

# The Geometry of Spatial Analyses: Implications for Conservation Biologists

Victor Lemes Landeiro\* & William Ernest Magnusson

*Programa de Pós-graduação em Ecologia, Instituto Nacional de Pesquisas da Amazônia – INPA, Manaus, AM, Brazil*

## Abstract

Most conservation biology is about the management of space and therefore requires spatial analyses. However, recent debates in the literature have focused on a limited range of issues related to spatial analyses that are not always of primary interest to conservation biologists, especially autocorrelation and spatial confounding. Explanations of how these analyses work, and what they do, are permeated with mathematical formulas and statistical concepts that are outside the experience of most working conservationists. Here, we describe the concepts behind these analyses using simple simulations to exemplify their main goals, functions and assumptions, and graphically illustrate how processes combine to generate common spatial patterns. Understanding these concepts will allow conservation biologists to make better decisions about the analyses most appropriate for their problems.

**Key words:** Conservation Biology, Autocorrelation, Space, Variance Partitioning, SAR, Spatial Filters.

## Introduction

Spatial ecology has increasingly attracted the attention of ecologists and conservationists, and spatial analyses are frequently used in biodiversity conservation planning (Diniz-Filho & Telles 2002; Nams *et al.* 2006; Moilanen *et al.* 2008). For example, approximately 25% of the articles citing SAM software, a specialized spatial analysis software (Rangel *et al.* 2006), were concerned with biodiversity conservation (Rangel *et al.* 2010). Beale *et al.* (2010) listed four questions of interest to conservation biologists that potentially involve spatial analyses: 1) How does the spatial scale of human activity impact biodiversity or biological interactions? 2) How does the spatial structure of species' distribution patterns affect ecosystem services? 3) Can spatially explicit conservation plans be developed? 4) Are biodiversity patterns driven by climate? The third question is probably of most immediate concern to conservation biologists, and has spurred the development of complex algorithms to help land-use decision-making processes, such as Marxan with Zones (Watts *et al.* 2009). The mathematics associated with this type of question are usually normative (Colyvan *et al.* 2009), and designed to optimize the chances of obtaining a consensus decision.

Spatial ecology has opened many promising avenues of research for conservation. It has been used to extrapolate

and predict species occurrence (Austin 2002; Betts *et al.* 2006; De Marco *et al.* 2008), and may be used to predict the effects of global warming on biodiversity. However, one of the main strengths of spatial analysis in conservation is its capacity to describe the patterns of diversity at different spatial scales. Knowing what factors generate beta diversity, and at what spatial scales they act, can be of great importance to conservation planning (Legendre *et al.* 2005; Tuomisto & Ruokolainen 2006). Spatial analysis can be also used to identify patterns of genetic variability at different spatial scales and define operational units for conservation planning (Diniz-Filho & Telles 2002).

The rapid development and sophistication of spatial methods and their applications have enabled researchers to make predictions of species distributions and plan conservation efforts. For example, Bini *et al.* (2006) used simulation procedures to predict anuran species that could be discovered in the Cerrado biome by 2050, and showed that the predicted distributions lead to different priorities for placement of reserves than those based on currently known distributions of species. Some researchers have suggested that spatial interpolation to predict species distributions may be more effective than models based on environmental variables (Bahn & McGill 2007).

Arguably, all conservation related questions should be embedded in a landscape context (Metzger 2006). Chesson (2003, p. 253) commented "Would it not be more useful to focus on how physical environmental variation is translated into patterns exhibited by organisms?". However, recent

\*Send correspondence to: Victor Lemes Landeiro  
Programa de Pós-graduação em Ecologia,  
Instituto Nacional de Pesquisas da Amazônia – INPA,  
Av. André Araújo, 2936, CP 478, CEP 69011-970,  
Manaus, AM, Brazil  
E-mail: vllandeiro@gmail.com

discussion of spatial analyses in the scientific literature has focused on descriptive models that produce the parameters that can be used as inputs to more applied models. Beale *et al.* (2010, p. 246) asserted that “[...] many ecologists [...] often believe that spatial analysis is best left to specialists. This is not necessarily true and may reflect a lack of baseline knowledge about the relative performance of the methods available.” We suggest that, rather than being a problem of not understanding the relative performance of the methods, most conservationists focus on particular problems that can be approached with normative mathematics, and not on the problems in obtaining generally robust descriptive statistics that were derived from simulations using unrealistic ecological assumptions.

Most recent comparative evaluations of spatial methods used computer simulations to evaluate the relative utility of different methods (Dormann *et al.* 2007; Beale *et al.* 2010; Landeiro *et al.* 2011). These simulations are often difficult for biologists to appreciate because they are couched in terms of distance space and matrix algebra. In this paper, we use simple geometric models to illustrate the concepts behind regression analysis of distance data, and discuss what the results imply in terms of ecological processes that may be of interest to conservationists.

The leaders in spatial ecology usually explain ecology with the associated mathematics and statistics. However, ecologists and conservationists often find the explanations complex, due to the difference between space and most ecological variables. Ecological variables are generally treated as linearly additive by appropriate transformations or sampling procedures. That is, each variable represents a single dimension. However, space is usually measured in two or more dimensions in a coordinate system. The coordinates themselves do not necessarily represent the conceptual distance between two objects, which is usually the Euclidean distance. Some believe that space cannot be represented by linear additive combinations, and that joint analysis of spatial and ecological variables can only be undertaken by transforming the ecological variables to distances (Tuomisto & Ruokolainen 2006). Others claim that this procedure produces statistics that are difficult to interpret, and that space should be converted to linear additive components for inclusion in analyses (Legendre *et al.* 2005, 2008). Although we are inclined towards the latter, we wish to avoid these difficult conceptual problems because most of the concepts in spatial analysis can be understood in terms of simple one-dimensional spatial models (*e.g.* distances along a transect), and it is easier for an ecologist to appreciate the conceptual problems if they are first presented in models in which space is described in only one dimension.

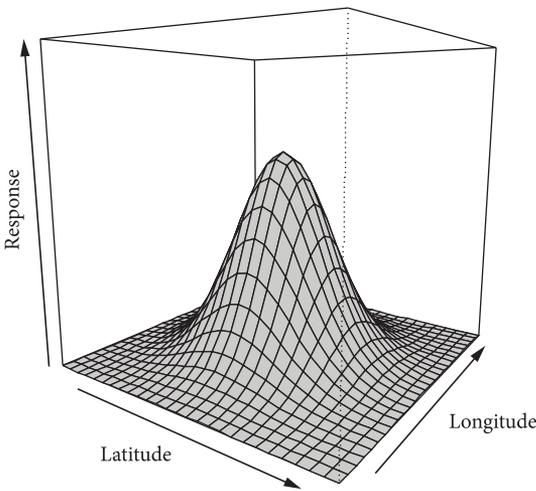
### Autocorrelation

Autocorrelation, as the name implies, is the correlation of a variable with itself. This correlation could be in time

or space. For example, values of a variable are temporally autocorrelated if the values of that variable at short time intervals are more or are less similar than expected for randomly associated pairs (Legendre & Legendre 1998). The same is true for spatial autocorrelation, in which values nearby are more similar than values from points separated by greater distances. There are several causes of spatial autocorrelation and this is the greatest source of confusion, because different definitions for spatial autocorrelation are used in relation to the process that generates it. For example, according to Peres-Neto & Legendre (2010, p. 175), autocorrelation results from “[...] spatial structure due to the dynamics of the species (or their communities) themselves (*e.g.*, via dispersal.)” Under this definition, spatial autocorrelation is not used for predictor variables but rather is used only for response variables that are autocorrelated by endogenous causes. The many definitions used in spatial ecology generate confusion, such that some authors have published their own glossary (Peres-Neto & Legendre 2010). The difference between the definition of Legendre & Legendre (1998), who defined autocorrelation in relation to pattern, and that of Peres-Neto & Legendre (2010), who defined autocorrelation in terms of process, is important, and reflects on another important concept, “stationarity.”

Fortin & Dale (2005, p. 11) defined stationarity as “[...] a process, or the model of a process, is stationary (or homogeneous) if its properties are independent of the absolute location and direction in space [...] the parameters of the process, such as the mean and variance, should be the same in all parts of the study area and in all directions.” However, whether this refers to the underlying process or the resulting pattern is unclear. Consider an organism that colonizes a point in a previously empty space, and then reproduces. Assuming that the organism and its descendants have limited dispersal, after a few generations the density of the species can be represented by a single peak in the previously empty space (Figure 1). The process that generated that peak was endogenous autocorrelation (we did not need information on anything but the density in neighboring sites in the previous generation to produce the peak), and the process was stationary (*i.e.* knowing the process, we only needed information on the densities in neighboring sites, independent on where we were in space).

A problem arises when we only have the pattern and are unsure of the process. Imagine no endogenous autocorrelation, but that the peak in population density corresponds to a physical peak in the landscape, which might happen if the density of the organism were related to temperature or some other correlate of altitude. The pattern is identical, but the density of the organism is a function of temperature and not a function of the density in neighboring sites. In this case, a combination of endogenous autocorrelation and an external driving variable can generate exactly the same pattern. The literature can be confusing because the interpretation of autocorrelation

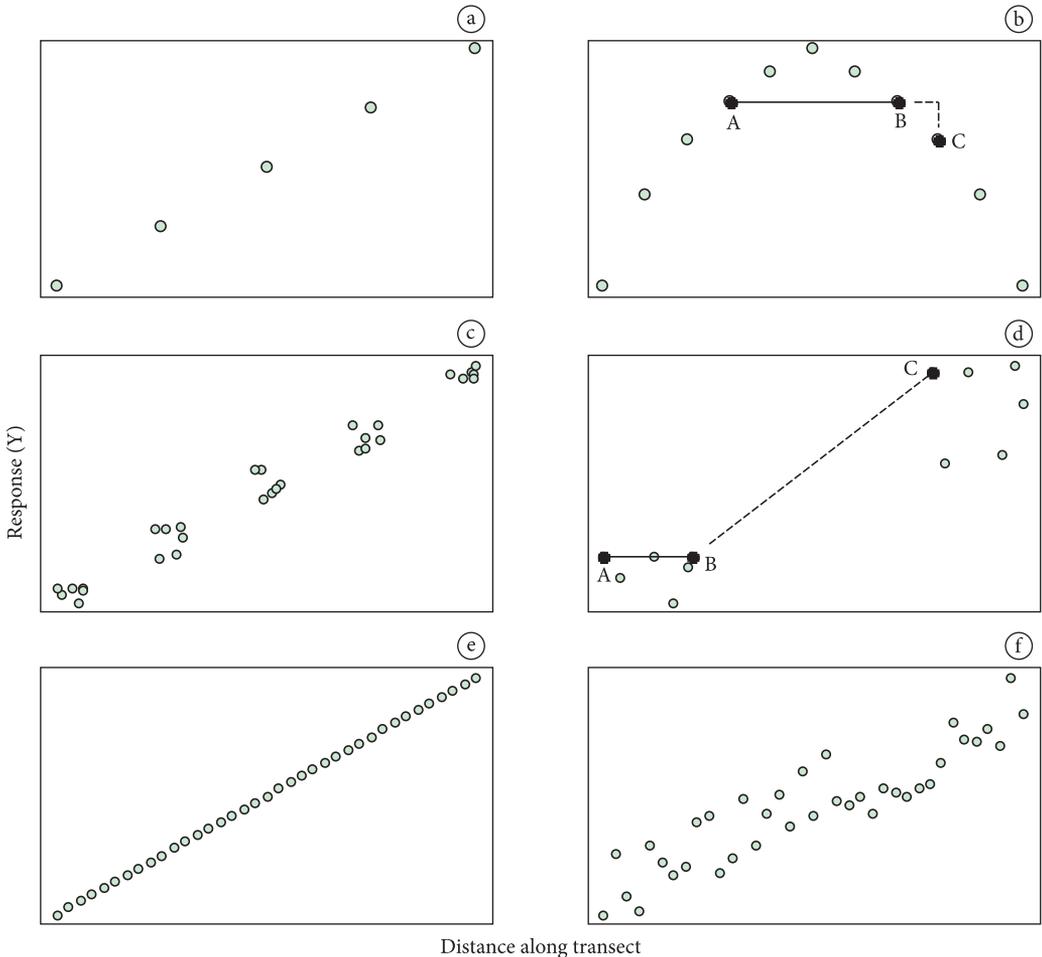


**Figure 1.** A peak of abundance representing the distribution of a species. Some factor associated with intrinsic biology of the species, such as reproduction or limited dispersion could create such pattern.

and stationarity depends on the researcher’s assumptions about the underlying processes, and we generally only have information on the pattern.

*Stationarity in one or two dimensions*

Consider a response variable (Y) that varies with distance along a transect (T), as shown in Figure 2a. An assumption of most spatial analyses is that the relationship between Y and space is stationary. That is, the variation of Y across T is the same independent of the observer’s position along the transect, and in any direction (*i.e.* the relationship is independent of the position in T). That condition can be seen to hold for the data in Figure 2a. Starting from any point, an increase in the distance along T of one unit, will increase the value of Y by a constant amount. Conversely, if we decrease T by one unit, we decrease the value of Y by the same constant amount.



**Figure 2.** The difference between a stationary process and a stationary pattern. a) a stationary pattern, where the effect of distance along a transect is independent of location or direction; b) a non-stationary pattern that could result from a stationary process acting over a limited time period; c) a pattern that could arise from a small-scale stationary process acting over a stationary pattern, such as reproduction with limited dispersal of the organisms illustrated in part A; d) magnification of A, a stationary process may create a non-stationary pattern; e) a stationary pattern similar to that in A, but the organisms are closer together. A small-scale stationary process, such as that illustrated in part C, does not produce a recognizably non-stationary pattern in this case, as seen in part f).

It is important to note that the only way for the observed relationship between Y and distance to be stationary is for the relationship between Y and T to be linear. Any nonlinear relationship will result in the effect of distance being dependent on spatial location (*i.e.* the value of T at which we start to measure the distance). This is illustrated in Figure 2b, where the relationship between Y and T is nonlinear. If we start at point B and move 1 unit forward along the T axis to C, Y is reduced by  $\sim 0.287$ . If we start at A and move four units forward along T, Y remains constant. In one dimension, the only way that the relationship between Y and distance can be stationary is for Y to have a linear relationship with distance. In two dimensions, the only way that the relationship between Y and distance can be stationary is if the value of Y can be represented in space by a flat plane with no curvature. Note that a small-scale stationary process, such as that described in Figure 1 can generate an apparently nonstationary pattern at a larger scale.

If the relationship between the value of a variable and space is linear in one dimension (*i.e.* the pattern is unambiguously stationary), it does not matter whether we use a conventional analysis or an analysis based on distances. For instance, we could calculate the differences between the values of Y ( $\delta Y$ ) for each pair of points and regress this against the distances between the points. A Mantel test uses the absolute value of the distance, but as we are only considering one dimension, we could use a positive or negative sign to indicate direction. The value of the slope of the regression (the amount that the dependent variable increases for a one unit increase in the independent variable) is logically the same whether the dependent variable is Y and the independent variable T, or whether the dependent variable is  $\delta Y$  and the independent variable  $\delta T$ . However, the values may only be the same if we use geometric mean regression for the second analysis, because we have artificially inflated the variance in T by using  $\delta T$ , and this biases the estimate of the slope downwards for least-squares regression (Zar 1996). The slope of the relationship is only representative of the “effect of distance” if the relationship with the resultant variable is stationary. As with any simple regression, if the underlying relationship is not linear (*i.e.* the effect of space is a variable and not a constant), estimating a single slope parameter is meaningless.

What this means for the construction of most conservation-related models is that useful parameters are only obtained if that parameter is a constant, unless we are willing to move to likelihood methods or Bayesian statistics and try to generate a probability distribution for the values of the parameter. We will use simple one-dimensional models to illustrate the recent discussion in the literature, and evaluate the relevance of those discussions to ecologists undertaking conservation research.

### Stationarity of pattern and stationarity of process

Note that a stationary process does not necessarily generate a stationary pattern (Fortin & Dale 2005). Let us imagine a secondary process that has a nonlinear relationship with space. For instance, each point on Figure 2a could represent a value for a single individual. If that individual reproduces, and dispersal is limited, we may see a pattern like that on Figure 2c, with similar values of Y (similar because of genetic similarity or maternal provisioning) at close by points in space. Although the process (reproduction with limited dispersal) is the same at each point (*i.e.* stationary), the resulting pattern is not stationary. This can be seen by amplifying the area around what were originally two individuals (Figure 2d). Although individuals vary in Y, the mean value of Y does not increase between points A and B. However, the effect of the same difference in Y between B and C is much greater. This point is important. A stationary process at one scale does not necessarily generate a stationary pattern at larger scales, and many analyses assume a stationary pattern.

We gave an example of a stationary process generating a nonstationary pattern in Figure 2a. Interpretation of a pattern generated by a nonlinear stationary process can be difficult, as can be seen from Hubbell's (2005) neutral theory of biogeography. By using simulations analogous to those we used to generate Figure 2c, but with many more potential species, Hubbell (2005) generated local communities that varied over a much larger metacommunity landscape. The overall analysis is very complicated, but the result of most relevance to spatial patterns is that this process led to similarity among local communities that decreased linearly with the log of distance. That is, the relationship of similarity (the complement of ecological distance) was nonlinear with distance, even though the process that generated that similarity was the same at each point.

It would appear easy to deal with this situation. We could carry out a Mantel test of the relationship between similarity and log distance, but transforming a distance matrix has complex implications for interpretation. The rules we use in mathematics generally conform to Euclidean geometry, but the geometry of curved surfaces is much more complex, and manipulation of such geometry is not a trivial task, even for geniuses, such as Einstein (Mlodinow 2001). If the “effect of distance” is not linear, the effect of a particular unit of distance (say the distance you walk from point 1 to point 2) depends on the position of the observer relative to those two points. This is the theory of relativity, and not the sort of problem that most ecologists are thinking of when they ask “How much does distance matter?”

The apparent effect of a secondary nonlinear process depends on the dispersal of the primary units (those generating the secondary response). The points in Figure 2a were widely scattered, and we assumed that these were the only individuals in the population (*i.e.* not the only ones sampled). Therefore, the secondary process of reproduction produced

clumps of points that reflected the autocorrelation. If the initial individuals were close together in relation to the extent of influence of the secondary process (Figure 2e), there may be no obvious clumping (*i.e.* the pattern is stationary) after the action of the secondary process (Figure 2f), even though the same mechanistic process generated the data. Pattern may be useful to indicate the probable action of a secondary process, but the absence of pattern is not necessarily evidence of the absence of that process. This is important because all spatial analyses are about detecting clumping and trying to determine what caused that clumping so that nuisance variables can be discounted (controlled) and interesting variables can be analyzed.

If clumps can be identified a priori, it may be possible to select the most probable hypotheses and discard the most unlikely (Barnett *et al.* 2010). However, most stationary positive autocorrelation processes will lead to an essentially uniform distribution of the dependent variable if left to act long enough in a homogeneous landscape. Strong clumping is usually strong evidence that a stationary positive autocorrelation process is not acting alone. Assumption of an autocorrelation process may lead to erroneous biological conclusions when some other process causes clumping (Barnett *et al.* 2010).

## Spatial Analysis

### *Clumping as an indication of the effect of space*

Most hypotheses about ecological communities attempt to explain spatial patterns (clumping). However, researchers seek independent evidence, and spatial proximity may cause pseudo-replication (Hurlbert 1984). Therefore, researchers face the quandary of forming hypotheses due to spatial clumping while attempting to avoid clumping to test those hypotheses. Space is not an ecological variable, but rather reflects some process that varies spatially (Diniz-Filho *et al.* 2003). Clumping may occur at any of a variety of scales, from large (Figure 2a) to small (Figure 2d) with many intermediate possibilities (Legendre & Legendre 1998; Legendre *et al.* 2002). It may be illogical to try to study many phenomena occurring at different scales in the same analysis (Fortin & Dale 2009), and all spatial analyses can be considered attempts to isolate the effects of particular independent variables from other processes that cause clumping.

General trends (which may be the only stationary patterns) might be excluded before undertaking spatial analyses, or removing effects of local patterns might be necessary. Regardless, the choice of which scales to study should be determined by the questions, not the analysis (Diniz-Filho *et al.* 2007; Fortin & Dale 2009). There is no scale at which only endogenous autocorrelation can be assumed, and endogenous autocorrelation does not necessarily occur only at one scale. Consider the distribution of individuals of a species of plant that is dispersed passively by gravity

and also by birds. This will result in two scales of clumping, both of which are endogenous. If the extent of the study is small in relation to the extent of endogenous autocorrelation, the autocorrelation may be manifest as a broad-scale trend across the study area (Beale *et al.* 2010). Removal of such a trend to obtain “stationarity”, as is frequently recommended in time-series analyses, may be totally inappropriate.

The first step in an investigation of the role of space in ecology is exploratory data analysis (EDA). In this step we do not invoke process and must only investigate pattern. Therefore, we use the definition of Legendre & Legendre (1998), which defines autocorrelation in terms of pattern, rather than that of Peres-Neto & Legendre (2010), which defines autocorrelation in terms of process, because distinguishing endogenous from exogenous autocorrelation requires knowledge of the process. In this step, we are asking questions, such as “Are my data spatially autocorrelated?” “Is the response variable, the predictor variable, or both autocorrelated?” “If yes, what is the extent of autocorrelation?” “Are model residuals autocorrelated?” “Should I use a spatial analysis to take autocorrelation into account (see below)?”

Measures of autocorrelation, such as Moran’s  $I$  and Geary’s  $c$  and their correlograms are used to explore these questions. Correlograms are used to detect statistically significant spatial structure (*i.e.*, the pattern, not the process) and to describe its general features. Combined with maps, they are used to assess the magnitude and the pattern of autocorrelation in data sets (Legendre & Legendre 1998). However, it is not obvious what criteria should be used to indicate when space needs to be taken into account, and several authors recommend the use of spatial analyses on the basis that they will always improve interpretation (Dormann *et al.* 2007; Beale *et al.* 2007, 2010).

### *Why undertake spatial analyses?*

When nearby values of variables are more similar than expected at random, a pattern of positive autocorrelation is assumed, and produces two major classes of problems in spatial analyses. The first is conceptual and related to the structure of the causal interpretation of the model being investigated. When we introduce “space” into the model, we are including it as surrogate for some biological or physical process, which induces spatial autocorrelation. If it is only a surrogate for a nuisance variable, then eliminating the effect of space will not affect our interpretation. However, if it is also a surrogate for a variable we wish to investigate, removing the “problem” of space may eliminate an effect that we wanted to study. Therefore, before analysis, it is necessary to decide which aspects of space we want to include in the analysis, and which aspects we want to discard. This decision is biological/conceptual and often very difficult when we know little about the functioning of the biological systems. However, it is also the most important decision, because it will affect all of our interpretations (Legendre 1993; Legendre *et al.* 2002).

The second class of problems is statistical/computational. Autocorrelated data can give the wrong estimates of degrees of freedom for conventional statistical tests and consequently gives inflated type I error rates (Legendre 1993). This effect is often called pseudoreplication, but it is very different from the pseudoreplication caused by confounding variables described in the previous paragraph. Spatial autocorrelation may also affect estimates of regression coefficients due to red shifts caused by spatial autocorrelation (Lennon 2000).

Discussion of the points alluded to in the preceding paragraphs (mainly the one related to coefficient shifts) is recent, and filled with controversies (Lennon 2000; Diniz-Filho *et al.* 2003; Hawkins *et al.* 2007; Bini *et al.* 2009). The second class of problems has been the focus of most of the recent discussions in the literature (Diniz-Filho *et al.* 2003; Dormann *et al.* 2007; Begueria & Pueyo 2009; Bini *et al.* 2009), but these aspects are also related to the practice of partitioning variance between interesting predictor variables and the possibly confounding factor “space” (Borcard *et al.* 1992; Legendre 1993; Legendre & Legendre 1998). Partitioning variance between “space” and ecological predictor variables is the focus of research on niche versus neutral models of community dynamics (Peres-Neto *et al.* 2006; Legendre *et al.* 2009a, 2009b; Peres-Neto & Legendre 2010), and the question of whether area or habitat is more important for reserve design.

Before deciding which spatial analysis to use, one must answer the following conceptual questions:

- 1) Do we only want to remove the possible effects of other variables that are spatially confounded with the predictor variable?
- 2) Do we want to partition the variance in the response variable into that which appears to be associated only with the predictor variable(s) and that which may be associated with the predictor variable and/or other variable(s) that are confounded with spatially-structured environmental variation?
- 3) Do we only want to use a spatial analysis to remove spatial autocorrelation in order to be able to use standard statistical tests?
- 4) Do we want to describe spatial patterns in response and predictor variables, relating them to a specific spatial scale where they are most affected by autocorrelation?

Some analyses do more than one of these simultaneously, but it is important that we recognize which problems are being resolved, because there is no general method that can solve all the conceptual and statistical problems simultaneously.

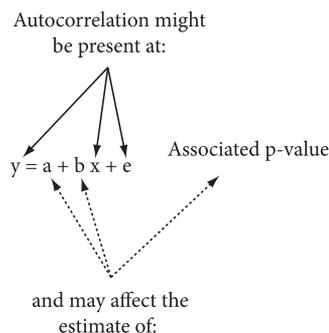
### Where is space in my model?

In general, the construction of an ecological model is a trade-off between complexity and utility (Levins 1966).

In the best-case scenario, the predictor variables should be orthogonal to space and therefore not autocorrelated; however this rarely occurs in observational studies. In the simplest form of statistical tests, inclusion of spatial variables decreases spatial autocorrelation in the residuals, but reduces degrees of freedom. When modeling, our data might have autocorrelation patterns in the response variable, in the predictor variables, and/or in the errors (residual) of the model (Figure 3).

An assumption of most statistical tests is that the errors are independent and identically distributed (the so called I.I.D. of errors), and it is common practice to say that residuals results from all factors not included in model; e.g. soil pH, land use history (Diniz-Filho *et al.* 2003). The assumption of I.I.D. of residuals (errors) is necessary to generate the distributions of statistics under null hypotheses for most tests. In the “error model,” residuals may be independent (first i in I.I.D.). In “residual” (ecological) models, residuals are known not to be independent because they have causal relationships with variables not included in model. At most, we can hope that they are independent of the variables included in the model. One of the external variables traditionally relegated to the residual variation is “space.”

When “space” affects variables in the analysis, the residuals may have a spatial pattern. Consequently, the decision to use spatial methods may come as a result of an evaluation of residuals. If residuals are autocorrelated then spatial analysis is used. However, statistical tests are compromised only when both the predictor and response variables are autocorrelated (Legendre *et al.* 2002). Therefore, residuals can be spatially structured without inducing statistical bias (P.R. Peres-Neto *Personal Communication*). In fact, the residuals may remain autocorrelated even after the use of the appropriate spatial analysis (Beale *et al.* 2010). Therefore, the choice of the appropriate test should not be based only on analyses of residuals, but by assessing whether both response and predictor variables are spatially structured.



**Figure 3.** The basic structure of a linear-regression equation. Autocorrelation might be present in the response ( $y$ ) and/or in the predictor ( $x$ ) variables, as well as in the errors ( $e$ ). When present, autocorrelation might affect the estimate of p-values, though the existence of shifts in the estimates of the intercept ( $a$ ) and the slope ( $b$ ) is debatable (Lennon 2000; Diniz-Filho *et al.* 2003; Hawkins *et al.* 2007).

Several solutions have been proposed to manage the spatial autocorrelation in ecological data. We distinguish among two groups of solutions: i) removers - autocorrelation is a problem that should be removed from data; and ii) includers - autocorrelation is a natural process that should be understood and studied as an ecological phenomenon, not as a statistical problem. Generally, “the removers” tend to delete sampling sites until the data are no longer autocorrelated (Legendre & Legendre 1998, p. 14, describe this process, but do not recommend it), or to apply some type of correction to obtain the geographically effective degrees of freedom (Dutilleul 1993; Dutilleul *et al.* 2008). “Removers” do not necessarily try to take out all of the autocorrelation, but may restrict analyses to data grouped in scales relevant to the question, and in which it is unnecessary to account for autocorrelation at other scales. The “inclusive methods” are based on statistical procedures that take spatial autocorrelation into account (Dormann *et al.* 2007), changing the way that the data are analyzed and interpreted (Legendre 1993).

## Simulations

### *What the simulations mean*

In the following sections we will use simple models with space represented by a single dimension (distance along a transect) to illustrate the results of some of the simulations in the literature, and their implications for different types of analyses. Basically we will generate 16 types of simulated data (Figure 4) and analyze these data using simple Ordinary Least Squares (OLS) regressions, Simultaneous Autoregressive (SAR) models (error, lagged, and mixed), Generalized Least Squares (GLS), and Spatial Filtering Techniques (using three different procedures to choose spatial filters to use in the model). Details of simulations and analysis are in the supplementary material. The effects of spatial autocorrelation on our interpretations depend on its strength and extent (Beale *et al.* 2010). We will discuss that later, and start with simple combinations of large-scale (a linear trend across the transect – Figure 2a), small-scale autocorrelation generated by local processes (such as in Figure 2d), and no autocorrelation (random association with space). Either or both of the dependent and independent variables may have no, large-scale, small-scale, or large- and small-scale autocorrelations. The possible combinations and resulting patterns in the relationships between dependent and independent variables are shown in Figure 4.

We can group the 16 graphs in three general scenarios: 1) Autocorrelation in either the dependent or predictor variable, but not in both (Figure 4b, c, d, e, i, m). 2) Both the dependent and independent variables are spatially autocorrelated, but they are orthogonal (independent in the sense that information on one relationship does not allow prediction of values generated by the other), and spurious relationships are unexpected for purely geometrical reasons

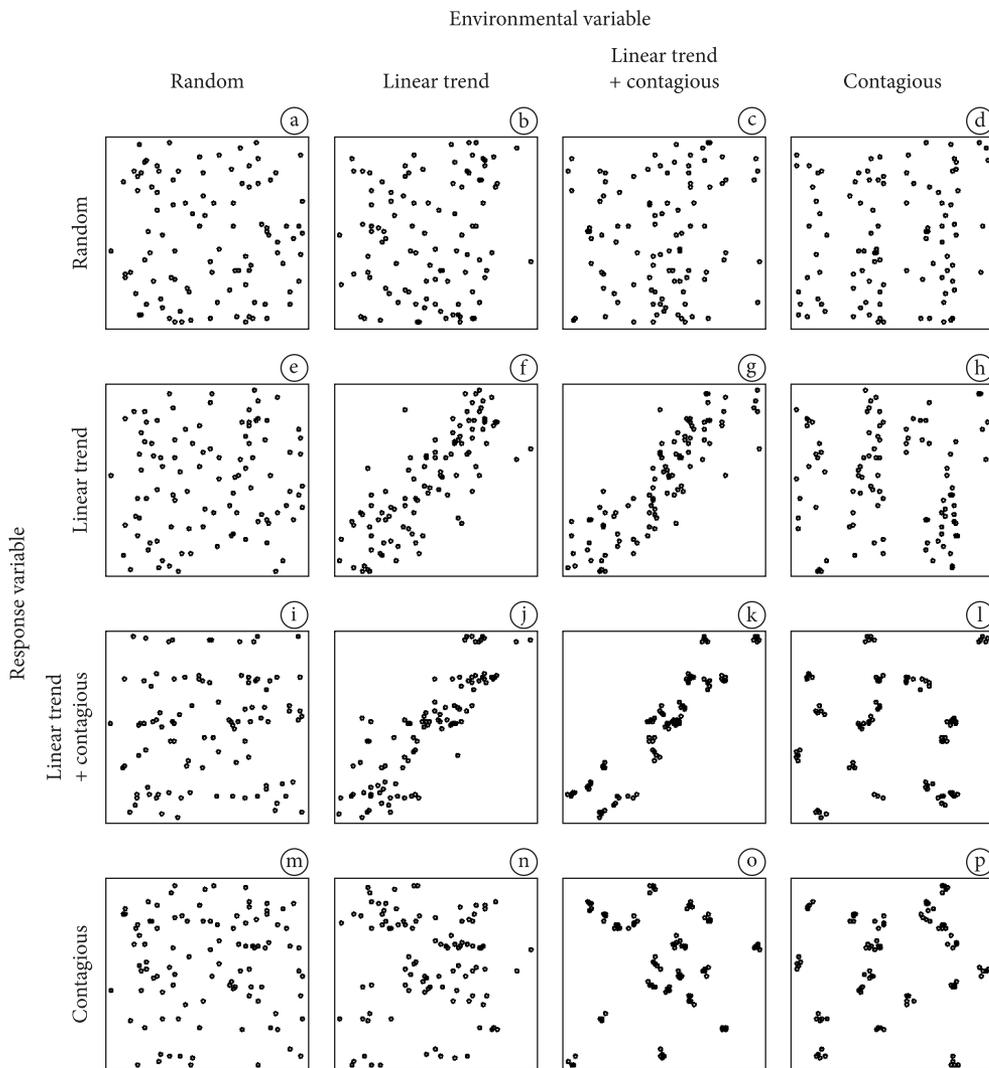
(Figure 4h, l, n, o, p). 3) Both variables are linearly related to space, resulting in a spurious relationship between them due to their common relationship with space (Figure 4f, g, j, k).

These data were generated to avoid a causal relationship between the dependent and independent variables. That is, information about the independent variable was not used to generate the dependent variable. Therefore, the ideal statistical test would not indicate a relationship between the dependent and independent variables. If we apply a test of the relationship between the dependent and independent variables many times (we used 1000 times in our simulations), they should give an apparently significant result only once in twenty times, if we use the conventional critical level to reject the null hypothesis of 0.05. While we do not recommend an arbitrary 0.05 “significance” level, it is commonly used to estimate the frequency of type I error (how often the null hypothesis is rejected erroneously).

**Scenario 1** – All of the combinations in scenario 1 involving autocorrelation in the dependent variable (Figure 4a, b, c, d, e, i, m) induce autocorrelation in the residuals of a regression of the dependent variable on the independent variable (Table 1), but conventional statistical tests produce about the correct level of type I error (0.05). This is expected because statistical tests are compromised only when both the predictor and response variables are autocorrelated (Legendre *et al.* 2002). However, advocates of spatial analyses claim that spatial analyses should be carried out always because spatial autocorrelation may affect the analyses even when statistical tests do not detect autocorrelation at the appropriate significance level. We will not enter into this debate, but it clearly would be beneficial to have diagnostic statistics to indicate when autocorrelation in the variables is likely to lead to compromised statistical tests.

**Scenario 2** – Both the dependent and independent variables are autocorrelated, but the processes that lead to autocorrelation are independent for each variable, such that we would not expect a relationship between them for geometric reasons (*i.e.* they are geometrically orthogonal, Figure 4h, l, n, o, p). This situation is probably rare in nature (Betts *et al.* 2009), and is not what worries most ecologists. However, this scenario has been used in simulations by most modelers (Dormann *et al.* 2007; Betts *et al.* 2009; Beale *et al.* 2010) because it gives a good example of how autocorrelation can give spurious statistical results, despite apparently orthogonal geometry. In this scenario, the null hypothesis of no relationship between the dependent and independent variables is true, but ordinary least squares (OLS) regressions indicate significant relationships (Table 1). Of the eight spatial methods frequently recommended, only three, those related to SAR methods, returned type I error rates close to the nominal 0.05 level (Table 2).

**Scenario 3** – Both the dependent and independent variables have spatial relationships that lead to a spurious relationship between them (Figure 4f, g, j, k). This is probably the most common case confronting ecologists and conservationists.



**Figure 4.** Sixteen combinations that can result from sampling different combinations of the structures described in Figure 2a-d. We sampled 200 equidistant points, spaced by five units, along the transect.

**Table 1.** Results of ordinary least squares (OLS) regression models for 1000 simulation runs combining samples taken from our four scenarios (Figure 4). Type I error rate/rate of times that the residuals were autocorrelated at the first distance class among 1000 simulation runs.

		Environmental			
		Random	Linear	Linear+Contagious	Contagious
Response	Random	0.059/0.065	0.045/0.040	0.047/0.041	0.054/0.053
	Linear	0.056/1.000	1.000/0.713	1.000/0.543	0.387/1.000
	Linear+Contagious	0.049/1.000	1.000/0.986	1.000/0.937	0.407/1.000
	Contagious	0.050/0.911	0.353/0.867	0.385/0.875	0.412/0.924

If all clumping (autocorrelation pattern) in the dependent variable is due to the effects of independent variables, there is no statistical problem due to the autocorrelated pattern (Beale *et al.* 2010). However, with real data, the cause of clumping is being inferred, and is not known before analysis. The clumping could be due to endogenous autocorrelation (a process affecting only the dependent variable), due to

independent variables included in the model, or other independent variables not included in the model. Researchers tend to assume that the spatial autocorrelation is totally attributed to endogenous processes (*i.e.* not due to habitat). However, that is a very sweeping assumption that should be supported by strong natural-history justifications.

**Table 2.** Proportion of simulation runs that had a p-value  $\leq 0.05$  out of 1000. Row names are the analysis used and column names are the variables used in the model. Y indicates a response variable and X a predictor one. Subscript  $c$  indicates contagious,  $l$  indicates linear, and  $l + c$  indicates linear plus contagious.

	$Y_c - X_l$	$Y_c - X_{l+c}$	$Y_c - X_c$	$Y_{l+c} - X_c$	$Y_l - X_c$
OLS	0.504	0.561	0.564	0.565	0.513
SAR <sub>error</sub>	0.051	0.047	0.054	0.050	-
SAR <sub>lagged</sub>	0.064	0.058	0.062	0.080	0.070
SAR <sub>mixed</sub>	0.053	0.046	0.054	0.061	0.051
GLS	0.241	0.409	0.417	0.405	0.399
ME	0.577	0.618	0.653	0.685	0.636
SF	0.566	0.659	0.657	0.669	0.775
PCNM	0.573	0.832	0.837	0.708	0.102

We used spatial confounding with a large-scale trend because it is easier to visualize, but confounding can result when autocorrelation is on a similar scale for the dependent and independent variables, independent of the scale of the autocorrelation. The problem of incorrectly estimated probabilities remains along with the extra problem of confounded effects. Hurlbert (1984) referred to the action of an unrecognized confounding variable as “demonic intrusion” If the objective of spatial analyses is to evaluate the possible effects of all spatially confounding variables by including them in the model as “space”, then space represents demonic intrusion. As we have seen, “space” is what we use to represent clumping. By including the effects of clumping, we are including the effects of all confounding variables that cause clumping.

In this case, the most we can do is to separate the variability in the dependent variable into parts that are generated by different processes. Part can be unambiguously attributed to the non-spatial independent variables included in the model, and part can be unambiguously attributed to spatially aggregated effects, which could be due to endogenous processes, such as limited dispersal of organisms, or spatially aggregated predictor variables not included in the model., Part of the variability cannot attribute to anything (residual), and the rest could be due to either the spatial predictors or the other independent variables included in the model (Figure 5). To separate the effects of space and predictor variables, we must model autocorrelation in the independent variable that corresponds to autocorrelation in the dependent variable. Borcard & Legendre (2002) has pioneered this type of analysis, mainly using a technique called Principal Coordinates of Neighbourhood Matrix - PCNM (Dray *et al.* 2006; Legendre *et al.* 2009a). However, any of the methods that take spatial autocorrelation into account in the independent variable may be used (Table 2).

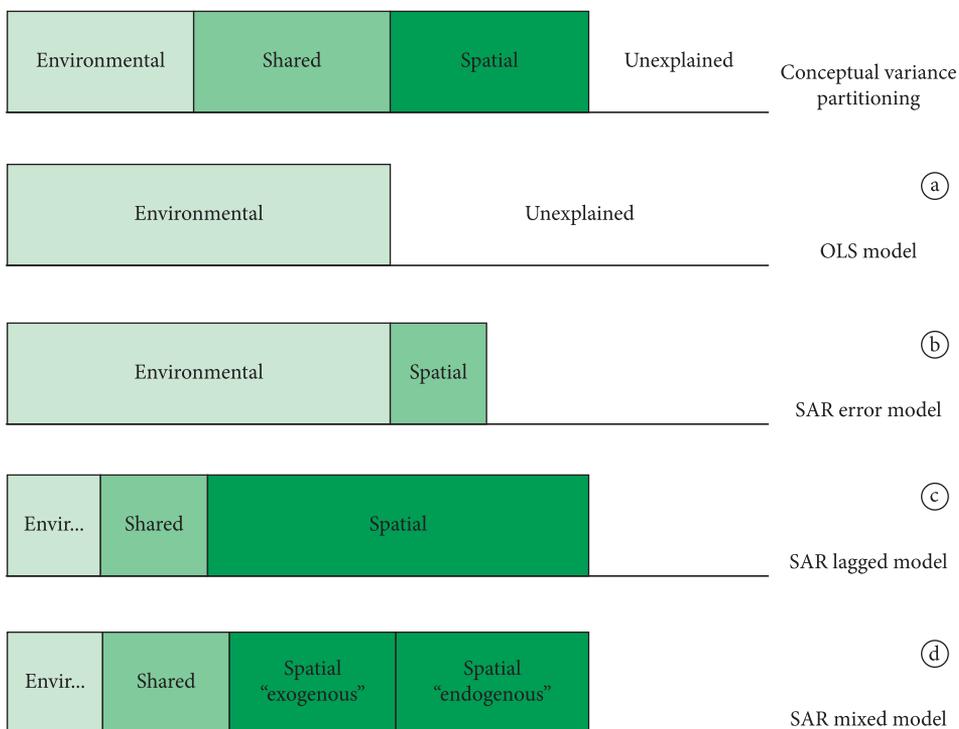
The down-side of taking into account the potentially confounding effect of space is that when we take out “space” we may be removing a true effect of the independent variable. This will affect our estimates of the regression coefficient for the independent variable. We have seen that, even when the effects of space and the independent variable are orthogonal, many of the spatial techniques, including

PCNM, may provide unbiased estimates of the slope of the regression, but with great cost in precision (Figure 6). This is important, because an imprecise estimate of the regression coefficient will lead to imprecise variance partitioning (*i.e.* the amount of potential confounding). Because researchers normally do one or a few studies, and have only one or a few estimates of the regression coefficient, it may not be very relevant that if they had done 1000 studies, the mean estimate of the regression coefficient would have been close to correct. Worse still, some of the best methods for dealing with the statistical problem of high rates of type I error for scenario 2 (*e.g.* autoregressive models) produce strongly biased estimates of the regression coefficient in scenario 3.

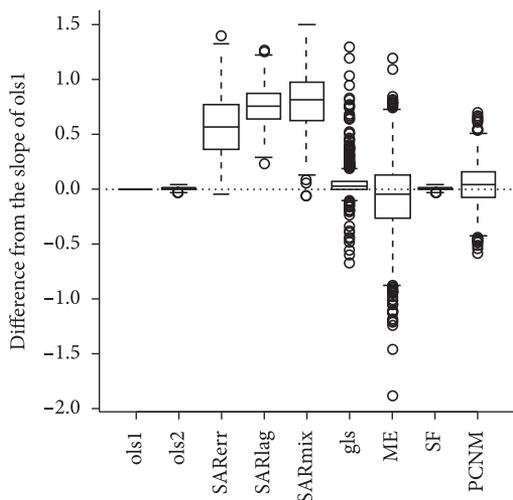
There is also a conceptual problem with the exercise of attributing proportions of variance to “space”. Beside the fact that the result will be biased if there is a miss match between the scale of sampling and the scale of effect of predictor variables (De Knecht *et al.* 2010), the answer must be scale specific. The amount of variance due to any variable is not a characteristic of the biological system, it is a characteristic of the sampling scale. Any discussion of the proportion of variance attributable to factors causing endogenous autocorrelation should be prefaced by an explanation of why that particular scale is of interest for the conservation problem in hand.

### More complex simulations

Beale *et al.* (2010) have carried out comprehensive simulations that are extensions of scenario 2, with collinear predictor variables, model selection algorithms and application of regression techniques designed to address problems derived from the violation of assumptions. In general, their conclusions are similar to those presented here, although some methods that work well under simple scenarios are not improved by use of model selection algorithms. Model selection for collinear variables is an extremely complex subject and perhaps more polemical than selection of spatial techniques (Taper & Lele 2004). The two most complex scenarios presented by Beale *et al.* (2010) were the scenarios in which none of the methods worked well and are useful to illustrate the limitations of spatial analyses in general.



**Figure 5.** Conceptual variation partitioning of OLS and SAR models. The first is the conceptual variation partitioning diagram, showing the environmental-only component, the environmental shared with the spatial component, the spatial-only component, and the unexplained variation. The remaining partitions are for: a) OLS models, in which there is considered to be only the environmental component and the unexplained variance; b) SAR error, in which a spatial variable is created to account for the autocorrelated errors, so this model conceptually has no shared component; c) SAR lagged, in which a spatial variable is created to explain spatial patterns of the response variable, so there is a shared component between environmental variables and the spatial component; and d) SAR mixed models, in which two spatial variables are created in a way that the spatial component might be interpreted as two spatial only components, one related to the endogenous autocorrelation  $\rho_{WY}$ , and the other related to the exogenous autocorrelation  $\gamma_{WX}$ .

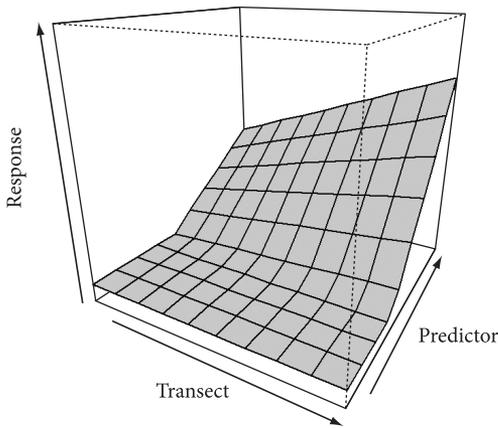


**Figure 6.** Boxplots representing the differences found in the slope (standardized coefficients) between OLS1 estimated parameters from the other analysis run after the data being "pseudoreplicated". The line inside the boxes is the median, the box indicates the first and third quartiles and whiskers which extend to the minimum and maximum values (points are outliers further from the mean than 1.5 times the box length).

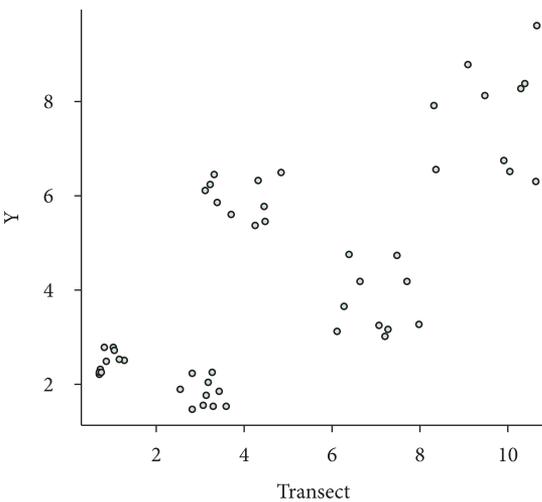
The first situation is where the relationship between the dependent and independent variables is nonstationary. As in Beale *et al.* (2010), we simulated no relationship between the dependent and independent variables on one side of the space (in our case, on one side of the transect) and a strong relationship on the other side (Figure 7). The lines in Figure 7 illustrate the relationship we are trying to describe. It is clear why a global model cannot describe this situation. The regression coefficient is not a constant, and any model that ignores that will be misleading. This is independent of the possible autocorrelation in the residuals or any other statistical problem. The model is so badly specified that it is meaningless to compare the utility of the different methods.

The second situation, which Beale *et al.* (2010) surprisingly considered worse than the first, is when the general model is correct, but the autocorrelation in the residuals is nonstationary. They modeled an increase in the extent of the autocorrelation across their spatial coordinates. This relationship is illustrated in one dimension in Figure 8. This situation is analogous to breaking the assumption of homogeneity of variance (heteroscedasticity) in a simple regression situation, with well-known consequences

(incorrect estimates of type I errors). The estimate of the regression coefficient is generally not badly affected by heteroscedasticity in a simple regression, but estimates of slopes with collinear predictor variables and heteroscedasticity may be very inaccurate (Beale *et al.* 2010). Although we agree with Beale *et al.* (2010) that nonstationarity of the autocorrelation in the residuals is a grave problem, we believe that, unlike the error in model specification described above, it is not inherently unsolvable, and, where individual clumps can be recognized, analyses such as those described by Barnett *et al.* (2010), which include different variances for each level of the predictor



**Figure 7.** Example of a situation in which there was no relationship between the dependent and independent variables (response [e.g. regression slope] = 0) on one side of the space (in our case, on one side of the transect) and there is a strong relationship on the other side.



**Figure 8.** An example where the extent of autocorrelation is non-stationary, which might occur in a situation where dispersal is more limited on one end of the transect. This results in points clumps being more aggregated at small distances along the transect.

variable, may lead to improved spatial analyses, as they do for repeated-measures analyses.

Most of the techniques we have discussed assume isotropy (the effect of distance is independent of direction). When the effect of distance depends on direction (usually), this needs to be taken into account in the analysis. Spatial filters are designed to capture any form of clumping, but most other analyses need information on the form and direction of the autocorrelation. Dendritic systems usually have connections that are not well modeled by Euclidean distance (Peterson & Ver Hoef 2010). Those authors describe how to take into account different forms of connectivity (dispersal), but as with most of the papers reviewed here, they only treated autocorrelation in the residuals, and not in the predictor variables (pseudoreplication *sensu* Hurlbert, 1984). We can expect further advances in modeling anisotropic systems in the near future.

## Conclusions

### Where to go from here?

Conservation biologists want to use the most powerful method, and recent studies of spatial analyses conclude that applying some of the techniques they describe is better than doing nothing (Dormann *et al.* 2007; Bini *et al.* 2009; Beale *et al.* 2010). However, conservation biologists must be clear about their objectives. Spatial autocorrelation is generally advantageous for specific normative studies, because it permits land-use zoning and the inclusion of considerations relating to costs of land acquisition and control of access (Watts *et al.* 2009). Many of the most promising spatial methods in conservation biology described in the introduction do not involve statistical problems of autocorrelation in the residuals, which has been the focus of much of the recent debate. It would be foolish to try to remove the effect of spatial aggregation before undertaking these studies.

Although conservation biologists may be concerned about the possibility of unmeasured and unknown confounding variables (demonic intrusion) leading to spurious conclusions, this has not been the focus of most of the recent debate. Simulations were specifically designed to create autocorrelation in the residuals without collinearity between “space” and the independent (predictor) variables (De Knegt *et al.* 2010). If the researcher is worried about confounding variables, they should use techniques that model space in the dependent or independent variables. However, no particular advantage may be obtained in allocating variance between “space” and environment, because, at most spatial scales of interest to conservation biologists, “space” generally just represents unknown environmental variables in the analysis. If a specific process, such as reproduction or dispersal, is thought to cause autocorrelation, it may be better to model that process, rather than calling it “space.” We have focused on simple examples, and assumed that sampling was undertaken at the scale appropriate for the questions. However, autocorrelation in the residuals is

likely to be caused by sampling at a scale inappropriate to the question (De Knecht *et al.* 2010). In this case, removing autocorrelation from the residuals instead of using it to redefine the question will result in analyses that are as biased and inappropriate as OLS regression.

If the researcher can assume that “space” does not represent confounding variables, and only wants to carry out valid statistical tests and estimate parameters (that cannot also be variables), then spatial techniques that focus on the residuals are the most appropriate and may greatly improve estimates (Beale *et al.* 2010 and references therein). Although we agree with Beale *et al.* (2010) that nonstationarity of the autocorrelation in the residuals is a grave problem, we believe that, unlike the model misspecification described in the previous paragraph, it is not inherently unsolvable, and, it may be possible to use covariates to model the residual structure (Zuur *et al.* 2009). Where individual clumps can be recognized, analyses such as those described by Barnett *et al.* (2010) may lead to improved spatial analyses, as they do for repeated-measures analyses.

Recent studies in landscape ecology suggest that the configuration of landscape elements may be important in itself, and there may be nonlinear “threshold” effects (Metzger 2006). There has been only limited progress in landscape ecology because of the difficulty of replicating landscapes. Internal validation (such as standard statistical tests) assumes that the ecological relationships are well known (and generally linear) and can be extrapolated to other landscapes. However, real-world landscapes are generally so complex, and with so many nonlinear relationships, that extrapolation to other systems based on past knowledge of a particular system is risky because of the likelihood of essentially unpredictable phenomena (“black swans” in the terminology of Taleb 2007). Conservation biologists should seek more substantive replication (*i.e.* the repetition of the study by other researchers in other landscapes) in order to have confidence in their models.

## Acknowledgements

We are grateful to Pedro Peres-Neto, Luis Mauricio Bini, Alexandre Diniz-Filho, James Roper, Thiago Rangel and Tadeu Siqueira for constructive comments on the earlier versions of this manuscript. We regret that space did not allow us to include all of their valuable suggestions.

## References

- Austin M, 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. *Ecological Modelling*, 157:101-118. [http://dx.doi.org/10.1016/S0304-3800\(02\)00205-3](http://dx.doi.org/10.1016/S0304-3800(02)00205-3)
- Bahn V & McGill BJ, 2007. Can niche-based distribution models outperform spatial interpolation? *Global Ecology and Biogeography*, 16:733-742. <http://dx.doi.org/10.1111/j.1466-8238.2007.00331.x>
- Barnett AG *et al.*, 2010. Using information criteria to select the correct variance-covariance structure for longitudinal data in ecology. *Methods in Ecology and Evolution*, 1: 15-24. <http://dx.doi.org/10.1111/j.2041-210X.2009.00009.x>
- Beale CM *et al.*, 2007. Red herrings remain in geographical ecology: a reply to Hawkins *et al.* (2007). *Ecography*, 30:845-847. <http://dx.doi.org/10.1111/j.2007.0906-7590.05338.x>
- Beale, CM *et al.*, 2010. Regression analysis of spatial data. *Ecology Letters*, 13:246. PMID:20102373. <http://dx.doi.org/10.1111/j.1461-0248.2009.01422.x>
- Beguéria S & Pueyo Y, 2009. A comparison of simultaneous autoregressive and generalized least squares models for dealing with spatial autocorrelation. *Global Ecology and Biogeography*, 18:273-279. <http://dx.doi.org/10.1111/j.1466-8238.2009.00446.x>
- Betts MG *et al.*, 2006. The importance of spatial autocorrelation, extent and resolution in predicting forest bird occurrence. *Ecological Modelling*, 191:197-224. <http://dx.doi.org/10.1016/j.ecolmodel.2005.04.027>
- Betts MG *et al.*, 2009. Comment on “Methods to account for spatial autocorrelation in the analysis of species distributional data: a review”. *Ecography*, 32:374-378. <http://dx.doi.org/10.1111/j.1600-0587.2008.05562.x>
- Bini LM *et al.*, 2009. Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. *Ecography*, 32:193-204. <http://dx.doi.org/10.1111/j.1600-0587.2009.05717.x>
- Bini LM *et al.*, 2006. Challenging Wallacean and Linnean shortfalls: knowledge gradients and conservation planning in a biodiversity hotspot. *Diversity and Distributions*, 12:475-482. <http://dx.doi.org/10.1111/j.1366-9516.2006.00286.x>
- 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. [http://dx.doi.org/10.1016/S0304-3800\(01\)00501-4](http://dx.doi.org/10.1016/S0304-3800(01)00501-4)
- Borcard D *et al.*, 1992. Partialling out the spatial component of ecological variation. *Ecology*, 73:1045-1055. <http://dx.doi.org/10.2307/1940179>
- Chesson P, 2003. Understanding the role of environmental variation in population and community dynamics. *Theoretical Population Biology*, 64:253-254. <http://dx.doi.org/10.1016/j.tpb.2003.06.002>
- Colyvan M *et al.* 2009. Philosophical Issues in Ecology: recent trends and future directions. *Ecology and Society*, 14:22.
- De Knecht HJ *et al.*, 2010. Spatial autocorrelation and the scaling of species-environment relationships. *Ecology*, 91:2455-2465. PMID:20836467. <http://dx.doi.org/10.1890/09-1359.1>
- De Marco P *et al.*, 2008. Spatial analysis improves species distribution modelling during range expansion. *Biology Letters*, 4:577-580. PMID:18664417. PMID:2610070. <http://dx.doi.org/10.1098/rsbl.2008.0210>
- Diniz-Filho JAF *et al.* 2003. Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, 12:53-64. <http://dx.doi.org/10.1046/j.1466-822X.2003.00322.x>
- Diniz-Filho JAF *et al.*, 2007. Are spatial regression methods a panacea or a Pandora's box? A reply to Beale *et al.* (2007). *Ecography*, 30:848-851.
- Diniz-Filho JAF & Telles MPC, 2002. Spatial autocorrelation analysis and the identification of operational

- units for conservation in continuous populations. *Conservation Biology*, 16:924-935. <http://dx.doi.org/10.1046/j.1523-1739.2002.00295.x>
- Dormann CF *et al.*, 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30:629-628.
- Dray S *et al.*, 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, 196:483-493. <http://dx.doi.org/10.1016/j.ecolmodel.2006.02.015>
- Dutilleul P, 1993. Modifying the t test for assessing the correlation between two spatial processes. *Biometrics*, 49:305-314. <http://dx.doi.org/10.2307/2532625>
- Dutilleul P *et al.*, 2008. Modified F tests for assessing the multiple correlation between one spatial process and several others. *Journal of Statistical Planning and Inference*, 138:1402-1415. <http://dx.doi.org/10.1016/j.jspi.2007.06.022>
- Fortin M-J & Dale MRT, 2005. *Spatial Analysis: a guide for ecologists*. Cambridge: Cambridge University Press.
- Fortin M-J & Dale MRT, 2009. Spatial autocorrelation in ecological studies: a legacy of solutions and myths. *Geographical Analysis*, 41:392-397. <http://dx.doi.org/10.1111/j.1538-4632.2009.00766.x>
- Hawkins BA *et al.* 2007. Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology. *Ecography*, 30:375-384.
- Hubbell SP, 2005. Neutral theory in community ecology and the hypothesis of functional equivalence. *Functional Ecology*, 19:166-172. <http://dx.doi.org/10.1111/j.0269-8463.2005.00965.x>
- Hurlbert SH, 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, 54:187-211. <http://dx.doi.org/10.2307/1942661>
- Landeiro VL *et al.*, 2011. Spatial eigenfunction analyses in stream networks: do watercourse and overland distances produce different results? *Freshwater Biology*, 56:1184-1192. <http://dx.doi.org/10.1111/j.1365-2427.2010.02563.x>
- Legendre P, 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74:1659-1673. <http://dx.doi.org/10.2307/1939924>
- Legendre P *et al.*, 2005. Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs*, 75:435-450. <http://dx.doi.org/10.1890/05-0549>
- Legendre P *et al.*, 2008. Analyzing or explaining beta diversity? Comment. *Ecology*, 89:3238-3244. <http://dx.doi.org/10.1890/07-0272.1>
- Legendre P *et al.*, 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography*, 25:601-615. <http://dx.doi.org/10.1034/j.1600-0587.2002.250508.x>
- Legendre P & Legendre L, 1998. *Numerical ecology*. Amsterdam: Elsevier.
- Legendre P *et al.*, 2009a. PCNM: PCNM spatial eigenfunction and principal coordinate analyses. R Forge. R package version 1.9.
- Legendre P *et al.*, 2009b. Partitioning beta diversity in a subtropical broad-leaved forest of China. *Ecology*, 90: 663-674. PMID:19341137. <http://dx.doi.org/10.1890/07-1880.1>
- Lennon JJ, 2000. Red-shifts and red herrings in geographical ecology. *Ecography*, 23:101-113. <http://dx.doi.org/10.1111/j.1600-0587.2000.tb00265.x>
- Levins R, 1966. The strategy of model building in population biology. *American Scientist*, 54:421-431.
- Metzger JP, 2006. How to deal with non-obvious rules for biodiversity conservation in fragmented landscapes? *Natureza & Conservação*, 4:11-23.
- Mlodinow L, 2001. *Euclid's Window: The Story of Geometry from Parallel Lines to Hyperspace*. New York: Touchstone.
- Moilanen A *et al.*, 2008. A method for spatial freshwater conservation prioritization. *Freshwater Biology*, 53:577-592. <http://dx.doi.org/10.1111/j.1365-2427.2007.01906.x>
- Nams VO *et al.*, 2006. Determining the spatial scale for conservation purposes - an example with grizzly bears. *Biological Conservation*, 128:109-119. <http://dx.doi.org/10.1016/j.biocon.2005.09.020>
- Peres-Neto PR & Legendre P, 2010. Estimating and controlling for spatial structure in the study of ecological communities. *Global Ecology and Biogeography*, 19:174-184. <http://dx.doi.org/10.1111/j.1466-8238.2009.00506.x>
- Peres-Neto PR *et al.*, 2006. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, 87:2614-2625. [http://dx.doi.org/10.1890/0012-9658\(2006\)87\[2614:VPOSDM\]2.0.CO;2](http://dx.doi.org/10.1890/0012-9658(2006)87[2614:VPOSDM]2.0.CO;2)
- Peterson EE & Ver Hoef JM, 2010. A mixed-model moving-average approach to geostatistical modeling in stream networks. *Ecology*, 91:644-651. PMID:20426324. <http://dx.doi.org/10.1890/08-1668.1>
- Rangel TF *et al.*, 2006. Towards an integrated computational tool for spatial analysis in macroecology and biogeography. *Global Ecology and Biogeography*, 15:321-327. <http://dx.doi.org/10.1111/j.1466-822X.2006.00237.x>
- Rangel TF *et al.*, 2010. SAM: a comprehensive application for Spatial Analysis in Macroecology. *Ecography*, 33:46-50. <http://dx.doi.org/10.1111/j.1600-0587.2009.06299.x>
- Taleb NN, 2007. *The Black Swan*. New York: Random House.
- Taper ML & Lele SR, 2004. *The Nature of scientific of evidence*. Chicago: University of Chicago Press.
- Tuomisto H & Ruokolainen K, 2006. Analyzing or explaining beta diversity? Understanding the targets of different methods of analysis. *Ecology*, 87:2697-2708. [http://dx.doi.org/10.1890/0012-9658\(2006\)87\[2697:AOEBDU\]2.0.CO;2](http://dx.doi.org/10.1890/0012-9658(2006)87[2697:AOEBDU]2.0.CO;2)
- Watts ME *et al.*, 2009. Marxan with Zones: Software for optimal conservation based land- and sea-use zoning. *Environmental Modelling & Software*, 24:1513-1521. <http://dx.doi.org/10.1016/j.envsoft.2009.06.005>
- Zar JH, 1996. *Biostatistical analysis*. New Jersey: Prentice Hall.
- Zuur A *et al.*, 2009. Mixed effects models and extensions in Ecology with R. New York: Springer. <http://dx.doi.org/10.1007/978-0-387-87458-6>

Received: March 2010  
 First Decision: April 2010  
 Accepted: February 2011

