

RESEARCH  
PAPER



# Modelling geographical patterns in species richness using eigenvector-based spatial filters

José Alexandre Felizola Diniz-Filho<sup>1,2\*</sup> and Luis Mauricio Bini<sup>1</sup>

<sup>1</sup>Departamento de Biologia Geral, ICB, Universidade Federal de Goiás, CP 131, 74.001–970, Goiânia, GO, Brazil; and <sup>2</sup>Departamento de Biologia/MCAS-VPG, Universidade Católica de Goiás, Goiânia, GO, Brazil

## ABSTRACT

**Aim** To test the mechanisms driving bird species richness at broad spatial scales using eigenvector-based spatial filtering.

**Location** South America.

**Methods** An eigenvector-based spatial filtering was applied to evaluate spatial patterns in South American bird species richness, taking into account spatial autocorrelation in the data. The method consists of using the geographical coordinates of a region, based on eigenanalyses of geographical distances, to establish a set of spatial filters (eigenvectors) expressing the spatial structure of the region at different spatial scales. These filters can then be used as predictors in multiple and partial regression analyses, taking into account spatial autocorrelation. Autocorrelation in filters and in the regression residuals can be used as stopping rules to define which filters will be used in the analyses.

**Results** Environmental component alone explained 8% of variation in richness, whereas 77% of the variation could be attributed to an interaction between environment and geography expressed by the filters (which include mainly broad-scale climatic factors). Regression coefficients of environmental component were highest for AET. These results were unbiased by short-scale spatial autocorrelation. Also, there was a significant interaction between topographic heterogeneity and minimum temperature.

**Conclusion** Eigenvector-based spatial filtering is a simple and suitable statistical protocol that can be used to analyse patterns in species richness taking into account spatial autocorrelation at different spatial scales. The results for South American birds are consistent with the climatic hypothesis, in general, and energy hypothesis, in particular. Habitat heterogeneity also has a significant effect on variation in species richness in warm tropical regions.

## Keywords

Eigenanalysis, macroecology, spatial autocorrelation, spatial filtering, species richness, South American birds.

\*Correspondence: José Alexandre Felizola Diniz-Filho, Departamento de Biologia Geral, ICB, Universidade Federal de Goiás, CP 131, 74.001–970, Goiânia, GO, Brazil.  
E-mail: diniz@icb.ufg.br

## INTRODUCTION

Tests of the mechanisms driving species diversity have usually been made using multiple regression and related statistical approaches, such as path analysis, in which species richness is regressed against sets of environmental variables or other surrogates of ecological and historical processes (Hawkins & Porter, 2003a,b; Hawkins *et al.*, 2003a; Diniz-Filho *et al.*, 2004; and references therein). However, the lack of independence between pairs of observations across geographical space (spatial autocorrelation) is commonly found in ecological data, in such a way that more complex strategies for data analyses are required (Legendre, 1993).

As recently pointed out by Griffith (2003), spatial autocorrelation can be interpreted in two slightly different, not mutually exclusive, ways. First, methods of spatial autocorrelation assessment, especially the Moran's *I* correlograms, can be used to explore spatial patterns in data (e.g. gradients, trends, mosaics or patches) or to verify the presence of latent predictors, which are also spatially autocorrelated, in a multiple regression, at different spatial scales (the map pattern). For example, Hawkins & Porter (2003a) recently showed that, after fitting annual actual evapotranspiration (AET) to richness data of North American birds, a significant amount of autocorrelation remained in the regression residuals. After mapping these residuals, they realized that historical effects

accounted for the remaining spatial patterns. When age of the cells after the last deglaciation was included in the model, residual autocorrelation almost disappeared.

Second, spatial autocorrelation can be interpreted as redundant information on data, because of the 'contagious' effect among adjacent spatial units. Thus, degrees of freedom are overestimated because spatial units are considered pseudo-replication of the phenomenon (e.g. see Dutilleul, 1993). This interpretation leads to the well known problem of statistical inference under spatial autocorrelation, i.e. if there is autocorrelation in multiple regression residuals, due to unknown factors or intrinsic dependence in data, the standard errors of regression coefficients are usually underestimated and, consequently, Type I errors may be strongly inflated. Diniz-Filho *et al.* (2003) discussed the origins of spatial autocorrelation in grid data usually used in geographical ecology. It may be due to unknown factors explaining richness at a given spatial scale, as found by Hawkins & Porter (2003a). However, and perhaps more importantly, short-distance autocorrelation may be due to simple dependence caused by the way data is collected (for example, richness is frequently derived from the overlap of extents of occurrence, and once ranges are larger than cell size, adjacent cells usually have rather similar species composition and, consequently, similar species richness). In this case, introducing other predictors will rarely be able to overcome the autocorrelation effects, in such a way that more complex generalized least squares (GLS) or autoregressive models that incorporate the spatial patterns into model structure or regression residuals must be used, providing in principle unbiased regression coefficients. These different forms of spatial models are complex, computationally intensive and sometimes difficult to implement (Haining, 1990; Griffith, 2003; but see Lichstein *et al.*, 2002). Moreover, in some situations, the application of autoregressive methods does not ensure that all autocorrelation in data is taken into account (see Tognelli & Kelt, 2004 for a recent example).

More recently, Borcard & Legendre (2002) and Griffith (2003) proposed that eigenvector-based spatial filters could be a more simple solution to the autocorrelation problem. The basic idea is to extract eigenvectors of a connectivity (or distance, see below) matrix among spatial units (e.g. cells in a grid), and use these eigenvectors, which describe the spatial structure of the region under study at different scales, as additional predictors of the response variable. This way, any remaining spatial structures in regression residuals would be taken into account, and so these models would not be affected by the problem of spatial autocorrelation. Griffith (2003) showed that the advantage of this approach is the use of a multiple regression framework, which is familiar to most ecologists and evolutionary biologists. The approach by Borcard & Legendre (2002) is based on a principal coordinate analysis of the truncated geographical distance matrix among spatial units (see also Diniz-Filho *et al.*, 1998; for a similar application to phylogenetic comparative analyses), whereas Griffith (2003) used transformations of a connectivity (binary) matrix among spatial units.

Our goal here is to discuss the eigenvector-based spatial filtering and analyse geographical patterns in species richness of South American birds and correlate them with climatic and environmental

variation across the continent. The spatial filters were used in a multiple-partial regression designed to partition between purely environmental-climatic and geographical effects, and the results were compared with more commonly used partial regression based on polynomial expansions (i.e. incorporating a cubic trend surface into the model). We also compared regression coefficients of climatic-environmental factors as predictors of species richness, obtained after incorporating spatial filters in the multiple regression, with the coefficients of the factors in a generalized least squares (GLS) approach that incorporates spatial structure into model residuals. In this way, we explore many of the analytical frameworks available for ecologists in dealing with the complex issue of relating species richness and environmental factors at broad spatial scales.

## METHODS

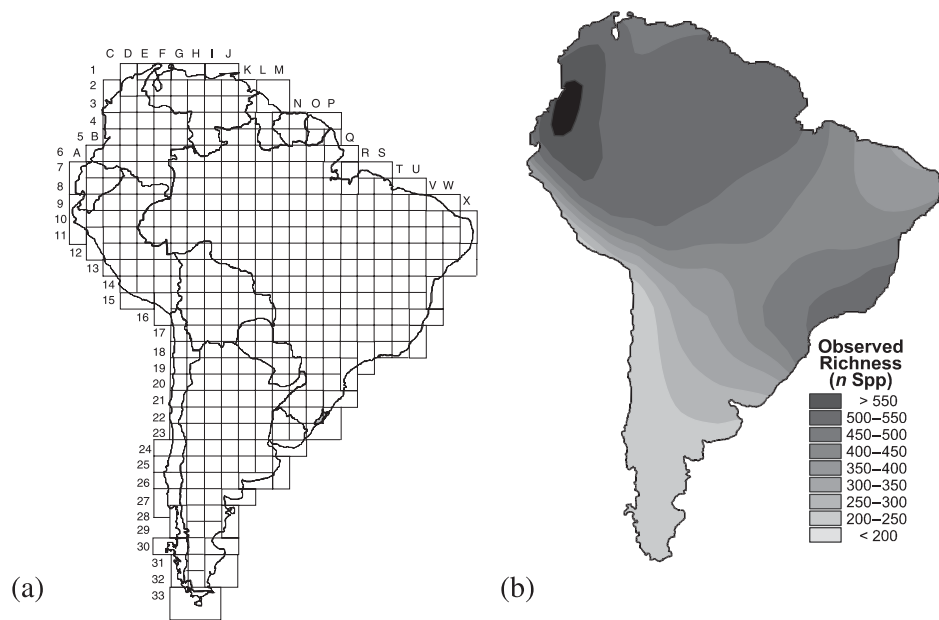
### Species richness and environmental data

South America was divided into 374 equal area cells 220 km × 220 km (2° × 2° at the equator). The geographical ranges of the 2894 species of terrestrial birds present on the continent were redrawn over this grid, and the presence of each species in each cell was recorded (see Hawkins *et al.*, 2003b; for details). Coastal cells were combined to obtain areas approximately equal in size to inland cells. Thus, area was held as constant as possible and was not included explicitly in the analysis (Fig. 1a). Due to computational limitations of SPACEMAKER 2 (see below), we randomly deleted 14 cells and all further analyses were based on a grid with 360 cells.

Five climatic variables that have been shown to be associated with broad-scale richness gradients were compiled from various sources: (1) potential evapotranspiration (PET); (2) actual evapotranspiration (AET); (3) mean daily temperature in the coldest month; (4) annual mean temperature; and (5) annual rainfall (see Diniz-Filho *et al.*, 2003; Hawkins *et al.*, 2003a; for details). We also included range in elevation (Pergamon World Atlas, 1968), estimated to the nearest 50 m, to estimate topographic variability. Further, we incorporated an interaction term between topographic variability and minimum temperature to capture the idea that topographic variability is only important in warm environments, creating strong environmental effects at the mesoscale (see Rahbek & Graves, 2001 for a similar approach using the interaction between latitude and topographic variability). This variable was called climatic heterogeneity (Diniz-Filho *et al.*, 2004). Finally, we counted the number of habitats in each cell using remotely sensed AVHRR (Advanced Very High Resolution Radiometer) land cover data (Hawkins & Porter, 2003b; Diniz-Filho *et al.*, 2004), creating an explicit measurement of habitat variety.

### Extracting and interpreting spatial filters

Geographical coordinates (latitude and longitude) of each cell covering South America were used to construct a pairwise matrix of geographical distances among cells (*G*), which was truncated



**Figure 1** (a) Distribution of the 360 cells that were used to analyse the spatial variation of bird species richness in South America; and (b) spatial variation in the richness of breeding terrestrial bird species in South America.

(see below) at a distance of 1000 km (truncation distance,  $D$ ). This truncation distance is important because it gives more weight to short-distance effects, after the filtering process. This initial truncation value was based on our previous analyses of spatial autocorrelation in residuals of species richness after fitting environmental data (Hawkins *et al.*, 2003a). Thus, distances larger than 1000 km were replaced by  $4 \times D$  ( $= 4000$  km), whereas distances  $< 1000$  km were kept as they were calculated (see Fig. 1 of Borcard & Legendre, 2002). The truncated  $G$  matrix was then submitted to a principal coordinate analysis (PCORD) (Legendre & Legendre, 1998), which consisted of performing an eigenanalysis of the double-centred  $G$  matrix.

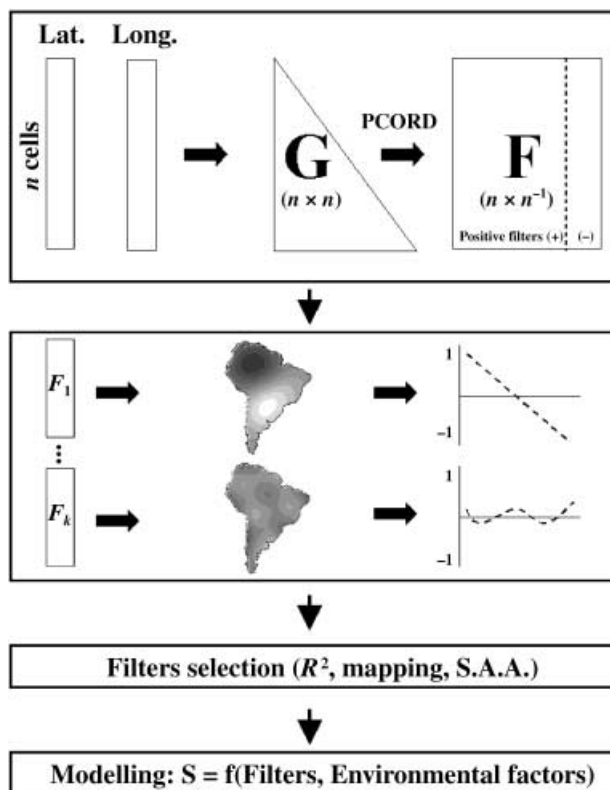
The eigenvectors associated with positive eigenvalues of the double-centred  $G$  matrix represent the spatial relationship among cells covering South America, at different spatial scales. The first eigenvectors represent broad-scale variation, whereas eigenvectors derived from small eigenvalues represent fine-scale variation. These vectors are then new orthogonal variables (called filters by Griffith, 2003) that capture, at different scales, the geometry of the grid covering the region under analyses (i.e. South America) and that can be incorporated into multiple regression approach in different ways, taking into account spatial autocorrelation and allowing an unbiased estimation of regression parameters. Figure 2 shows a schematic view of how filters are calculated and interpreted [see also Borcard & Legendre (2002) for a similar representation of this technique].

The next step of the analytical protocol discussed here includes the selection of the eigenvectors that should enter as predictors in the model. Borcard & Legendre (2002) suggested testing the significance of all the partial regression coefficients and retaining only the eigenvectors that are significant. Griffith (2003) showed that using all eigenvectors in the analyses might 'overcorrect' for

spatial autocorrelation, and proposed some strategies to choose some of the vectors. These strategies included: (i) maximization of the regression multiple correlation coefficient ( $R^2$ ); (ii) minimization of residual spatial autocorrelation; and (iii) a significant correlation between response variable and each selected eigenvector (see Diniz-Filho *et al.*, 1998; for a different approach in comparative analyses, based on broken-stick null distribution of eigenvalues).

We mapped the spatial filters and analysed spatial patterns in both filters and in the residuals from a multiple regression fitting species richness successively to these filters, using spatial correlograms defined using Moran's  $I$  coefficients calculated for 15 distance classes (Legendre & Legendre, 1998; Diniz-Filho *et al.*, 2003). This way, only filters that in fact contain important parts of the geometry of the continent will be used in the analysis, avoiding excess of predictors in the multiple regression (see Rohlf, 2001).

In summary, filters can be considered as different and independent propositions of how cells are geographically related or connected to each other, expressed as new variables derived from geographical distances and indicating the spatial relationships among cells. Mapping and running spatial autocorrelation analysis on filters and regression residuals can help to interpret which part of spatial structure among cells are captured by each filter. Besides, including filters in environmental models minimizes the undesirable effects of subadjacent spatial structures that were not captured by environmental factors (see below). Suppose that the response variable (species richness in our case) is spatially patterned, for example, with low and high values alternating in space, forming patches (e.g. in a fragmented landscape). In this case, it is highly probable that one of the filters will account for this complex spatial pattern if no environmental factor does this. Thus, this patchy structure will not be present in the residuals



**Figure 2** General description of the method used in this study. The first panel shows spatial filters calculation based on geographical coordinates of the  $n$  cells, which are transformed in a Euclidean distance matrix ( $G$ ). Borcard & Legendre (2002) truncated this matrix at a given value ( $D_T$ ) and replaced all distance higher than this value by  $4 \times D_T$ . After establishing  $G$ , a principal coordinate analysis (PCORD) is computed. Eigenvectors (filters, matrix  $F$ ) associated with  $k$  positive eigenvalues can then be analysed by mapping and by spatial autocorrelation analysis (SAA) (second panel), and results of this interpretation permit decisions on which filters (filters selection) should be entered into different types of statistical analysis (modelling).

and the regression model will not be biased by spatial autocorrelation. Indeed, according to Borcard & Legendre (2002), filters are spatial descriptors of the response variable and can be incorporated into analytical frameworks in different ways depending on the context (see below).

### Incorporating spatial filters into multiple regression and comparison with other methods

Partial regression analyses have been used to establish the amount of variation in richness that can be explained independently by environmental variables and geography (Legendre & Legendre, 1998; Balmford *et al.*, 2001; van Rensburg *et al.*, 2002; Ruggiero & Kitzberger, 2004). In these analyses, spatial structure has been expressed by latitude and longitude and their polynomial expansions, usually up to the third order (i.e. a cubic trend surface analysis). However, these polynomial expansions usually take into account only broad-scale effects, and short-distance

autocorrelations remain, in such a way that even the estimation of spatial trends can be biased (see Haining, 1990 for methods to correct this problem). But this problem should also be avoided if 'space' or 'geography' is expressed by a given number of spatial filters, since they express spatial structure at different spatial scales. In this paper, we compared the partial regressions obtained using polynomial expansions and filters as expression of 'geography', both in terms of their relative components and in terms of their residual autocorrelation.

Finally, we analysed the partial regression coefficients of climatic–environmental variables after including the important spatial filters into the regression model, hoping that this allows a correct estimate of the effects of each variable, unbiased by spatial autocorrelation in residuals (as frequently occur with ordinary least square regression, or OLS). We compared these regression coefficients with those obtained from different generalized least squares (GLS) estimates that incorporate spatial structure in the model residuals, assuming exponential, spherical or Gaussian relationship between error term and geographical distances (see Selmi & Boulinier, 2001; Diniz-Filho *et al.*, 2003; Gaston & Evans, 2004). Model fit and their performance in relation to standard OLS of species richness against climatic–environmental variables was tested using Akaike criterion.

Principal coordinate analysis of  $G$ -matrix was performed using SPACEMAKER 2 (<http://www.fas.umontreal.ca/biol/legendre/>), and spatial autocorrelation analyses were performed in SAAP 4.3 (Wartenberg, 1989). Partial and multiple regressions were performed in SYSTAT 10 (SYSTAT Inc. 2000), while GLS estimates were obtained in Proc Mixed routine of SAS (Littell *et al.*, 1996).

## RESULTS

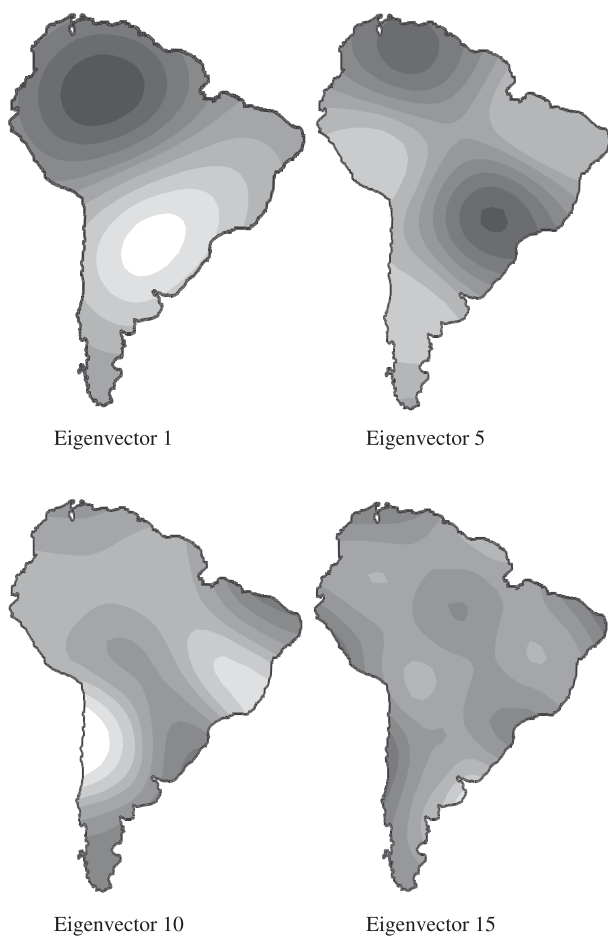
Geographic variation of bird species richness in South America was strongly autocorrelated, showing a monotonic decrease of Moran's  $I$  coefficient across distances. This can be clearly associated with broad-scale variation in richness across the continent (Fig. 1b), in which distant areas have very distinct richness values (e.g. north–south extremes, and east–west clines due to the Andes). The climatic–environmental multiple regression model, using all seven climatic and environmental predictors, explained 85% of variation in bird species richness, with larger coefficients for AET and for the interaction between topographic variability and minimum temperature (Table 1). However, as previously found using a smaller set of predictors (Hawkins *et al.*, 2003a), the residuals of this model still had short-distance positive autocorrelation (Fig. 4). This indicates that broad scale spatial patterns in species richness were fully explained by environmental factors, but that short-distance variation still remains unexplained. So, other analytical strategies are necessary to take into account these spatial structures and allow unbiased estimation of regression parameters.

Out of 359 eigenvalues extracted from the  $G$  matrix, 175 were positive and the eigenvectors (filters) associated with these positive eigenvalues were extracted and analysed. The scree plot (Fig. 5) showed that the eigenvalues tended to stabilize after

Riqueza

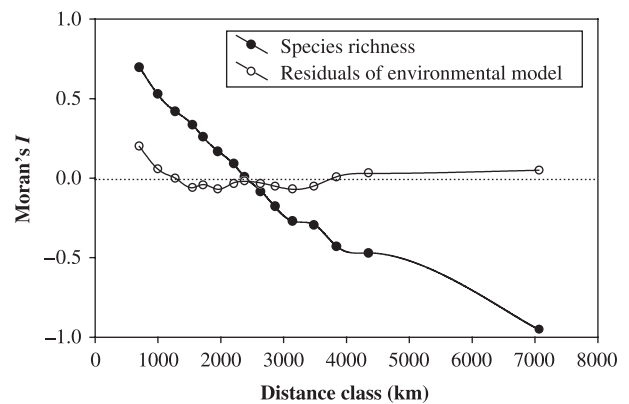
**Table 1** Partial regression coefficients of the multiple regression models ( $b$ ),  $t$  statistics and associated  $P$ -values for bird species richness regressed against environmental factors in South America, according to ordinary least square (OLS), generalized least-squares (GLS, assuming a spherical model to describe spatial patterns in residuals) and according to the filters approach. In the Filter model, spatial structure was incorporated into the multiple regression by adding 15 eigenvectors (not shown)

Variables	OLS			GLS			Filters		
	$b$	$t$	$P$	$b$	$t$	$P$	$b$	$t$	$P$
Annual rainfall (1)	0.021	3.48	0.001	0.003	0.72	0.471	0.009	1.82	0.070
Annual mean temperature	0.143	1.34	0.181	0.372	5.66	0.000	0.467	5.18	0.000
Range in elevation	0.011	3.04	0.003	0.005	1.41	0.161	-0.003	-0.63	0.527
Potential evapotranspiration	0.064	3.06	0.002	0.017	0.88	0.380	0.011	0.46	0.646
Actual evapotranspiration	0.194	10.52	0.000	0.056	3.44	0.001	0.139	8.34	0.000
Habitat variety	-2.599	-1.00	0.319	-0.282	-0.14	0.886	4.653	1.86	0.063
Minimum temperature	0.271	0.22	0.830	-4.204	-4.50	0.000	-4.706	-3.64	0.000
Interaction (range in elevation * minimum temperature)	0.001	4.54	0.000	0.001	4.74	0.000	0.001	5.51	0.000



**Figure 3** Patterns depicted by 5 spatial filters showing the relationship among cells at distinct spatial scales. Increasing darkness of shading indicates the largest numerical values of the eigenvectors.

c. 15–20 extracted axes. Moran's  $I$  coefficients were used to evaluate spatial patterns in filters as an effective measure of their spatial structure (Fig. 6a). The 15 first filters had relatively high Moran's  $I$  coefficients in the first distance class (i.e.  $I > 0.1$ ), and their

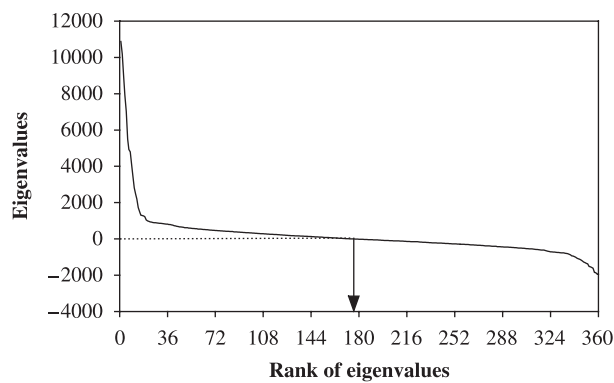


**Figure 4** Moran's  $I$  correlogram for bird species richness and the residuals of multiple regression model including the seven environmental predictors.

spatial complexity and variable shape of correlograms showed how they reflected different spatial structures at different spatial scales (see Fig. 3 for some examples of maps of spatial filters and Fig. 6b, for the spatial correlograms of the first 15 filters). High levels of spatial autocorrelation in filters, as in the first eigenvector (Fig. 6a), tended to be portrayed by a map pattern containing two major clusters of similar values (one cluster with relatively high values and the other with relatively low values; see Griffith, 2003). These clusters tended to be positioned at opposite ends of a map, suggesting some type of gradient (Fig. 3a). As the degree of positive spatial autocorrelation decreased, the map pattern became more fragmented (see Fig. 3d for the map pattern and Fig. 6a for the spatial correlograms). In short, the first filters possess strong clinal patterns, with high positive and negative Moran's  $I$  in the first and last distance classes, respectively. The filters with smaller eigenvalues have successively smaller Moran's  $I$  in the first distance class and start to create an 'oscillatory' pattern across the space.

The coefficient of determination of the regression models relating species richness and filters, successively added to the



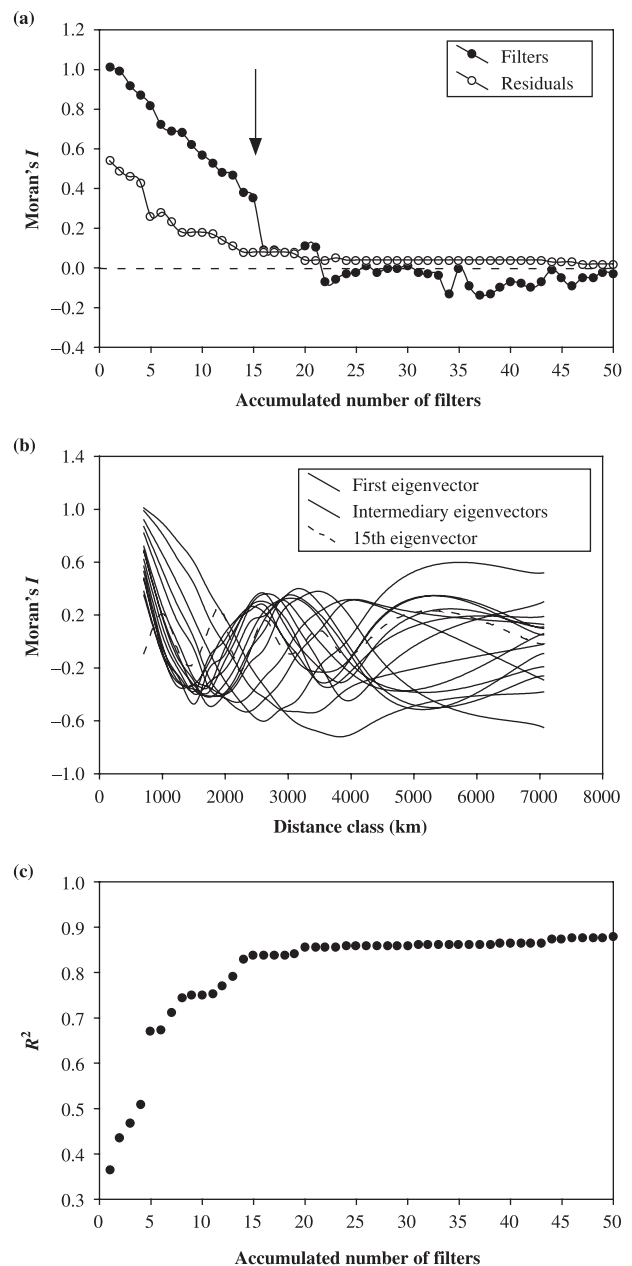


**Figure 5** Scree plot of the eigenvalues resulting from the principal coordinate analysis applied to the truncated distance matrix among cells covering South America.

model, increased fast in the first 3 or 4 axes, and started to stabilize around 10–15 filters (Fig. 6c). These 15 filters together explained around 84% of variation in species richness, and Moran's  $I$  coefficients of residuals indicated that spatial structures disappear when all these filters were added as predictors in the multiple regression (Fig. 6a). This way, we confirmed their ability to take into account patterns in richness at different spatial scales.

Thus, it is possible to assume, based on all these previous analyses, that the first 15 filters contain important and complex spatial patterns of the South American continent, and that fitting these filters to species richness is sufficient to remove autocorrelation structure from the residuals. Since these filters are sufficient to take into account most of spatial structure in species richness at different spatial scales, they can be used safely in the multiple regression framework, together with other environmental factors, to avoid the undesirable effects of spatial autocorrelation in hypothesis testing and providing unbiased regression coefficients.

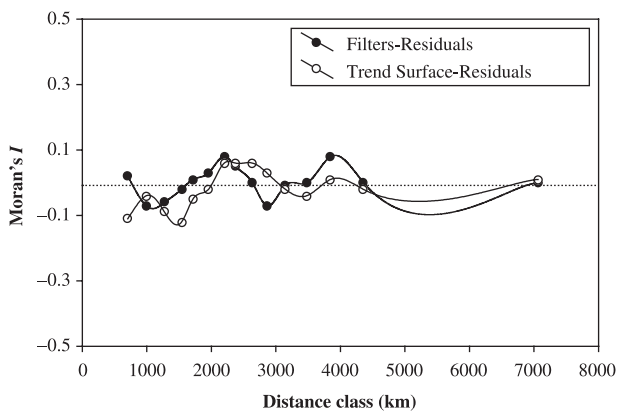
The 15 filters can be used as an expression of 'space' or 'geography' in the framework of variation partitioning. Adding both filters and the environmental variables as predictors in the multiple regression furnished an  $R^2$  of 0.917 and, after partitioning the variation in the different components, the following results were obtained: 8% of the variation in species richness can be explained by environment alone, whereas 6.7% can be explained by spatial structures alone. However, 77% of the variation in richness refers to the interaction between environment and spatial structure, especially broad-scale ones. These results are very similar to the OLS partial regression using a third-order polynomial expansion of latitude and longitude as expression of 'space', in which 10.9% of the variation in species richness can be explained by environment only, whereas 4.9% can be explained strictly by spatial structures. The spatially structured environment effect (interaction term) explained 74.1% of variation in species richness. However, using filters provides a slightly better reduction of spatial autocorrelation in the residuals at short distances, up to the fourth distance class. Moran's  $I$  in the first distance class reduced from 0.20 in the environmental OLS model to  $-0.11$ , when using



**Figure 6** (a) Moran's  $I$  coefficients of the first 50 filters estimated at the first distance class. The same is shown for the residuals of multiple regression models after successively including filters; (b) complete spatial correlograms for the first 15 filters (scores of principal coordinate analysis or eigenvectors); (c) coefficient of determination ( $R^2$ ) as a function of the number of filters entering in the multiple regression model. All three figures allow interpretation of the 15 first filters as important descriptors of spatial structure in data.

cubic polynomial expansions in the model, and to 0.02 when using filters (Fig. 7).

Partial regression coefficients of the environmental variables from the model including 15 spatial filters allowed an estimation of which predictors were more significantly associated with bird species richness, independently of spatial structures at distinct scales (Table 1). When compared to the OLS model, which does



**Figure 7** Moran's  $I$  coefficients of the residuals of multiple regression models after including filters and environmental data. The same is shown for the residuals of multiple (polynomial) regression.

not incorporate spatial structure, the best predictors (AET and climatic heterogeneity) remained highly significant when filters were added to the model. Three variables (annual rainfall, range in elevation and PET) that were highly significant in the original OLS model, became unimportant after adding filters, whereas habitat variety and both minimum and annual temperature tended to be significant in the spatial model.

The comparison between these regression coefficients and those obtained by different GLS models may also reveal how filters work. The Akaike criterion shows that, out of the three models tested, the spherical provides the best fit ( $AIC = -1809.32$ ) although all are considered better than the non-spatial OLS model ( $AIC = -1973.66$ ). Regression coefficients of GLS are similar to those obtained when filters are added into the regression model, especially because the effects of both annual and minimum temperature are also significant in the best GLS model, whereas rainfall and habitat variety remain not significant. On the other hand, the main difference between GLS models and the filter model is that AET is not the predictor with high coefficient in GLS, although it is still significant and relatively high when compared with annual and minimum temperature and with the coefficient for the interaction term between minimum temperature and range in elevation.

## DISCUSSION

Grid-based richness data sets are always spatially autocorrelated and, although fitting environmental models usually take into account broad-scale patterns, the remaining short-distance autocorrelation structure may be enough to create bias in regression coefficients and eventually change interpretations of the ecological mechanisms associated with richness (Diniz-Filho *et al.*, 2003). In this paper, we presented how spatial filtering, as proposed by Borcard & Legendre (2002) and Griffith (2003), can solve this problem by creating spatial variables and incorporating them into linear models. This method is also much simpler, both in computation and implementation, than more sophisticated

GLS and autoregression models (Lichstein *et al.*, 2002; see Griffith, 2003).

One first problem to the application of filtering is how to define which eigenvectors extracted from a double-centred  $G$ -matrix (after excluding those associated with negative eigenvalues) should be used as filters. In this paper, we followed Griffith (2003) and initially evaluated both spatial autocorrelation in filters and regression residuals after fitting these filters to richness data. These analyses can indicate which eigenvectors contain significant spatial patterns and at which scale they are effective in taking into account most of the spatially structured variation in richness, as indicated by the absence of spatial autocorrelation in the residuals. At the same time, the regression  $R^2$  can indicate stabilization of the influence of spatial patterns on richness. These criteria, if interpreted in isolation or together, can furnish stopping rules to introduce new eigenvectors into multiple regression and, consequently, help avoid overcorrection of spatial autocorrelation in data (Griffith, 2003).

Borcard & Legendre (2002) pointed out some aspects of the extension of their method to data sampled across a surface (bi-dimensional spatial data). Basically, they indicated that the properties of the spatial variables, in the bi-dimensional case (e.g. lattice), were similar to those detected in the one-dimensional case. In a recent contribution, Borcard *et al.* (2004) also applied the method to study the effects of environmental factors on phytoplankton biomass data collected over a bi-dimensional sampling grid. They showed that the method was useful to indicate how the influence of environmental factors on phytoplankton biomass varied across spatial scales. Probably, further investigations may be required on the mathematical and geometrical properties of filters calculated in bi-dimensional for both regular and irregular sampling grids (but see Griffith, 2003; for a detailed evaluation of a very similar method).

Using spatial filters in partial regression may be similar to the more conventional approach based on geographical coordinates and their polynomial expansions (quadratic, cubic, and so on) (e.g. Lobo *et al.*, 2002; Balmford *et al.*, 2001; van Rensburg *et al.*, 2002; Chown *et al.*, 2003; Ruggiero & Kitzberger, 2004). However, it is important to note that, since these partial regression analyses are based on trend surfaces, only broad-scale spatial structures are taken into account, and more local spatial autocorrelation effects are still not considered, eventually creating biased analyses or generating autocorrelated residuals at short distance classes (see Haining, 1990). Indeed, we observed in our analyses that, although adding cubic trends to model reduced first-order autocorrelation in residuals from 0.20 (in non-spatial regression) to 0.11, spatial filtering was more effective in taking these short scale structures into account. Besides, an important problem with polynomial regression is that individual terms in the regression model are highly correlated, which creates multicollinearity problems and difficulties for the analyses of spatial structures at different scales (Borcard & Legendre, 2002). In short, the main advantage of the approach shown here is that the eigenvector-based spatial filters are designed to capture spatial structures at different scales (see Figs 3 and 6b), in such a way that significance tests of regression coefficients are not biased by autocorrelation at these scales.

On the other hand, the interpretation of partial regression coefficients of environmental and climatic predictors after adding filters should be made carefully. As recently pointed out by Hawkins & Diniz-Filho (2004), the introduction of geographical space as a 'variable' or as a 'factor' in a model (especially latitudinal coordinates only) does not improve ecological interpretation, except in the sense that regression parameters should not be biased by local autocorrelation effects. Of course, it is important to verify at which scale spatial autocorrelation is being taken into account, but if filters capture better the broad-scale spatial autocorrelation (usually those with largest eigenvalues), they may covary with broad-scale climatic factors and the relative importance of regression coefficients may be shifted towards more local and meso-scale factors. On the other hand, a significant fraction of variation attributed to the spatial filters only suggests that some ecological process has generated the identified spatial pattern, even though the causal process (as indicated by some relevant ecological predictor) are unknown and, in this way, not included in the model (Borcard *et al.*, 2004).

Our analysis using spatial filters, although corrected efficiently for autocorrelation in residuals, did not show a scale-shift in regression coefficients towards more local scales (see Lennon, 2000; Diniz-Filho *et al.*, 2003; see also Fotheringham *et al.*, 2002). Note that although these estimates are shifted toward 'local' or 'semilocal' scales, they are still global estimates (i.e. averages across the entire studied area) and should not be confused with local analyses of spatial patterns (Fotheringham *et al.*, 2002; Foody, 2004). Since different factors affect richness at different spatial scales, it is usually expected that some methods designed to control and take into account spatial effects, such as GLS and autoregressive models, may tend to increase the magnitude of factors affecting richness at smaller spatial scales, and at the same time reduce the importance of climatic factors at broad scales (Lennon, 2000; Diniz-Filho *et al.*, 2003). However, our comparison between filtering and GLS did not clearly show this scale-shift effect, since the best predictors in GLS is annual temperature, which is also an environmental variable affecting richness at broad spatial scales. This reinforces our previous suggestion (Diniz-Filho *et al.*, 2003) that comparisons of these different statistical models may be difficult with these broad-scale environmental data because of the strong multicollinearity effects, and that indiscriminate use of these methods can cause confusion in terms of teasing apart the relative importance of these factors. As pointed out by Griffith (2003) and Borcard & Legendre (2002), we expect that spatial filtering methods can help ecologists to overcome one of the most debatable issues in geographical ecology, namely, the influence of spatial scales on the presumable mechanisms explaining spatial patterns in species richness.

Finally, our analyses with South American birds using the eigenvector-based spatial filters support many previous studies which showed that bird species richness is more associated with AET at broad scales, giving support to the climatic hypothesis in general, and to energy hypothesis in particular (Hawkins *et al.*, 2003a,b; Diniz-Filho *et al.*, 2004; Mathias *et al.*, 2004). Hawkins *et al.* (2003a) used a stepwise approach to show that AET is the best predictor for species richness, using this same dataset. How-

ever, since we used here the full set of predictors, we also showed that annual temperature, minimum temperature, and climatic heterogeneity are also important predictors of bird species richness (see also Diniz-Filho *et al.*, 2004). Marginally significant coefficients ( $P < 0.10$ ) also appear for rainfall and habitat variety. Moreover, the present analysis using eigenvector-based spatial filtering shows that these conclusions were not biased by the autocorrelation structures in data at different spatial scales. However, comparison between filtering and GLS suggest that these interpretations may be sensitive to the strong multicollinearity effects in data, although they all reinforce the multiple causal mechanisms driving species richness at different spatial scales.

## ACKNOWLEDGEMENTS

LMB and JAFDF are CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) researchers and acknowledge the grants from this agency. We thank Brad Hawkins and Thiago F. L. V. B. Rangel for critical reading of many versions of this manuscript, Alexandre Siqueira Guedes Coelho for programming GLS routine in SAS and Breno Vasconcellos for help in running Proc MIXED. Work by JAFDF and LMB was also supported by Fundação de Amparo à Pesquisa da Universidade Federal de Goiás (FUNAPE/UFG).

## REFERENCES

- Balmford, A., Moore, J.L., Brooks, T., Burgess, N., Hansen, L.A., William, P. & Rahbek, C. (2001) Conservation conflicts across Africa. *Science*, **291**, 2616–2619.
- Borcard, D. & Legendre, P. (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.
- Borcard, D., Legendre, P., Avois-Jacquet, C. & Tuomisto, H. (2004) Dissecting the spatial structure of ecological data at multiple scales. *Ecology*, **85**, 1826–1832.
- Chown, S.L., van Rensburg, B.J., Gaston, K.J., Rodrigues, A.S.L. & van Jaarsveld, A.S. (2003) Energy, species richness, and human population size: conservation implications at a national scale. *Ecological Applications*, **13**, 1233–1241.
- 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.
- Diniz-Filho, J.A.F., De Sant'ana, C.E.R. & Bini, L.M. (1998) An eigenvector method for estimating phylogenetic inertia. *Evolution*, **52**, 1247–1262.
- Diniz-Filho, J.A.F., Rangel, T.F.L.V.B. & Hawkins, B.A. (2004) A test of multiple hypotheses for the species richness gradient of South American owls. *Oecologia*, **140**, 633–638.
- Dutilleul, P. (1993) Modifying the t test for assessing the correlation between two spatial processes. *Biometrics*, **49**, 305–314.
- Foody, G.M. (2004) Spatial non-stationarity and scale dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. *Global Ecology and Biogeography*, **13**, 315–320.



- Fotheringham, A.S., Brunsdon, C. & Charlton, M. (2002) *Geographically weighted regression: the analysis of spatially varying relationships*. Wiley, Chichester.
- Gaston, K.J. & Evans, K.L. (2004) Birds and people in Europe. *Proceedings of the Royal Society of London Series B, Biological Sciences*, **271**, 1649–1655.
- Griffith, D.A. (2003) *Spatial autocorrelation and spatial filtering. Gaining understanding through theory and visualization*. Springer-Verlag, Berlin.
- Haining, R. (1990) *Spatial data analysis in the social and environmental sciences*. Cambridge University Press, Cambridge.
- Hawkins, B.A. & Diniz-Filho, J.A. (2004) 'Latitude' and geographic patterns in species richness. *Ecography*, **27**, 268–272.
- Hawkins, B.A., Field, R., Cornell, H.V., Currie, D.J., Guegan, J.F., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., Oberdorff, T., O'Brien, E.M., Porter, E.E. & Turner, J.R.G. (2003a) Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, **84**, 3105–3117.
- Hawkins, B.A. & Porter, E.E. (2003a) Relative influences of current and historical factors on mammal and bird diversity patterns in deglaciated North America. *Global Ecology and Biogeography*, **12**, 475–481.
- Hawkins, B.A. & Porter, E.E. (2003b) Water–energy balance and the geographic pattern of species richness of western Palearctic butterflies. *Ecological Entomology*, **28**, 678–686.
- Hawkins, B.A., Porter, E.E. & Diniz, J.A.F. (2003b) Productivity and history as predictors of the latitudinal diversity gradient of terrestrial birds. *Ecology*, **84**, 1608–1623.
- Legendre, P. (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology*, **74**, 1659–1673.
- Legendre, P. & Legendre, L. (1998) *Numerical ecology*. Elsevier, Amsterdam.
- Lennon, J.J. (2000) Red shifts and red herrings in geographical ecology. *Ecography*, **23**, 101–113.
- Lichstein, J.W., Simons, T.R., Shiner, S.A. & Franzreb, K.E. (2002) Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, **72**, 445–463.
- Littell, R.C., Milliken, G.A., Stroup, W.W. & Wolfinger, R.D. (1996) *SAS system for mixed models*. SAS Institute, Cary, NC.
- Lobo, J.M., Lumaret, J.P. & Jay-Robert, P. (2002) Modelling the species richness distribution of French dung beetles (Coleoptera, Scarabaeidae) and delimiting the predictive capacity of different groups of explanatory variables. *Global Ecology and Biogeography*, **11**, 265–277.
- Mathias, P.V.C., Mendonça, C.V., Rangel, T.F.L.V.B. & Diniz-Filho, J.A.F. (2004) Sensitivity of macroecological patterns of South American parrots to differences in data sources. *Global Ecology and Biogeography*, **13**, 193–198.
- Pergamon World Atlas (1968) *Pergamon World Atlas*. Pergamon Press, New York.
- Rahbek, C. & Graves, G.R. (2001) Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences, USA*, **98**, 4534–4539.
- van Rensburg, B.J., Chown, S.L. & Gaston, K.J. (2002) Species richness, environmental correlates, and spatial scale: a test using South African birds. *American Naturalist*, **159**, 566–577.
- Rohlf, F.J. (2001) Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution*, **55**, 2143–2160.
- Ruggiero, A. & Kitzberger, T. (2004) Environmental correlates of mammal species richness in South America: effects of spatial structure, taxonomy and geographic range. *Ecography*, **27**, 401–416.
- Selmi, S. & Boulinier, T. (2001) Ecological biogeography of Southern Ocean islands: the importance of considering spatial issues. *American Naturalist*, **158**, 426–437.
- Systat Inc (2000) *Statistics*, I, v. 10.0. Chicago, IL.
- Tognelli, M.F. & Kelt, D.A. (2004) Analysis of determinants of mammalian species richness in South America using spatial autoregressive models. *Ecography*, **27**, 427–436.
- Wartenberg, D. (1989) SAAP4.3: spatial autocorrelation analysis program. Exeter softwares, New York.

#### BIOSKETCHES

**José Alexandre F. Diniz-Filho** is interested in statistical methods applied to macroecology and evolutionary biology. Current projects involve the application of spatial autocorrelation analysis and phylogenetic comparative methods to understand ecological processes associated with latitudinal gradients in species richness. He is also interested in the application of spatial statistics to define operational units for intraspecific conservation, based on genetic data, and to establish priority areas for conservation.

**Luis Mauricio Bini** is interested in statistical methods applied to biodiversity analyses and limnology. Current projects involve the analysis of spatial population synchrony of aquatic assemblages in reservoirs and floodplains. He is also interested in how population dynamics is linked with more general biodiversity patterns, mainly the relationship between species diversity and ecosystem stability.