is then doubly centred (i.e. by rows and columns) and diagonalised. This last step generates $n - 1$ eigenvectors (where n is the number of sampling sites) associated with eigenvalues directly proportional to Moran's I coefficients. The first eigenvector has the highest Moran's I value and thus models broader structures than the second eigenvector, itself displaying a broader spatial pattern than the third eigenvector, and so on. These eigenvectors (hereafter referred to as MEM or MEM variables; see the MEM.ALL matrix in Fig. 1) can then be used 1) as predictors to study the multiscale spatial patterns of the response data (e.g. species abundances; Declerck et al. 2011, Bauman et al. 2016, Vleminckx et al. 2017), or 2) as covariables (spatial filters) to remove spatial autocorrelation from residuals (spatial eigenvector mapping, SEVM; Griffith 2003, Diniz-Filho and Bini 2005, Dormann et al. 2007, Bini et al. 2009, Corkeron et al. 2011). However, all eigenvector-based methods, including MEM, face a common issue. Specifically, a variable selection is required to avoid model overfitting and a loss of statistical power to detect the environmental contribution to the variability of the response data (Griffith 2003, Dray et al. 2006, Blanchet et al. 2008, Peres-Neto and Legendre 2010, Diniz-Filho et al. 2012). Although several procedures have been proposed relatively separately in the scopes of MEM, SEVM, or PVR, there is still no consensus on a most suitable eigenvector selection method. The goal of this study is to test and compare the most common eigenvector selection methods to define a guideline of the best practices, and rule out some biased and underpowered approaches.

Selecting a subset of eigenvectors

A review of the literature showed that three main approaches to select spatial variables have been widely used in the framework of eigenvector-based methods. The first method, further referred to as the 'AIC approach', originates from Dray et al. (2006). This procedure relies on the computation of a corrected version of the Akaike information criterion (AIC_c; Hurvich and Tsai 1989) extended to the case of multivariate response data (Godinez-Dominguez and Freire 2003). This approach starts by reordering the complete set of MEM (MEM.ALL) by decreasing (canonical) coefficient of

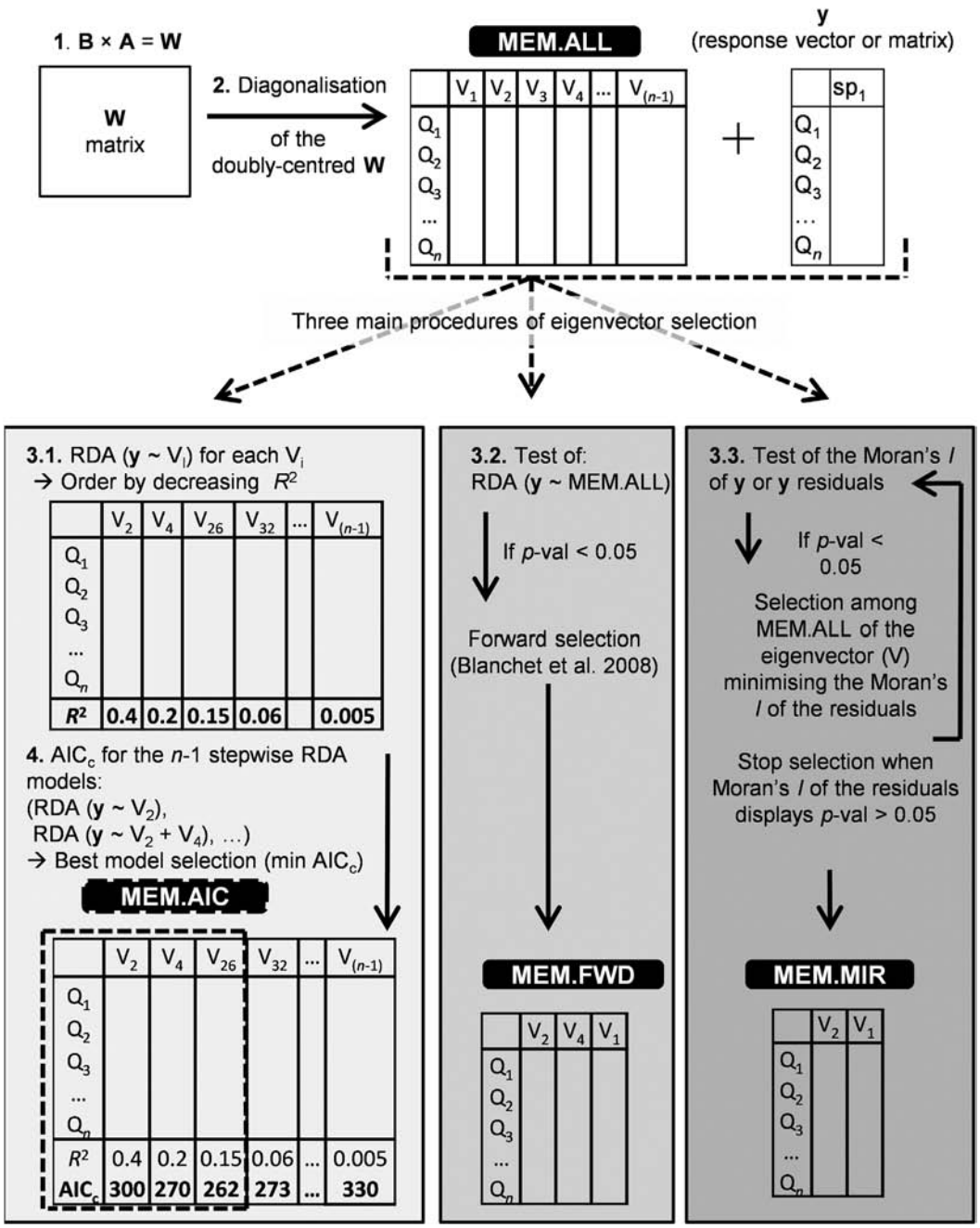

Figure 1. Schematic illustration of the three selection procedures for a univariate response vector **y**. Q: quadrat; V: spatial, temporal or phylogenetic eigenvector; sp: species; RDA: redundancy analysis. The first step corresponds to the Hadamard product of a connectivity (**B**) and weighting matrix (**A**). The spatial weighting matrix **W** is then doubly centred and diagonalised, leading to the MEM.ALL matrix, a matrix of $n - 1$ eigenvectors. At this point, one model selection has to be used to reduce the number of eigenvectors. 1) Step 3.1. In the AIC approach of Dray et al. (2006), the R^2 of each MEM is computed by constrained ordinations, most often a redundancy analysis (RDA; multivariate response), or linear regression models (univariate response), and the MEM variables are ordered by decreasing R^2. At step 4, the ordered MEM are included one by one to the model and an AIC_c value is computed at each step. The subset of MEM corresponding to the min AIC_c, although initially designed only to select a **W** matrix, is then used as the best spatial model of the corresponding **W** matrix (MEM.AIC). 2) In the forward selection developed by Blanchet et al. (2008; step 3.2.), a global significance test is computed with all MEM (associated to positive and negative eigenvalues separately, using a p-value correction for multiple tests). If the global model is significant, a forward selection is performed with the p-value and global adjusted R^2 as stopping criteria, providing a subset of MEM (MEM.FWD) to be used for further analyses. 3) The step 3.3. corresponds to the MIR approach (Griffith and Peres-Neto 2006), and aims at finding the smallest subset of MEM variables minimising the autocorrelation (Moran's *I*) in the residuals. The method first tests the significance of the Moran's of **y** with a permutation test, and if significant, looks for the predictor best minimising it. The significance of the Moran's *I* is then recomputed, and if significant, the procedure searches for the next eigenvector best minimising the Moran's *I* of the residuals of the model of **y** as a function of the first added eigenvector. The procedure goes on until the Moran's *I* is not significant anymore.

determination (R^2; Step 3.1 in Fig. 1). The reordered MEM are then added one by one in an explanatory model, and the AIC_c is computed at each step. The subset of MEM corresponding to the minimum AIC_c (MEM.AIC in Fig. 1) is then selected.

The AIC approach was originally developed by Dray et al. (2006) to select a spatial weighting matrix **W** among a range of candidates. However, it has been utilised to select a subset of MEM in nearly all the works using the method since 2006 (see 'Results'). This misuse emerged from the subtlety that, although the method selects a subset of MEM (Step 4 in Fig. 1), the only purpose of the latter is to detect the best **W** matrix. Contributing to this ambiguity, Borcard et al. (2011) defined the 'champion model' as the MEM subset best minimising the AIC_c among the different **W** matrices tested. Hence, they recommended the use of the AIC approach both for selecting a **W** matrix and a 'very best model' within it (and provided an R code to do so). Our study will therefore consider this misuse of the procedure and will assess its suitability for spatial eigenvector selection.

The second approach, proposed by Blanchet et al. (2008) and further referred as the 'FWD approach', is based on a forward selection with double stopping criterion. The method selects a subset of explanatory variables while maintaining a correct type I error rate and avoids model overfitting (MEM.FWD, Step 3.2 in Fig. 1, also presented in Borcard et al. 2011). Blanchet et al. (2008) showed that the type I error rate of the stepwise procedure was correct if and only if a global test of significance was conducted previous to applying the forward selection (i.e. testing the model containing all MEM variables). Moreover, adding to the classical p-value stopping criterion of the forward selection, Blanchet et al. (2008) introduced the adjusted coefficient of determination (R^2_{adj}) of the global model as additional criterion, preventing the selection of models from explaining more than the complete set of explanatory variables, and showed that it allowed avoiding an inflated R^2_{adj}. If both positive and negative spatial autocorrelations are of interest, then the method is performed separately on two subsets of MEM (displaying positive and negative eigenvalues, respectively), and is followed by a p-value correction for multiple tests (Blanchet et al. 2008).

The third most widely used variable selection method consists of selecting the smallest subset of predictors that best minimises the spatial autocorrelation in model residuals (Griffith and Peres-Neto 2006, Tiefelsdorf and Griffith 2007). It was shown to be a reliable and efficient way to control spatial (Bini et al. 2009) and phylogenetic autocorrelation (Diniz-Filho et al. 2012), and to maintain a high statistical power. The method consists of removing spatial effects from the residuals of explanatory models (e.g. linear models) that relate a univariate response variable to a set of explanatory predictors (environmental factors, for instance). The method starts by computing the autocorrelation (Moran's *I*) of the residuals and tests its significance by a permutation procedure. If the test is significant, each MEM is added separately to the regression model and the Moran's *I* of the residuals is recomputed for each model. The procedure is repeated until the residuals display no more significant spatial autocorrelation. This minimisation of Moran's *I* in the residuals (MIR approach) both ensures the independence of the residuals and maintains a high statistical power for the explanatory variables of interest as it adds as few MEM variables as possible (Griffith and Peres-Neto 2006). Whereas the FWD and AIC approaches are designed for studies where space is considered a legacy (their criteria maximise the spatial fit of the model), the MIR approach is oriented towards the spatial nuisance viewpoint (MEM used as covariables or spatial filters; Corkeron et al. 2011, Siesa et al. 2011).

In this paper, we decided to compare the AIC, FWD, and MIR approaches only in a context where space is considered a legacy. This choice is justified because our review demonstrates that these three methods are the most used in the literature and that the MIR approach has also been applied for spatial legacy purposes even if it was not designed for this objective. To test the suitability of this method for a 'spatial legacy approach', we slightly adapted the method so that no environmental variables were needed. To do so, we introduce the MEM variables by the procedure described above in a model that contains only an intercept term (Step 3.3 in Fig. 1).

Other criteria were proposed but are much less used (e.g. using all MEM, only those significantly related to the response data, displaying a Moran's *I* or R^2 greater than a given threshold value, displaying a significant Moran's *I* value (permutation test), or a combination of these criteria).

Our study is structured into three sections. We present a representative ten-year review (2006–2016) of the peer-reviewed literature using eigenvector-based spatial methods. Then, we use simulated data to evaluate the type I error rate, statistical power, and R^2 estimation accuracy of the AIC, FWD, and MIR procedures. Lastly, we explore a real dataset to compare to what extent the different selection methods can affect the results and their ecological interpretation.

Material and methods

Review of the literature

The review aimed to give a representative view of the selection practices of spatial predictors in the past 10 yr. A research of the peer-reviewed articles was carried out with the terms 'Moran's eigenvector maps', 'Moran eigenvector maps', and 'spatial eigenvector mapping' in separate Google Scholar searches (on January 2017). We limited the research to these two criteria, as they covered the principal selection methods and consisted of a large number of articles (301). All 'Material and methods' sections of the articles reviewed were carefully analysed to determine how the spatial predictors were selected.

Spatial eigenvector selection: type I error rate

We considered a study area of 1000 × 500 cells for both a regular (117 sites positioned on a 13 × 9 grid covering the entire area) and a random sampling design (117 randomly chosen cells, Supplementary material Fig. A1). The type I error rates were computed for the three selection approaches (Fig. 1) by considering a random univariate response variable **y** generated from four different distributions: uniform, normal, exponential, and cubed exponential, following Anderson and Legendre (1999) and Manly (2007).

We built a spatial weighting matrix **W** (Fig. 1) generated with distance-based criteria to define the **B** and **A** matrices. Two sites were considered connected if they were found within a threshold distance corresponding to the shortest distance that keeps all sites connected (i.e. the length of the largest edge of the minimum spanning tree). The links were then weighted by $f(D_{ij}) = 1 - (D_{ij}/4t)^2$, where D_{ij} is the Euclidean distance between sites i and j, and t is the threshold value. The MEM predictors were then obtained by the diagonalisation of the doubly centred spatial weighting matrix **W**. This produced 116 eigenvectors (MEM.ALL in Fig. 1). Dray et al. (2006) demonstrated that this particular case of specification of the **W** matrix produced distance-based MEM (db-MEM) that correspond roughly to the original PCNM method. The main difference is that the PCNM approach is originally defined by computing a principal coordinate analysis of a truncated Euclidian distance matrix. Thus, it does not return the full set of $n - 1$ eigenvectors and does not respect the strict equivalence between eigenvalues and Moran's index of autocorrelation (more details in Dray et al. 2006).

The three selection methods (FWD, AIC, MIR; Fig. 1) were then applied on the full set of spatial predictors (MEM.ALL). The tests of significance were performed with 999 permutations, and 10 000 simulations were conducted for each scenario. Type I error rates were computed as the proportions of significant results (for a significance level of 0.05) among the 10000 simulations.

Spatial eigenvector selection: power and accuracy

To test the power and a possible bias of the R^2_{adj}, the three selection methods were applied in a second set of simulations. We used the same simulation design as for estimating type I error rates except for building the response variable **y**. Instead of using random numbers, we built spatially positively autocorrelated **y** as the sum of a linear combination of three MEM variables with a random normal noise, following Jombart et al. (2009; details in Supplementary material Appendix 1). We considered three cases where the response variable **y** was structured either at broad (first three MEM), medium (MEM 25 to 27), or fine spatial scale (the last three positively autocorrelated MEM; Supplementary material Appendix 1), for both the regular and random sampling designs. Then, a linear regression explaining **y** by the three corresponding MEM variables was computed, and the resulting R^2_{adj} was considered the reference value of the spatial signal contained in **y**. The three selection approaches were then applied separately for selecting the subset of MEM variables that best explained **y**, and the adjusted R^2 of the resulting models were compared to the reference value, providing a ΔR^2 ($R^2_{\text{AIC, FWD, or MIR}} - R^2_{\text{reference}}$). This procedure was repeated 10 000 times. The power was then computed as the proportion of simulations leading to a significant model (significance level of 0.05), while the mean of ΔR^2 was used as a measure of the accuracy of the selection approaches. All R^2 values presented in this study are adjusted R^2 (Ezekiel 1929). Also note that, in this work, 'scale' exclusively refers to the spatial feature 'focus', sensu Scheiner (2011), that is, the dimension of the aggregated grains of a spatial pattern, and is never used to refer to the 'extent' or 'grain' of the study to avoid confusion.

Illustration on a real dataset: tree species of a Miombo Woodland

The three methods of selection were also applied to the Mikembo forest data (Muledi et al. 2017). This dataset corresponds to an exhaustive census and mapping of all individual trees (≥ 10 cm diameter at breast height; 4604 individuals) in a 10-ha tropical dry woodland (500 × 200 m) located in the eastern part of Upper Katanga (Democratic Republic of the Congo). The plot was divided into a grid of 160 quadrats of 25 × 25 m in which 36 soil parameters were measured (texture, soil depth, soil chemistry, etc; see Muledi et al. 2017 for further information on the dataset). The MEM variables were generated and used in a univariate context for the 24 most abundant tree species of the forest (using db-MEM based on the PCNM criterion as in the simulation study), and the selection was performed using the three selection methods (Fig. 1). We compared the proportion of species displaying a significant spatial structure and the corresponding R^2 for the subset of MEM variables selected by each method. Additionally, a variation partitioning analysis (VP; Borcard et al. 1992, Peres-Neto and Legendre 2010) was applied separately for each species displaying a significant spatial structure to assess the pure and shared effects of soil and spatial variables on species distributions. The VP were performed with each MEM subset to illustrate to what extent ecological conclusions can be affected by the different procedures of selection. We focused on fractions corresponding to the effect of a spatially structured environment ($R^2_{ENV\text{-}SPA}$) and the pure spatial fraction ($R^2_{SPACE.PURE}$). A shared effect of soil and spatial variables ($R^2_{ENV\text{-}SPA}$) may indicate an induced spatial dependence (i.e. the spatial signature of an influent environmental factor; Legendre and Legendre 2012), while pure spatial effects ($R^2_{SPACE.PURE}$) may relate to dispersal limitations, biotic interactions, or unmeasured environmental parameters. For each species, we computed $\Delta R^2_{\text{AIC-FWD}}$ ($R^2_{\text{MEM.AIC}} - R^2_{\text{MEM.FWD}}$), $\Delta R^2_{\text{AIC-MIR}}$ ($R^2_{\text{MEM.AIC}} - R^2_{\text{MEM.MIR}}$), and $\Delta R^2_{\text{FWD-MIR}}$ ($R^2_{\text{MEM.FWD}} - R^2_{\text{MEM.MIR}}$), for both $R^2_{ENV\text{-}SPA}$ and $R^2_{SPACE.PURE}$. These indices allowed comparing how the selection procedures influenced the estimations of the spatial fractions in VP.

All the analyses were performed in the R environment (ver. 3.3.2., R Development Core Team) using the packages

'vegan' (Oksanen et al. 2017), 'spdep' (Bivand 2006), and 'adespatial'. The R code used to run the simulations is provided in Supplementary material Appendix 2. The R function 'MEM.moransel' applying the MIR approach on a model that only contains an intercept term is provided in Supplementary material Appendix 3.

Results

Review of the literature

We analysed 301 articles published between 2006 and 2016 using MEM, PCNM, and SEVM (and PVR to a lesser extent; Supplementary material Appendix 4 for the list of references, and Supplementary material Table A1 for the methodological details). The AIC approach was used alone in 15% of the studies, and was probably used in an additional 13% of articles lacking accurate information regarding the MEM selection procedure performed (e.g. citing Dray et al. (2006) for the MEM variable selection methodology without explicitly specifying whether the AIC approach was used, or referring to Dray et al. 2006 for further details; 'AIC' in Fig. 2). The AIC approach was also used in combination with the FWD approach in 4% of the studies. The latter generally used the AIC procedure for the selection of a spatial weighting matrix **W** and the FWD approach to select MEM variables within the chosen **W** matrix (another 2% likely did the same but lacked methodological specification; 'AIC + FWD' in Fig. 2a). Therefore, up to 32% of the reviewed studies used the AIC procedure. The FWD approach ('FWD' in Fig. 2) was used alone in 21.5% of the studies. Among all the studies applying this latter method (27.2%; i.e. alone or combined with another approach), few of them explicitly mentioned whether a global test of significance had been conducted. The classic forward selection (i.e. no global test of significance performed) was used in 5.6% of the studies. The minimisation of Moran's I in the residuals was used in 24.5% of the reviewed studies ('MIR' in Fig. 2). Among other approaches, 4.0% of the studies selected the MEM variables displaying a significant Moran's I ('Signif. Moran's I' in Fig. 2a). Also, 5.3, 4.6, and 1.6% of the studies used spatial predictors significantly related to the response variable(s), displaying a Moran's I superior to a threshold value (generally 0.1), and associated with an R^2 superior to a threshold value, respectively ('Others' for the three last categories; Fig. 2a). The use of these methods was not exclusive; thus, they represented 9.6% of the reviewed works. All MEM associated with positive eigenvalues were used without variable selection in 10.0% of the studies ('No selection' in Fig. 2).

As illustrated in Fig. 2b, while the number of studies using eigenvector-based methods clearly increased over the years, we observed no clear tendency of an increasing or decreasing use of one method at the expense of the others among the three most-used methods. Finally, only 62% of the studies properly characterised the chosen **W** matrix, and 57% of them used the original PCNM method. Among the latter, 43% claimed they used db-MEM while they used the original PCNM.

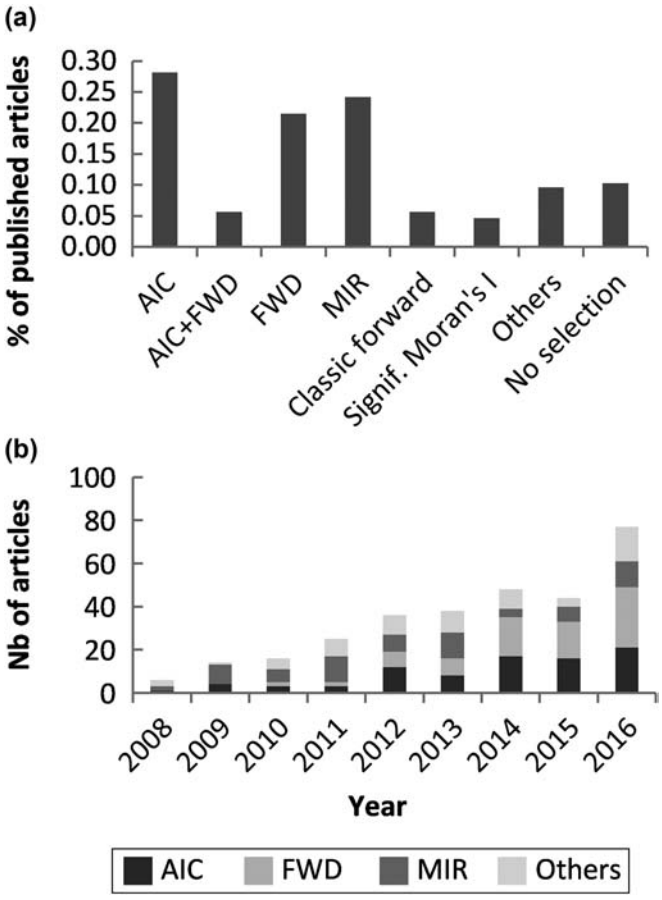

Figure 2. Review of the eigenvector selection methods used from 2006 to 2016 (number of studies: 301). (a) Percentage of the articles reviewed using MEM.AIC (AIC), MEM.FWD (FWD), a combination of the two former (FWD + AIC), the MIR approach (MIR), the classic forward selection (without a global test and with the p-value as only stopping criterion), all MEM variables displaying a significant Moran's I (Signif. Moran's I), all the positively autocorrelated MEM variables (No selection), or one of the following criteria: all MEM significantly related to **y**, or displaying a Moran's I or a R^2 superior to a threshold value (Others). (b) Number of articles that used MEM.AIC, MEM.FWD, MEM.MIR, or one of the remaining methods from 2008 to 2016 (only few studies were recorded before 2008, and are therefore not represented here).

Simulation study: type I error rate

The simulations performed using the four distribution types provided very similar results, so that we reported only the results for the uniform distribution (Supplementary material Table A2 for the complete results).

Figures 3a illustrates the type I error rate of the MEM variables selected by the three selection approaches. While the FWD and MIR methods presented correct type I error

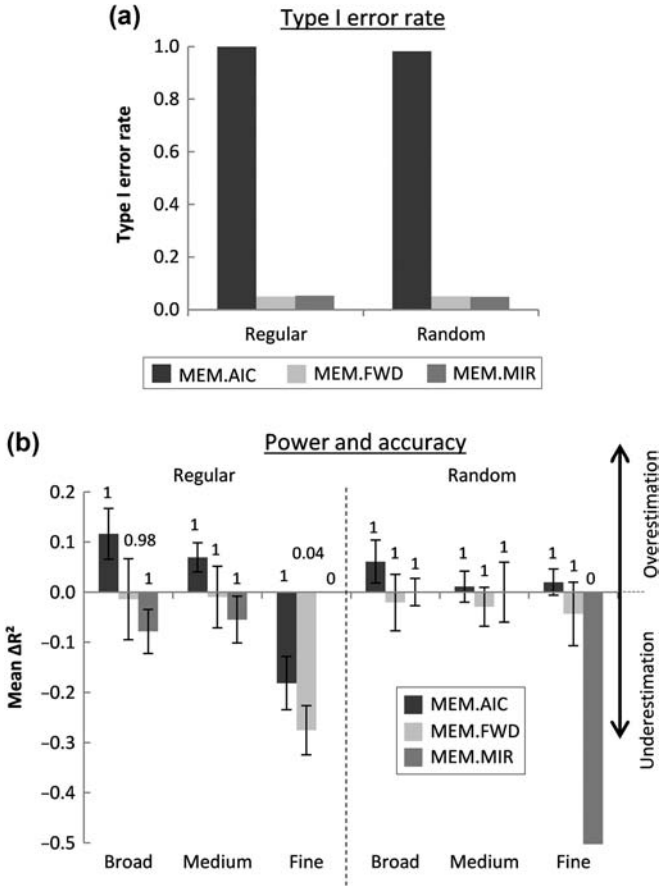

Figure 3. (a) Type I error rate of the three eigenvector selection methods. (b) Power and R^2 estimation accuracy of the AIC, FWD, and MIR procedures computed on 10 000 simulated response variables **y** structured at broad, medium, or fine scale, both in a regular and in a random sampling design. The y-axis gives the mean of the ΔR^2, that is, $R^2_{\text{simulated}} - R^2_{\text{reference}}$. A positive value indicates an overestimation of the reference R^2, while a negative value corresponds to an underestimation. The absence of a bar for the MIR approach with a response vector structured at fine scale and sampled in the regular sampling design indicates that the ΔR^2 could not be computed, as the ΔR^2 is only computed on the basis of the significant simulations and the statistical power of this scenario was null. Vertical bars represent standard deviations. The number above each standard deviation bar corresponds to the statistical power.

rates (~0.05), the AIC procedure detected a significant spatial signal in the response **y** in 100% of the simulations.

Simulation study: power and accuracy

Figure 3b shows the power of the selection methods and the mean ΔR^2 (simulated value – real value) obtained after 10 000 simulations for broad, medium, and fine spatial scales. The AIC approach always presented a power of 1, regardless of the spatial scale or type of sampling design, and systematically produced overestimated R^2, except at fine spatial scales for the regular sampling design (strong underestimation of the actual R^2 value). The FWD procedure always presented a power of ~1 and allowed retrieving a very accurate estimation of the real values of R^2 regardless of the spatial scale or type of sampling design, except for the fine-scale structure for the regular design. In the latter case, the power was very low (0.04), and R^2 was strongly underestimated. Finally, the MIR approach had a power of 1 at broad and medium scales, regardless of the type of sampling design. The method moderately underestimated the actual R^2 when using the regular design, but provided accurate estimates with the random design. However, the power of this approach to detect fine-scale spatial structures was null for both sampling designs. Regarding the number of MEM, the average number of selected variables was 6 ± 2.5, 6 ± 0, and 3 ± 1, with the FWD, the AIC, and the MIR approach, respectively.

Selection method effects on ecological interpretations of a real dataset

The three selection methods were used to select a subset of spatial predictors for the 24 most abundant tree species of Mikembo ($20 \leq n \leq 1239$). We observed significant spatial structures for around 50% of the species using the FWD and MIR approaches, while the AIC approach detected highly significant structures for all species. The FWD selection detected a weak spatial pattern for one species that the MIR approach did not detect, and the MIR approach also detected weak patterns for three species that the FWD did not detect. The main results (Fig. 4) therefore focus on the 11 species that displayed spatial patterns according to both the FWD and MIR methods. The R^2 obtained with MEM.AIC was systematically higher than that obtained with MEM.FWD (up to 23% more) and MEM.MIR (up to 33% more, Fig. 4a) for all species. Regarding VP, as illustrated in Fig. 4b, c, most of the R^2 overestimation of the AIC approach ended up in the pure spatial fraction of the VP (ΔR^2 up to 0.19) and, to a minor extent, in the structured environmental fraction (ΔR^2 up to 0.07). The results of $\Delta R^2_{\text{AIC-MIR}}$ were very similar to those of $\Delta R^2_{\text{AIC-FWD}}$ (Fig. 4b) and are therefore not presented here.

Discussion

AIC approach: a highly biased selection method

The simulations revealed that the AIC approach always detected spatial structures when there actually were none and that it always overestimated the actual proportion of spatially structured variability in **y**. The inferential outcomes of these results were illustrated on the real data, showing that spatial structures were wrongly detected for half of the species, while all spatial R^2 were overestimated when using the AIC selection (up to 23% with respect to the most accurate method, the FWD approach, Fig. 3b). This resulted in spurious ecological interpretations regarding the nature of the processes driving the spatial distribution of organisms

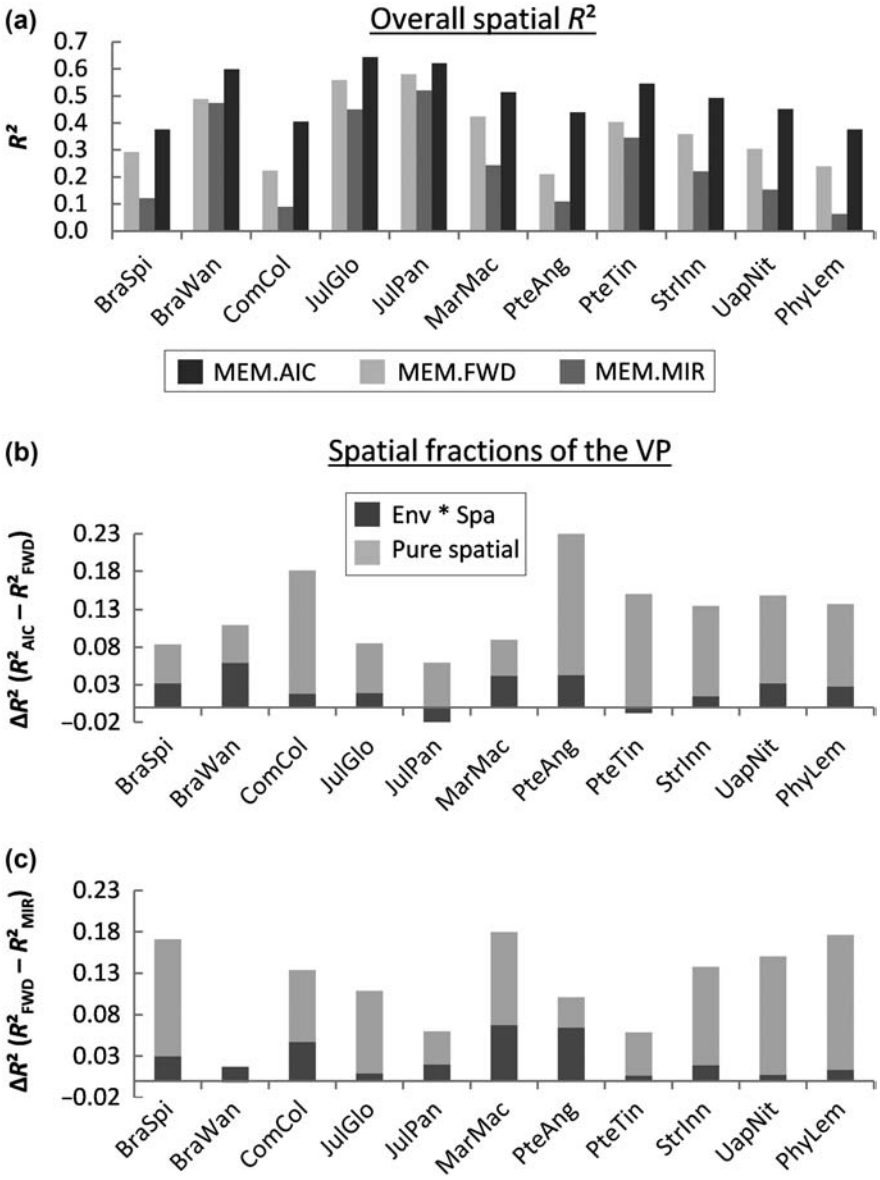

Figure 4. Illustration of the different selection methods on a real dataset. (a) Spatial R^2 obtained using MEM.FWD, MEM.AIC, and MEM.MIR for all Mikembo tree species displaying a significant spatial structure according to both the FWD and the MIR procedures (i.e. the methods with a correct type I error rate, Fig. 3a). (b) and (c) Proportions of the R^2 overestimation ($\Delta R^2_{\text{AIC-FWD}}$ and $\Delta R^2_{\text{FWD-MIR}}$, respectively) ending up in the pure spatial and in the shared environment-space fractions of the VP ('Pure spatial' and 'Env * Spa', respectively). The ΔR^2 were obtained by computing the difference between the R^2 values of MEM.AIC and MEM.FWD (b), and MEM.FWD and MEM.MIR (c). The x-axis corresponds to the tree species abbreviations (details in Muledi et al. 2017) of the species displaying a significant spatial structure according to both the FWD and MIR procedures.

(mainly an overestimation of community-intrinsic processes or unmeasured environmental parameters, Fig. 4b, c). These results show that the AIC approach systematically selects the MEM variables that best fit **y** among a high number of orthogonal explanatory variables, regardless of the global model significance, just as the classical forward selection does (Westfall et al. 1998, Blanchet et al. 2008). This explains why this method displays such a huge type I error rate. Moreover, while the FWD approach controls for R^2 inflation using the R^2 of the global model as second stopping criterion, the AIC approach does not apply any kind of equivalent control, which may partly explain why it always explained more than it should have in the simulations (Fig. 3b). We applied a global test of significance and added the adjusted R^2 stopping criterion to the AIC approach before using it on the simulated structured **y**. The results were nearly identical to those obtained with the FWD approach (Supplementary material Table A3), hence indicating that the origin of the biases is not the AIC per se, but the absence of a global test of significance and a criterion controlling for model overfitting (such

as the global R^2). However, while the classic (univariate) AIC is computed on the basis of a parametric model, Godinez-Dominguez and Freire (2003) proposed a rather simplistic way to compute AIC for the redundancy analysis (RDA; multivariate context). They make a simple but wrong analogy between the univariate residual sum of squares (RSS) and its multivariate analogue, replacing the RSS by the RSS_{RDA} in the classical definition of AIC for linear models. However, as they do not clearly assume any distributional assumptions (e.g. multivariate normal as Pech and Laloë 1997) in the RDA model, their definition of AIC is probably wrong and not applicable. We therefore advocate the use of the R^2_{adj} as a criterion of selection when using the FWD approach, following Blanchet et al. (2008). The AIC approach should be abandoned for future works involving spatial, temporal, or phylogenetic predictors, whether for selecting a subset of predictors (biases highlighted by our results), and probably also for choosing a best-suited **W** matrix (lack of statistical ground for multivariate response data). In addition, the optimisation of the choice of a **W** matrix is likely to be biased (inflated type I error rate) if not associated to a p-value correction for multiple testing, as the chance of detecting a significant **W** matrix by chance is expected to increase with the number of **W** matrices compared.

Fine-scale positively autocorrelated patterns: limited power and underestimated R^2

Our results confirmed a limited power of MEM at fine spatial scales (Layeghifard et al. 2015), and highlighted a systematic underestimation of the actual R^2. A combined approach using MEM for detecting broad and medium-scale patterns and spatial point pattern analysis (i.e. an individual-based approach; see Velázquez et al. 2016 for a review) for fine-scale patterns may be a good strategy to avoid the limited power issue at fine scales for positively autocorrelated structures. The advantage of exploring spatial data through these different methods is that, while MEM is most powerful at broad and medium scales, the spatial point pattern analysis successfully highlights patterns and processes occurring at very local scales, and could therefore compensate the lack of power of MEM. Nevertheless, this would require knowing the spatial location of all individuals, a condition sometimes difficult to meet. In addition, further work is still needed to test and explore the potential of combining individual-based and site-based methods (such as spatial point pattern analysis and MEM) to detect the spatial structures of species and communities and interpret them in terms of ecological processes.

In this study, we focused on positively autocorrelated spatial structures, and the fine scale patterns were therefore positively autocorrelated and generated with the MEM variables associated to small positive eigenvalues. However, although much less investigated by ecologists, fine scale patterns can also be negatively autocorrelated. To explore this point, we also simulated response variables using the last three MEM variables that are the most negatively autocorrelated (smallest eigenvalues) and tested the statistical power and accuracy of the different eigenvector selection approaches as for the other simulation scenarios. These supplementary analyses revealed that the power was close to 1 and that the R^2 estimation accuracy was good for both the FWD and the MIR approaches (Supplementary material Table A4). Hence, the statistical power of MEM actually decreases with the absolute value of the eigenvalue associated to spatial predictors, indicating that the relation between statistical power and spatial scale is not straightforward. The relation between power and eigenvalue also explains why the power of the FWD approach was much higher for the fine scale patterns on a random sampling design than on a regular sampling design. Indeed, there are less positive MEM variables when the design is irregular and the last eigenvectors with positive eigenvalues therefore have a higher eigenvalue than the last positively autocorrelated eigenvectors of a regular sampling design. Further works will be necessary to better investigate the relation between the eigenvalues of the spatial eigenvectors, the scales of the corresponding patterns, and the statistical performances of MEM to detect these patterns.

It is also worth mentioning that, although the ecological example chosen to illustrate the simulated results concerned the selection of spatial predictors at a small extent (500 × 200 m), the results would likely be the same for temporal or phylogenetic eigenvectors or if the study were conducted at a macroecological extent. Indeed, as it was shown by the simulations, the differences among the selection approaches arose from the control of the type I error rate, and from the number of selected spatial predictors that either aimed at minimising the spatial autocorrelation (SAC) in the residuals with a minimum number of variables (MIR approach) or at describing the SAC as accurately as possible with a slightly higher number of variables (FWD approach). Hence, the spatial, temporal, or phylogenetic nature of the autocorrelation does not matter; what does matter is whether there is autocorrelation, at what scale (i.e. at what focus, sensu Scheiner 2011), and the characteristics of the sampling design. For instance, an irregular spatial sampling will cause variations in inter-site distances, an irregular time-series will induce differences in sampling time intervals, and a phylogenetic tree with varying branch lengths will highlight that some species are separated from the rest of the species by a much longer evolutionary time. These irregularities will affect the average inter-site distance of the study in the same way, and will therefore influence the ability to detect a pattern (Legendre and Legendre 2012), be it spatial, temporal, or phylogenetic. However, this irregularity should be considered with respect to the geographical extent of the study, so that only the relative position and distance among sites matter and not the extent per se (e.g. the same results could be obtained for a 500 × 200 m or 1000 × 400 km study area if the relative sampling designs have similar characteristics with respect to the extent of the study). These considerations about the extent of the sampling design and the nature of the autocorrelation also hold for the low statistical power that we emphasised at fine scales; the

pattern can be spatial, but also temporal or phylogenetic, and the notion of 'fine' is relative to the sampling design or phylogeny used (it was of 25–50 m² in our example, but could be of several hundreds of square kilometres in a metacommunity analysis).

The low and null power obtained for fine scale positively autocorrelated patterns using the FWD and MIR approaches, respectively, indicate that ecologists must bear in mind that their pattern detection may be biased towards broad and medium scales. This difference of power across scales is likely to bias the interpretations of many studies. In evolutionary biology for instance, the general tendency of the phylogenetic signals detected in traits, using PVR, is therefore expected to indicate that most traits diverged a long time ago (broad and medium-scale phylogenetic structures), while the traits displaying no signal could have diverged more recently (undetected fine-scale patterns). Consequently, the specific (S) and phylogenetic (P) components of the trait variability (T = S + P, see Diniz-Filho et al. 2012) will often be artificially over-, or underestimated, respectively. Diniz-Filho et al. (2012) reviewed several selection methods in the scope of phylogenetic autocorrelation control and highlighted that the selection method of Griffith and Peres-Neto (2006; MIR approach) was the best performing approach to control for phylogenetic autocorrelation in residuals but that using a small number of eigenvectors may be insufficient to correctly describe the phylogenetic autocorrelation in some cases, mainly for fine-scale complex structures. Our results indeed indicate that, although it successfully removed the autocorrelation from the model, the MIR approach is likely to perform poorly at fine phylogenetic scales. This also questions the suitability of the Moran's *I* to detect fine-scale spatial structures. The FWD approach would be an alternative providing higher statistical power and a better R^2 estimation accuracy at fine scales. Moreover, it usually selects a slightly higher number of eigenvectors than the MIR approach, therefore ensuring a better characterisation of the phylogenetic signal (Diniz-Filho et al. 2012).

Summarising ten years of eigenvector selection practices

Blanchet et al.'s (2008) forward selection appeared to be the selection method yielding both the highest power and R^2 estimation accuracy while displaying correct type I error rates. However, the review revealed that most studies using this method did not specify whether a global test of significance had been conducted, hence possibly obtaining and interpreting spurious results.

Additionally, up to 32% of the studies reviewed used the AIC approach. The number of studies using this method has not decreased over the years (Fig. 2b), hence suggesting more biased studies to come. The review also highlighted a widespread lack of methodological specification regarding the definition of the **W** matrix used, and whether a variable selection was used or not.

Diniz-Filho et al. (2012) showed that selecting all eigenvectors significantly related to the response variable produced over-fitted models, and yet this approach was used in 5.3% of the studies considered here. Similarly, the classical forward selection was used in 5.6% of the reviewed studies in spite of its high type I error rate and model overfitting issues (Blanchet et al. 2008). All together, these 'bad practices' of selection concerned 38.9% of the published works reviewed in this study. Many studies may therefore have been subject to inflated type I error rates and potentially wrongly detected spatial structures, making the reliability of their results uncertain both quantitatively and in terms of ecological interpretations. These methodological biases are likely to have caused false positives and/or overfitted signals (mostly in the pure spatial fraction of VP), and these may impact meta-analyses (Soininen 2016, for a meta-analysis of the degree of spatial structure across different types of organisms and ecosystems). For future reliable meta-analyses and comparisons to be possible, unbiased methods of eigenvector selection well suited for the biological questions of interest will have to be adopted.

Recommendations: eigenvector selection and objective of the study

The results showed that both the FWD and the MIR methods were well adapted to select spatial predictors. The FWD approach returned more accurate R^2 estimates and had more power at fine scales than the MIR approach, but it also selected a slightly higher number of predictors. As both methods displayed a low type I error rate, the choice to use one or the other variable selection method will mostly depend on the purpose of the study and on the univariate or multivariate nature of the model. For a univariate response variable, if the only purpose is to control the spatial autocorrelation in the model residuals, the priority is to introduce as few predictors as possible in the model to avoid losing power; thus, the MIR approach should be preferred. If the response is multivariate, or if it is univariate but the purpose is to model as accurately as possible the multiscale spatial, temporal, or phylogenetic structures, then priority should be given to precision, and we would advocate Blanchet et al.'s (2008) forward selection. A tutorial is provided in Supplementary material Appendix 5 to help users select and code the most suitable selection method in R depending on their purpose and type of data. Additionally, the FWD approach may be a promising solution to the issue of phylogenetic eigenvector selection in PVR (Rohlf 2001, Freckleton et al. 2011, Diniz-Filho et al. 2012) for multivariate response data and when the purpose is to describe trait phylogenetic patterns precisely.

In this study, we disentangle unbiased and accurate from biased and underpowered selection methods by considering the relations between a response variable and the spatial predictors. Although we considered an additional set of environmental predictors in the real dataset, we did not investigate with simulations how the selection methods influenced the type I error rates and estimate accuracy of the VP fractions. When considering VP, previous studies showed that variable

selection procedures could lead to inflated type I error rates of the total and pure environmental fractions (Peres-Neto and Legendre 2010, Smith and Lundholm 2010). This can be explained by the fact that 'space' is not as accurately described with a subset of spatial predictors as with the complete set of positively autocorrelated predictors, resulting in an incomplete modelling of the spatial signal of the environmental component. However, a selection procedure is essential, as the power to detect the pure environmental fraction can strongly decrease when all MEM variables are used (Peres-Neto and Legendre 2010), and using all MEM variables overcorrects the spatial autocorrelation in the data (Griffith 2003). A solution to this problem would be to 1) apply a forward selection of the spatial predictors (Blanchet et al. 2008), and 2) introduce them in a VP in which the fractions are tested and estimated on the basis of an R^2 corrected by a null model constrained with Moran spectral randomisation (Wagner and Dray 2015) to maintain the spatial structures of the data (Clappe et al. pers. comm.).

The present study focused on the selection of an optimal subset of MEM within a given **W** matrix. However, **W** matrices can be built in many ways and can lead to contrasting results for irregular and clustered sampling designs. The selection of a **W** matrix is therefore a key step of all eigenvector-based methods (Dray et al. 2006), and yet this step appeared in our review as seldom explicitly tackled. Hence, a great proportion of studies are likely to have obtained underpowered results, as the authors used the original PCNM, only used distance-based MEM (while the latter may be an unsuitable choice for irregular or clustered sampling designs, Dray et al. 2006), did not compare different **W** matrices, and, most often did not even specify the **W** matrix that was used. The next step will therefore be to set up an accurate and unbiased optimisation procedure allowing users to test diverse types of **W** matrices and select the most adapted to their dataset.

Acknowledgements – We are grateful to three anonymous reviewers and to the anonymous subject editor for their constructive suggestions and discussions.
Funding – This research was supported by the Belgian National Fund for Scientific Research (F.R.S.-FNRS) to DB. The computational resources have been provided by the Shared ICT Services Centre, Univ. Libre de Bruxelles.

References

Anderson, M. J. and Legendre, P. 1999. An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. – J. Stat. Comput. Simul. 62: 271–303.
Baho, D. L. et al. 2015. Assessing temporal scales and patterns in time series: comparing methods based on redundancy analysis. – Ecol. Complex. 22: 162–168.
Barbier, N. et al. 2008. Spatial decoupling of facilitation and competition at the origin of gapped vegetation patterns. – Ecology 89: 1521–1531.
Bauman, D. et al. 2016. Multiscale assemblage of an ectomycorrhizal fungal community: the influence of host functional traits and soil properties in a 10-ha miombo forest. – FEMS Microbiol. Ecol. 92: fiw151.
Bini, L. M. et al. 2009. Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. – Ecography 32: 193–204.
Bivand, R. 2006. spdep: spatial dependence: weighting schemes, statistics and models. – R package ver. 0. 6-13.
Blanchet, F. G. et al. 2008. Forward selection of explanatory variables. – Ecology 89: 2623–2632.
Borcard, D. and Legendre, P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. – Ecol. Model. 153: 51–68.
Borcard, D. et al. 1992. Partialling out the spatial component of ecological variation. – Ecology 73: 1045–1055.
Borcard, D. et al. 2011. Numerical ecology with R. – Springer.
Corkeron, P. J. et al. 2011. Spatial models of sparse data to inform cetacean conservation planning: an example from Oman. – Endanger. Species Res. 15: 39–52.
Cormack, R. M. and Ord, J. K. 1979. Spatial and temporal analysis in ecology. – Int. Congr. Ecol. 1978.
Declerck, S. A. J. et al. 2011. Scale dependency of processes structuring metacommunities of cladocerans in temporary pools of High-Andes wetlands. – Ecography 34: 296–305.
Diniz-Filho, J. A. F. and Bini, L. M. 2005. Modelling geographical patterns in species richness using eigenvector-based spatial filters. – Global Ecol. Biogeogr. 14: 177–185.
Diniz-Filho, J. A. F. et al. 1998. An eigenvector method for estimating phylogenetic inertia. – Evolution 52: 1247–1262.
Diniz-Filho, J. A. F. et al. 2012. On the selection of phylogenetic eigenvectors for ecological analyses. – Ecography 35: 239–249.
Dormann, C. F. et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. – Ecography 30: 609–628.
Dray, S. et al. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). – Ecol. Model. 196: 483–493.
Ezekiel, M. 1929. The application of the theory of error to multiple and curvilinear correlation. – J. Am. Stat. Assoc. 24: 99–104.
Freckleton, R. P. et al. 2011. Comparative methods as a statistical fix: the dangers of ignoring an evolutionary model. – Am. Nat. 178: E10–E17.
Godinez-Dominguez, E. and Freire, J. 2003. Information-theoretic approach for selection of spatial and temporal models of community organization. – Mar. Ecol. Prog. Ser. 253: 17–24.
Griffith, D. A. 1996. Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data. – Can. Geogr. 40: 351–367.
Griffith, D. 2003. Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization. – Springer.
Griffith, D. A. and Peres-Neto, P. R. 2006. Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. – Ecology 87: 2603–2613.
Hurvich, C. M. and Tsai, C.-L. 1989. Regression and time series model selection in small samples. – Biometrika 76: 297–307.
Jombart, T. et al. 2009. Finding essential scales of spatial variation in ecological data: a multivariate approach. – Ecography 32: 161–168.

Layeghifard, M. et al. 2015. Spatial and species compositional networks for inferring connectivity patterns in ecological communities. – Global Ecol. Biogeogr. 24: 718–727.
Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? – Ecology 74: 1659–1673.
Legendre, P. and Fortin, M. J. 1989. Spatial pattern and ecological analysis. – Vegetatio 80: 107–138.
Legendre, P. and Legendre, L. 2012. Numerical ecology. – Elsevier.
Manly, B. F. J. 2007. Randomization, bootstrap and Monte Carlo methods in biology. – Chapman and Hall/CRC.
McIntire, E. J. B. and Fajardo, A. 2009. Beyond description: the active and effective way to infer processes from spatial patterns. – Ecology 90: 46–56.
Muledi, J. I. et al. 2017. Fine-scale habitats influence tree species assemblage in a miombo forest. – J. Plant Ecol. 10: 958–969.
Oksanen, J. et al. 2017. Package "vegan". – Community ecology package ver. 2.4-3.
Pech, N. and Laloë, F. 1997. Use of principal component analysis with instrumental variables (PCAIV) to analyse fisheries catch data. – ICES J. Mar. Sci. 54: 32–47.
Peres-Neto, P. R. and Legendre, P. 2010. Estimating and controlling for spatial structure in the study of ecological communities. – Global Ecol. Biogeogr. 19: 174–184.
Pinto, S. M. and MacDougall, A. S. 2010. Dispersal limitation and environmental structure interact to restrict the occupation of optimal habitat. – Am. Nat. 175: 675–686.
Rohlf, F. J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. – Evolution 55: 2143–2160.
Scheiner, S. M. 2011. Musings on the Acropolis: terminology for biogeography. – Front. Biogeogr. 3: 62–70.
Siesa, M. E. et al. 2011. Spatial autocorrelation and the analysis of invasion processes from distribution data: a study with the crayfish *Procambarus clarkii*. – Biol. Invasions 13: 2147–2160.
Smith, T. W. and Lundholm, J. T. 2010. Variation partitioning as a tool to distinguish between niche and neutral processes. – Ecography 33: 648–655.
Soininen, J. 2016. Spatial structure in ecological communities – a quantitative analysis. – Oikos 125: 160–166.
Tiefelsdorf, M. and Griffith, D. A. 2007. Semiparametric filtering of spatial autocorrelation: the eigenvector approach. – Environ. Plan. A 39: 1193–1222.
Velázquez, E. et al. 2016. An evaluation of the state of spatial point pattern analysis in ecology. – Ecography 39: 1042–1055.
Vleminckx, J. et al. 2017. The influence of spatially structured soil properties on tree community assemblages at a landscape scale in the tropical forests of southern Cameroon. – J. Ecol. 105: 354–366.
Wagner, H. H. and Dray, S. 2015. Generating spatially constrained null models for irregularly spaced data using Moran spectral randomization methods. – Methods Ecol. Evol. 6: 1169–1178.
Westfall, P. H. et al. 1998. Forward selection error control in the analysis of supersaturated designs. – Stat. Sin. 8: 101–117.

Supplementary material (Appendix ECOG-03380 at < www.ecography.org/appendix/ecog-03380 >). Appendix 1–5, Fig. A1, Table A1, A2, A3, and A4.