

## Spatial autocorrelation and the scaling of species–environment relationships

H. J. DE KNEGT,<sup>1,5</sup> F. VAN LANGEVELDE,<sup>1</sup> M. B. COUGHENOUR,<sup>2</sup> A. K. SKIDMORE,<sup>3</sup> W. F. DE BOER,<sup>1</sup>  
I. M. A. HEITKÖNIG,<sup>1</sup> N. M. KNOX,<sup>3</sup> R. SLOTOW,<sup>4</sup> C. VAN DER WAAL,<sup>1</sup> AND H. H. T. PRINS<sup>1</sup>

<sup>1</sup>*Resource Ecology Group, Wageningen University, P.O. Box 47, 6700 AA Wageningen, The Netherlands*

<sup>2</sup>*Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, Colorado 80523-1499 USA*

<sup>3</sup>*Department of Natural Resources, International Institute for Geo-Information Science and Earth Observation, P.O. Box 6, 7500 AA Enschede, The Netherlands*

<sup>4</sup>*School of Biological and Conservation Sciences, University of KwaZulu-Natal, Private Bag X54001, Durban 4000, South Africa*

**Abstract.** Issues of residual spatial autocorrelation (RSA) and spatial scale are critical to the study of species–environment relationships, because RSA invalidates many statistical procedures, while the scale of analysis affects the quantification of these relationships. Although these issues independently are widely covered in the literature, only sparse attention is given to their integration. This paper focuses on the interplay between RSA and the spatial scaling of species–environment relationships. Using a hypothetical species in an artificial landscape, we show that a mismatch between the scale of analysis and the scale of a species’ response to its environment leads to a decrease in the portion of variation explained by environmental predictors. Moreover, it results in RSA and biased regression coefficients. This bias stems from error–predictor dependencies due to the scale mismatch, the magnitude of which depends on the interaction between the scale of landscape heterogeneity and the scale of a species’ response to this heterogeneity. We show that explicitly considering scale effects on RSA can reveal the characteristic scale of a species’ response to its environment. This is important, because the estimation of species–environment relationships using spatial regression methods proves to be erroneous in case of a scale mismatch, leading to spurious conclusions when scaling issues are not explicitly considered. The findings presented here highlight the importance of examining the appropriateness of the spatial scales used in analyses, since scale mismatches affect the rigor of statistical analyses and thereby the ability to understand the processes underlying spatial patterning in ecological phenomena.

**Key words:** *landscape context; omitted variable bias; scale; spatial autocorrelation; spatial regression; spatially lagged predictor; species–environment relationships.*

### INTRODUCTION

Understanding the relationships between organisms and their environment is of paramount importance for understanding the mechanisms behind (spatial) variation in ecological phenomena (Currie 2007, de Knecht et al. 2007, 2008, McIntire and Fajardo 2009). Critical to this understanding are issues of pattern and scale (Levin 1992), because two general characteristics inherent to ecological data can complicate analyses of species–environment relationships. First, ecological phenomena are often spatially autocorrelated, leading to problematic statistical inference if left unaccounted for (Cliff and Ord 1981, Legendre 1993, Dormann et al. 2007). Second, organismal response to environmental cues hinges on the scales that individuals can perceive and respond to, leading to sensitivity in the quantification of species–environment relationships to the spatial per-

spectives chosen (Levin 1992, Wu 2007, Mayor et al. 2009). Hence, the importance of spatial autocorrelation (SAC) and spatial scale in the study of species–environment relationships has stimulated much research over the past decade. Unfortunately, integration between these fields has been limited, although different processes may create SAC at different scales depending on the scales of an organism’s response to its environment (Wagner 2004, Wagner and Fortin 2005). Here, we attempt to facilitate the integration between these important issues, as we argue that their interaction offers possibilities to achieve a more thorough understanding of species–environment relationships.

Ecological data may exhibit SAC due to “endogenous” (or “inherent”) community or demographic processes (e.g., dispersal, conspecific attraction), or spatial dependence of organisms to the underlying environmental conditions that are spatially structured (i.e., “exogenous” or “induced” SAC; Cliff and Ord 1981, Legendre 1993). If the sources of SAC are not fully accounted for in analyses (due to failure to include an important environmental driver, inadequate capture of

Manuscript received 27 July 2009; revised 6 November 2009; accepted 9 November 2009. Corresponding Editor: H. H. Wagner.

<sup>5</sup> E-mail: hdeknecht@hotmail.com

its nonlinear effect, or failure to account for endogenous processes), the unexplained spatial pattern will appear in the residual errors, leading to residual spatial autocorrelation (RSA). Consequently, the assumption of independently and identically distributed (i.i.d.) errors common to most statistical procedures is violated, creating biased Type I error estimates due to inflation of degrees of freedom (Clifford et al. 1989, Legendre et al. 2002). Moreover, parameter estimates may be biased or their sign even inverted (Lennon 2000, Kühn 2007, Bini et al. 2009).

Recent studies analyzing the scale sensitivity of species–environment relationships have formed the idea that species have “characteristic scales” of response to their environment (Dormann and Seppelt 2007). These studies typically analyze the importance of landscape characteristics by regressing response data against landscape variables measured at various spatial scales (i.e., ambit radii) around sampling locations (e.g., Van Langevelde 2000, Steffan-Dewenter et al. 2002, Holland et al. 2004, Mayor et al. 2007). Hence, the influence of the scale of landscape context (i.e., the characteristics of the landscape surrounding a site) on the phenomenon under study is being investigated (Brennan et al. 2002). However, these studies generally do not analyze neither the spatial structure of the environmental predictors (Dormann and Seppelt 2007) nor the spatial structure in the model residuals (i.e., RSA).

Yet, since organisms respond to environmental characteristics at specific (but often unknown) scales, the spatial patterns resulting from a species’ response to extraneous predictors may differ from the spatial structure of the landscape (Wagner and Fortin 2005). This potentially leads to RSA and problematic inference when the scale of analysis does not match the scale at which the focal species responds to its environment (Gotway and Young 2002). Thus, in order to account for the sources of SAC in ecological phenomena, scaling issues need to be considered. Moreover, because understanding the mechanisms that maintain the observed spatial patterns critically depends on our ability to decompose the spatial pattern into the contributions of different processes affecting it (Legendre et al. 2009), understanding the interactions between SAC and the spatial scaling of species–environment relationships is urgently required. Although several authors have studied the influence of the scale of environmental SAC on analyses of organismal response (e.g., Lennon 2000, Beale et al. 2007, Diniz-Filho et al. 2007, Hawkins et al. 2007), analyses regarding the influence of the scaling of a species’ response to exogenous predictors on RSA have not yet been carried out.

This paper therefore focuses on the interplay between RSA and the spatial scaling of species–environment relationships. We use a hypothetical species and artificial environmental data to examine this interplay by making statistical issues explicit. The use of artificial data allows us to vary, independently, the scale of analysis, the scale

of the species’ response, and the scale of environmental heterogeneity, as well as to control the relationship between the species and its environment. We frame our analyses around three central issues: (1) the link between RSA and the spatial scale of analysis, given the scale of a species’ response and the scale of environmental heterogeneity, (2) the consequences of a scale mismatch for parameter estimation in regression analyses, and (3) the robustness of several spatial regression methods, devised to account for the effects of RSA, when analyzing data at incorrect scales.

## METHODS

### *A virtual data set*

We used the data set of Dormann et al. (2007), which contains a regular grid with 1108 cells and two artificial explanatory variables: rain and jungle cover, hereafter referred to as  $R$  and  $D$ , respectively. The predictor variables are based on an elevation model of the Maunga Whau Volcano in New Zealand, where  $R$  is highly dependent on elevation (including a rain shadow in the east) and thus strongly autocorrelated, whereas  $D$  is dominated by a high noise component (Fig. 1; Dormann et al. 2007). The two variables are uncorrelated (Pearson’s  $r = 0.013$ ,  $n = 1108$ ,  $P = 0.668$ ), thus avoiding model instability due to correlated predictors. Both predictors were normalized to zero mean and unit variance prior to analyses.

We distributed a hypothetical species over this landscape based on the two predictor variables, while setting the spatial scale of its response. We did this by averaging the predictor variables in a circular focal neighborhood (or moving window) centered on each grid cell, with ambit radius or buffer size  $f$ . Hence, we refer to “scale” here as the radius within which the predictor variables are measured, thus being a measure of the area or inference space represented by each data point.

To simplify the modeling process and interpretation, we distributed the virtual species based on  $R$  with  $f=0$ , thus using only local grid cell information; however, its response to  $D$  was modeled using a radius of three cells ( $D_f$ , with  $f=3$ ; Fig. 1). Spatial scaling of the species’ response to  $D$  is not associated to a specific spatial causation here, but one could interpret the species’ response to  $D$  in relation to, for instance, proximity to nesting sites, risk contagion due to proximity of habitats associated to predators, or the ability to detect predators.

The abundance ( $y$ ) of the virtual species in each grid cell was modeled as

$$y = \beta_0 + \beta_1 R + \beta_2 D_f + \varepsilon \quad \varepsilon \sim \mathcal{N}(0, \sigma^2) \quad (1)$$

where  $\beta_0 = 15$ ,  $\beta_1 = 1$ ,  $\beta_2 = -4.29$ ,  $f = 3$ , and  $\varepsilon$  is a spatially uncorrelated (i.i.d.) Gaussian error term (“white” noise). We choose the parameter values in such manner that both predictors exerted equal influence

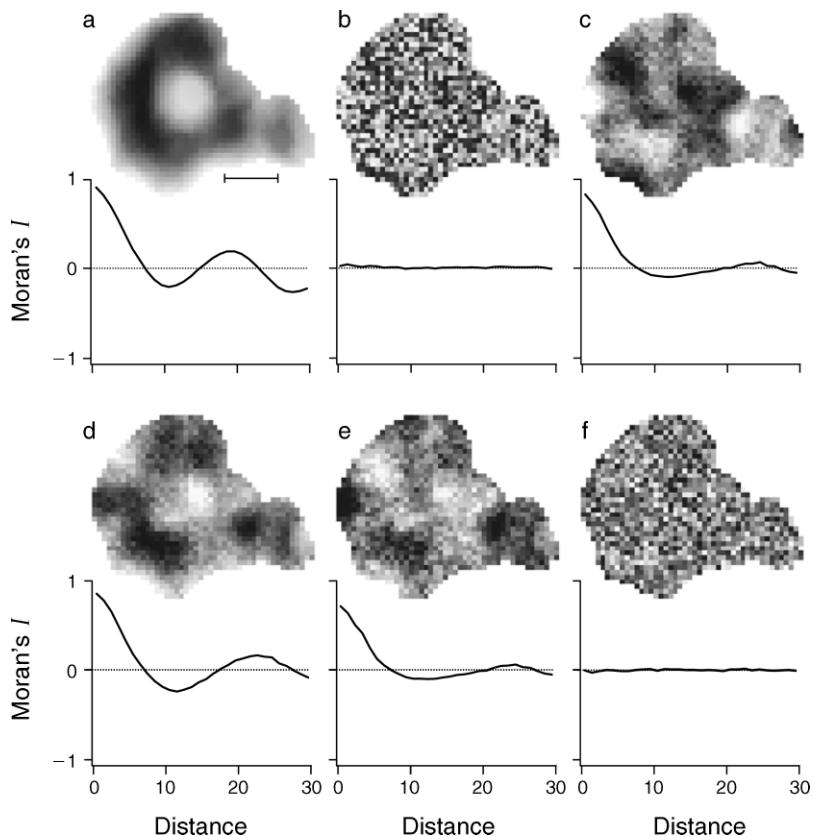


FIG. 1. Maps and corresponding correlograms for the predictor and response variables: (a) rain ( $R$ ); (b) jungle cover ( $D$ ); (c)  $D$  averaged within a circular focal neighborhood with a radius of three distance units (i.e.,  $D_f$  with  $f=3$ ); (d) simulated density of our hypothetical species ( $y$ ; Eq. 1 and 2); (e) residuals of an ordinary least-squares (OLS) model using only local information (i.e.,  $R$  and  $D$ ); (f) residuals of an OLS model using local and contextual information at the correct scales (i.e.,  $R$  and  $D_f$  with  $f=3$ ). The scale bar in panel (a) depicts 10 distance units.

on the response variable (i.e., equal standardized coefficients). We conducted analyses where we changed the signal-to-noise ratio, through varying the variance of the error term relative to the variance of the deterministic part of Eq. 1, such that (error  $\sigma$ ): (deterministic  $\sigma$ ) = 0.10, 0.25, or 0.50.

#### Analyses

We analyzed the abundance of our virtual species by regressing it against the two environmental predictor variables using ordinary least-squares (OLS) regression, and compared the parameter estimates to their “true” values for different scenarios. We used all grid cells in the statistical analyses. The spatial patterns in the residuals were examined using residual maps and correlograms that plot Moran’s  $I$  coefficients (e.g., Fortin and Dale 2005) as a function of separation distance between paired observations up to a distance of 30 distance units (Fig. 1). All analyses were carried out in the statistical software R (R Development Core Team 2009), using the libraries *spdep* (Bivand 2009), *ncf* (Bjornstad 2009), *nlme* (Pinheiro et al. 2008), and *RandomFields* (Schlather 2008).

#### Spatial scaling and RSA

In order to make the link between SAC and the spatial scaling of species–environment relationships explicit, we write Eq. 1 equivalently as a spatial cross-regressive model. Essentially, Eq. 1 is a nonspatial regression model augmented by a cross-regressor in the form of a spatially lagged explanatory variable:

$$y = \beta_0 + \beta_1 R + \beta_3 D + \lambda \mathbf{W}D + \varepsilon \quad (2)$$

where  $\lambda$  is the regression coefficient of the spatially lagged predictor  $D$  connected through the spatial weight matrix  $\mathbf{W}$  (here a row-standardized binary contiguity matrix with non-zero elements if the distance between grid cell centroids is less than or equal to three distance units),  $\beta_3 = \beta_2/(n+1) \approx -0.17$  and  $\lambda = n\beta_2/(n+1) \approx -4.12$ , where  $n$  is the number of neighborhood cells. Hence, the species’ response to its environment does not only depend on the characteristics within the grid cell ( $R$  and  $D$ , hereafter referred to as “local” environmental characteristic), but also on the characteristics of the surrounding landscape regarding  $D$  up to a distance of three distance units (i.e.,  $\mathbf{W}D$ ; the “landscape context” or neighborhood). We can thus refer to  $\beta_3$  and  $\lambda$  as the,

respectively, local and contextual landscape effects regarding  $D$ .

Hence, correlating the distribution of our virtual species to the predictor variables using only local information leads to the omission of the effect of landscape context ( $\mathbf{WD}$ ). Thus, the actual regression becomes

$$y = \beta_0 + \beta_1 R + \beta_3 D + \mu$$

where

$$\mu = \varepsilon + \lambda \mathbf{WD}. \quad (3)$$

Consequently, the statistical model is misspecified, because it is inconsistent with the data-generating process (Eq. 1 and 2). The error term  $\mu$  contains the omitted variable ( $\mathbf{WD}$ ), thus any spatial pattern in  $\mathbf{WD}$  will be reflected in the residual errors, leading to RSA. Alternatively, analyzing the distribution of our virtual species in relation to both predictors while including the effect of landscape context regarding  $D$ , but at a scale larger than the species' scale of response to  $D$ , leads to spatial smoothing of data and hence RSA (Gotway and Young 2002, Keitt et al. 2002, Diniz-Filho et al. 2003, Dormann et al. 2007).

Thus, a mismatch between the spatial scale of analysis and the scale of organismal response results in misspecification of the statistical model and RSA. Hence, we expect that plotting the level of RSA as function of the scale of analyses provides a clue regarding the characteristic scale of the species' response to its environment: that is the scale where the level of RSA is minimized. We therefore analyzed the distribution of our virtual species in relation to both  $R$  and  $D$ , using only local information regarding  $R$ , while varying the buffer size for predictor  $D$  from 0 up to six distance units. At each scale, the predictors  $D_f$  (with  $f = [0-6]$ ) are uncorrelated with  $R$  (all Pearson's  $|r| < 0.035$  and  $P > 0.2$ ,  $n = 1108$ ). Following Kissling and Carl (2008), we quantified the total level of RSA ( $\text{RSA}_{\text{tot}}$ ) as the summation of the absolute Moran's  $I$  values of the correlogram up to a distance of 30 distance units. Furthermore, we also analyzed Akaike's information criterion (AIC) values for the different scales. Note that the analyses carried out at different scales were not independent tests, but served to determine at which scale  $\text{RSA}_{\text{tot}}$  and AIC were minimized. We replicated these analyses 1000 times, each time with a different realization of the error term  $\varepsilon$ .

#### *Scale effects on parameter estimation*

In spatially structured landscapes, the presence of spatially lagged explanatory predictors can induce correlation between predictors (in our case, between  $D$  and  $\mathbf{WD}$ ), even if the environmental predictors themselves ( $R$  and  $D$ ) are uncorrelated (Haining 2003). This could encumber analyses and threaten their statistical and inferential interpretation (Graham 2003). Thus, besides resulting in RSA, model misspecification due to

the erroneous omission of spatially lagged predictors is a typical example of the omitted variable problem in a spatial context (Florax and Folmer 1992). Consequently, the OLS estimator of the regression coefficients may be biased, the residual variance overestimated, and inference procedures invalid (Florax and Folmer 1992, Anselin et al. 2004).

In the case of our virtual species, the local information regarding  $D$  and its spatially lagged counterpart  $\mathbf{WD}$  are only moderately correlated (Pearson's  $r = 0.09$ ,  $n = 1108$ ,  $P = 0.002$ ). However, as even low levels of correlation between predictors can bias analyses (Graham 2003), we expect that analyzing our virtual species in relation to  $R$  and  $D$  while omitting  $\mathbf{WD}$  results in an over-estimation of the regression coefficient for  $D$ .

Moreover, the level of correlation between local and spatially lagged predictors depends on the scale at which a species responds to the environmental predictors and is modified by the scale of SAC of the landscape variable (e.g., Gotway and Young 2002, Wong 2009). Hence, the bias that results from omitting a spatially lagged predictor depends on both the scale of environmental SAC and the scale of the species' response to the environmental predictor. To quantify this bias, we simulated landscapes similar in function as  $D$ , but with a varying spatial scale of SAC, which we refer to as the dominant scale (DS) of landscape heterogeneity (e.g., de Knecht et al. 2008). Then, through omitting the spatially lagged predictor and comparing the estimated regression parameters to their "true" values, we quantified the influence of the interaction between the DS of landscape heterogeneity and the scale of organismal response on the regression analyses. We simulated landscapes with a DS of 0.5, 2, and 10 distance units, and a virtual species whose scale of response ( $f$ ) to the simulated landscapes ranged from 1 to 6 distance units (with an i.i.d. error term with  $[\text{error } \sigma]:[\text{deterministic } \sigma] = 0.1$ ). For each combination of DS and  $f$ , we quantified: (1) the correlation between local and contextual information; (2) the difference in AIC values ( $\Delta\text{AIC}$ ) between a spatial model with landscape context at the correct scale and a nonspatial model that uses only local information; (3) the difference in regression coefficients between the nonspatial and spatial OLS models ( $\Delta\beta = \beta_{\text{OLS}_{\text{ns}}} - \beta_{\text{OLS}_{\text{s}}}$ ); and (4) the total level of RSA when omitting landscape context as measured by  $\text{RSA}_{\text{tot}}$ . All analyses were iterated 1000 times, with different realizations of the simulated landscapes and error term.

#### *Dealing with RSA resulting from a scale mismatch*

Ecologists facing RSA commonly rely on regression-based approaches that are intended either to live with the problem (e.g., by spacing sampling locations further apart, adjusting the degrees of freedom, or adjusting the effective sample size; see, e.g., Dutilleul 1993, Holland et al. 2004, Fortin and Dale 2005), or to model the spatial process causing the autocorrelation as part of the regression analysis. The latter approach allows a

correction of the parameter estimates, and resulted in the development of various forms of spatial regression that are increasingly becoming part of the standard toolbox for ecologists. Given their widespread use, we tested several of these techniques for their reliability in estimating the species–environment relationships when analyzing the distribution of our hypothetical species. We confined ourselves to two techniques widely used in ecological studies: methods based upon simultaneous autoregression (SAR) and generalized least squares (GLS); and a relatively new and emerging spatial approach belonging to the class of eigenvector-based spatial filtering techniques: spatial eigenvector mapping (SEVM). We refrain from technical discussion of these techniques here and only provide a nontechnical synopsis, since many of them have received exhaustive review, comparison and discussion elsewhere (Cliff and Ord 1981, Keitt et al. 2002, Fortin and Dale 2005, Schabenberger and Gotway 2005, Griffith and Peres-Neto 2006, Dormann 2007, Dormann et al. 2007, Kissling and Carl 2008, Anselin 2009, Bini et al. 2009).

SAR models operate with spatial weight matrices (viz.  $W$  in Eq. 2) that specify the strength of interaction between neighboring sites. Depending on where this spatial interaction is thought to occur, SAR models specify the relationship between the response variable (SAR-lag) or residual errors (SAR-error) at each location and those at neighboring locations (Anselin 2002, 2009, Schabenberger and Gotway 2005, Dormann et al. 2007). As demonstrated by Kissling and Carl (2008), the performance of SAR models depends on the neighborhood distance and coding styles of the spatial weight matrices used. Because it is difficult to decide a priori which neighborhood structure is most efficient, Kissling and Carl (2008) recommended to test a wide variety of SAR model specifications, and to identify the best model based on AIC and  $RSA_{tot}$ . Hence, we fitted both SAR-lag and SAR-error models using five different neighborhood sizes (1 to 3 distance units with increments of 0.5) and three coding styles (binary, row standardized, and variance stabilized), and used both AIC and  $RSA_{tot}$  to evaluate these model specifications.

Methods based on GLS include spatial interaction in the regression model by incorporating SAC explicitly into the variance–covariance structure, assuming a parametric correlation function as estimated from a semi-variogram of the OLS residuals (Keitt et al. 2002, Dormann et al. 2007). Three frequently used correlation functions are exponential, Gaussian and spherical representations (Dormann et al. 2007). We analyzed our data set using these three representations, and evaluated their performance using AIC and  $RSA_{tot}$ .

Eigenvector-based spatial filtering techniques seek to avoid the complications involved in estimating autoregressive parameters, as well as to exploit established OLS theory. They are based on the eigenfunction decomposition of spatial connectivity matrices, either binary or distance based (Griffith and Peres-Neto 2006,

Dormann et al. 2007). The resulting spatial filters translate the spatial arrangement of data points into explanatory variables that capture spatial effects at different scales, which can be included in regression analyses to capture the dependencies among the residuals (Borcard and Legendre 2002, Borcard et al. 2004, Diniz-Filho and Bini 2005, Griffith and Peres-Neto 2006, Tiefelsdorf and Griffith 2007). We used the distance-based eigenvector procedure as described by Griffith and Peres-Neto (2006) and Dormann et al. (2007), and included eigenvectors as spatial predictors into the linear model until RSA was no longer significant at  $\alpha = 0.05$ .

We used these spatial regression methods to analyze the distribution of our virtual species when only using local information (Eq. 3). We iterated the analyses of the SAR and GLS models 1000 times, each time with a different realization of the error term (with [error  $\sigma$ ]:[deterministic  $\sigma$ ] = 0.10). However, the analyses using SEVM were iterated 100 times, since these were computationally intensive. We compared the parameter estimates to their “true” values, assessed model fit using AIC, and checked for residual SAC using  $RSA_{tot}$ . Moreover, we tested how the different spatial methods behave in case of a scale mismatch under varying DS and  $f$ , with DS = 0.5, 2, and 10 distance units and  $f$  ranging from 2 to 6 distance units, while analyzing the deviance of the estimated regression coefficients from their true values ( $\Delta\beta$ ). We iterated each combination of DS and  $f$  500 times for the SAR and GLS models, yet 15 times for SEVM, due to its time-consuming computations.

## RESULTS

### *Spatial scaling and RSA*

When omitting  $WD$ , only 51% of the variation was explained by the predictor variables (Appendix), and the residuals exhibited strong SAC (Table 1, Fig. 1e). Because  $WD$  is uncorrelated with  $R$  (Pearson's  $r = 0.034$ ,  $n = 1108$ ,  $P = 0.261$ ) the OLS estimator was unbiased regarding the influence of  $R$  ( $\beta_1 \approx 1$ ; Table 1). However, since  $WD$  is correlated to  $D$ , the erroneous omission of  $WD$  led to an overestimation of the influence of  $D$ :  $\beta_3 \approx -0.25$  (cf.  $-0.17$ ; Eq. 2, Table 1).

Including both  $R$  and  $D$  in the analysis, while varying the scale of analysis for  $D$ , showed a pronounced scale dependency of  $RSA_{tot}$  (Fig. 2). The scale where  $RSA_{tot}$  was minimized corresponds to the scale at which the virtual species was set to respond to  $D$ , i.e., a buffer size of three cells. Only when analyzing the distribution of our virtual species at this scale could 99% of the variation be explained by the predictor variables (Appendix), while yielding the i.i.d. noise that we included in Eq. 1 and 2 (Fig. 1f), as well as the correct estimates of regression coefficients (Table 1). A scale mismatch thus not only induced RSA, but also reduced overall model fit as measured by AIC (Fig. 2).

A larger contribution of the noise term led to lower levels of  $RSA_{tot}$  in case of a scale mismatch (Fig. 2),

TABLE 1. Summary statistics of the regression models analyzed using 1000 iterations (except for SEVM: 100 iterations).

Model	$\beta_0$	$\beta_1$	$\beta_3$
Spatial OLS	15.00*** $\pm$ 0.00013	1.00*** $\pm$ 0.00014	-0.165*** $\pm$ 0.00013
Nonspatial OLS	14.99*** $\pm$ 0.00013	0.97*** $\pm$ 0.00014	-0.253*** $\pm$ 0.00013
GLS	15.05*** $\pm$ 0.00026	1.00*** $\pm$ 0.00048	-0.007ns $\pm$ 0.00015
SAR-error	15.00*** $\pm$ 0.00020	0.93*** $\pm$ 0.00055	-0.014ns $\pm$ 0.00015
SAR-lag	1.18*** $\pm$ 0.00120	0.15*** $\pm$ 0.00009	-0.033† $\pm$ 0.00014
SEVM	14.99*** $\pm$ 0.00043	0.80*** $\pm$ 0.00090	0.019† $\pm$ 0.00041

Notes: The unstandardized regression coefficients (mean  $\pm$  SE) refer to Eq. 2, and  $I$  is the Moran's  $I$  coefficient for the first distance class. The top row shows the regression parameters for the correct model (i.e., landscape context included at the correct scale), whereas the other rows depict models that use only local information. The significance levels are based on the number of iterations yielding significant effects (with  $\alpha = 0.05$ ). Abbreviations are: AIC, Akaike information criterion;  $RSA_{tot}$ , total level of residual spatial autocorrelation.

†  $P < 0.1$ ; \*\*\*  $P < 0.001$ ; ns, not significant.

since it was set to be i.i.d. However,  $RSA_{tot}$  and AIC showed similar patterns in their dependency on the scale of analysis in qualitative terms. The level of  $RSA$  as measured by  $RSA_{tot}$  showed no dependency on the signal-to-noise ratio when our virtual species was analyzed at the correct scale.

Scale effects on parameter estimation

The correlation between the local and contextual information increased with increasing DS of landscape heterogeneity, yet decreased with increasing scale of a species' response (Fig. 3). This was due to the increase in environmental SAC with increasing DS, while the similarity between observations separated in space decreased with increasing  $f$ . Consequently, the difference in model fit between a spatial and nonspatial model ( $\Delta AIC$ ) showed exactly the opposite pattern, because erroneously omitting landscape context resulted in less information loss with increasing DS or decreasing  $f$ .

The bias in regression coefficient ( $\Delta\beta$ ) when erroneously omitting the influence of landscape context resulted from a trade-off between the level of correlation between local and contextual information on the one hand, and the relative influence of the omitted variable in determining the response variable on the other hand. With increasing  $f$ , the importance of landscape context relative to the focal cell increased, whereas the level of correlation decreased. Hence, since the level of correlation between local and contextual characteristics was positively linked to DS,  $\Delta\beta$  showed a hump-shaped response relative to  $f$ , and increased with increasing DS (Fig. 3).

Dealing with RSA resulting from a scale mismatch

For the SAR models, a neighborhood structure with row-standardized coding for a distance of 1.5 cells (a "queen" contiguity matrix) explained the data best, yielding the lowest AIC and  $RSA_{tot}$  values for all tested specifications. For the GLS models, a spherical relationship between the error term and geographical distance gave the best performance as measured by AIC and  $RSA_{tot}$ . Below, we report only the results using the best configuration for each modeling approach. The results are summarized in Table 1.

All spatial models were able to reduce  $RSA$  to nonsignificant levels and simultaneously yielded lower AIC values than the OLS model when omitting  $WD$ . However, they also showed bias (i.e., underestimation) in the estimated magnitude of the effect of  $D$  on the abundance of our virtual species, with SEVM even

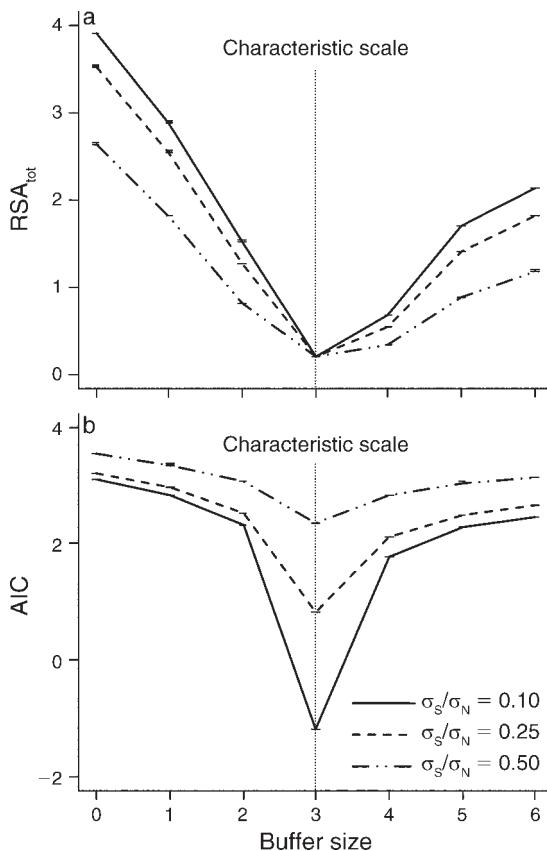


FIG. 2. Model diagnostics as function of the spatial scale used while regressing the species' distribution against  $D$ : (a)  $RSA_{tot}$ , i.e., the summation of absolute Moran's  $I$  values up to a distance of 30 units; and (b) Akaike's information criterion (AIC) values. The different lines represent different signal-to-noise ratios ( $\sigma_S/\sigma_N$ ). Means ( $\pm$ SE) are shown based on 1000 iterations. The vertical dotted lines represent the scale at which the virtual species was set to respond to its environment.

TABLE 1. Extended.

$\lambda$	AIC	RSA <sub>tot</sub>	<i>I</i>
-4.11*** ± 0.0173	-1105.6 ± 1.49	0.19 ± 0.00089	-0.003 <sup>ns</sup> ± 0.00049
	3100.3 ± 0.29	3.90 ± 0.00106	0.710*** ± 0.00009
	1500.0 ± 0.75	0.44 ± 0.00067	-0.079 <sup>ns</sup> ± 0.00015
	1547.6 ± 0.72	0.34 ± 0.00046	-0.071 <sup>ns</sup> ± 0.00013
	1584.4 ± 0.70	0.30 ± 0.00047	-0.057 <sup>ns</sup> ± 0.00013
	1391.3 ± 2.22	0.32 ± 0.00201	0.006 <sup>ns</sup> ± 0.00028

estimating a positive relationship as opposed to the true negative relationship. Although most tested methods yielded appropriate estimates of the intercept and the effect of *R* on the species' abundance, application of SAR-lag yielded shifts in both estimates.

Analyses of the distribution of our virtual species, with varying *f* and DS, in relation to only local information using the spatial methods showed that the bias in regression coefficients ( $\Delta\beta$ ) varied between the different methods used, and depended on both *f* and DS for GLS and SEVM (Fig. 4). The SAR models performed relatively well in landscapes with large DS, yet, in relative terms, their estimates were still more than

30% off. Note that comparing the different spatial methods to the results of a non-spatial OLS drastically changes the results, yet both the nonspatial OLS as well as the tested spatial regression methods essentially used a mis-specified model to estimate the species–environment relationships.

DISCUSSION

In this paper, we have focused on the influence of a scale mismatch on the estimation of species–environment relationships, as it is important to understand the way such analyses are affected by the use of data at inappropriate scales (Gotway and Young 2002). Our

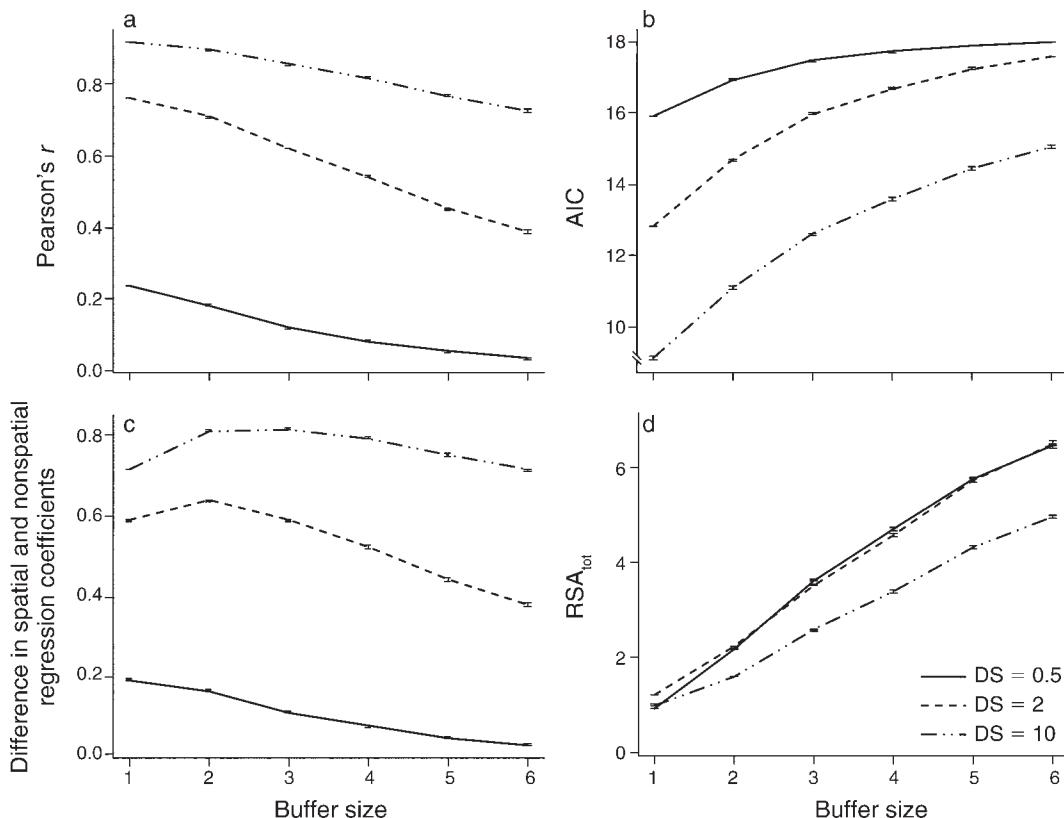


FIG. 3. Interactions between the spatial scale of a species' response (buffer size) and the spatial scale of spatial autocorrelation (SAC) in the environmental predictor (DS): (a) the correlation (Pearson's *r*) between local and contextual information; (b) the difference in Akaike's information criterion (AIC) values between a spatial (i.e., with landscape context at the correct scale) and nonspatial (i.e., using only local information) model; (c) the difference in regression coefficients between a spatial and nonspatial model; and (d) the total level of residual spatial autocorrelation (RSA<sub>tot</sub>) when omitting landscape context. Means (±SE) are shown based on 1000 iterations.

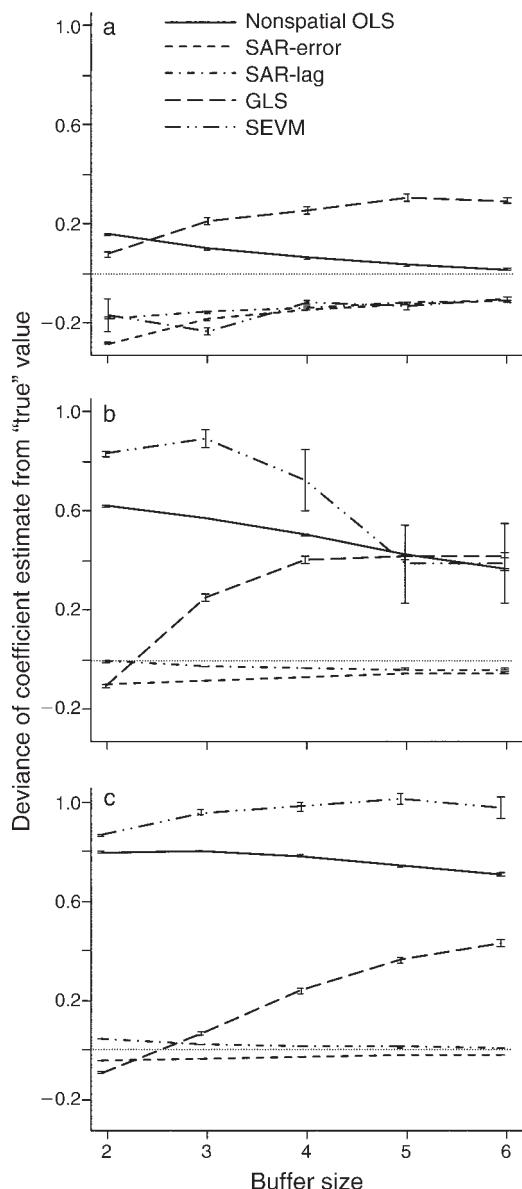


FIG. 4. Deviance of the coefficient estimates from their "true" values, for different models using only local information, as function of the scale of a species' response, and for landscapes with a scale of autocorrelation (DS) of (a) 0.5, (b) 2, and (c) 10 distance units. Means ( $\pm$ SE) are shown based on 500 iterations (except SEVM: 15 iterations). Abbreviations for methods: OLS, ordinary least squares; SAR, simultaneous autoregression; GLS, generalized least squares; SEVM, spatial eigenvector mapping.

analyses show that a scale mismatch leads to a reduction in the part of variation explained by landscape predictors and induces RSA. Although RSA is often seen as problematic, it implies structure in the residuals and therefore information about the processes not captured by the current model (Haining 2003, Fortin and Dale 2005, McIntire and Fajardo 2009). Thus, RSA is something one might not want to discard or correct

for: the problem is not its presence, but the absence of an explanation (Cliff and Ord 1981, Dormann 2007). Unfortunately, while RSA can be quantified, its origins cannot directly be identified: it may be the end product of an amalgam of interacting processes, with different processes creating patterns that may be observationally equivalent (Wagner and Fortin 2005, Anselin 2009).

#### *Inferring relevant scales from residuals*

As shown in Eq. 3, analyzing response data erroneously using only local information is essentially a type of model misspecification due to an omitted variable problem. This omitted variable, **WD** in case of our virtual species, contains two scale components: (1) the spatial structure of the landscape predictor *D* that is included at the wrong scale and (2) the spatial scale of the species' response to the predictor as captured by **W**. Hence, the pattern of RSA is the result of the interaction between these scale components, thus the scale of species-environment relationships cannot directly be inferred from spatial patterns in the residuals alone (Wagner and Fortin 2005, Dormann 2009), even when RSA exclusively stems from a scale mismatch.

However, because RSA can bear the fingerprint of a scale mismatch, varying the scale of analysis while analyzing RSA can reveal the scale of the species' response (Fig. 2). Using scales larger or smaller than the scale of a species' response leads to spatial smoothing or the omission of landscape context, respectively. While the influence of spatial smoothing on RSA has been noted by other authors (Gotway and Young 2002, Keitt et al. 2002, Diniz-Filho et al. 2003, Dormann et al. 2007, Tiefelsdorf and Griffith 2007), the influence of omitting landscape context has received only sparse attention (but see Florax and Folmer 1992, Wagner and Fortin 2005). In the following, we will therefore focus on this issue.

#### *Parameter shifts*

Besides leading to RSA, erroneously ignoring landscape context leads to a biased OLS estimator. However, this is not the result of RSA, but due to the non-zero covariance between regressor (*D*) and error ( $\mu$ ; Eq. 3) stemming from the correlation between local (*D*) and contextual (**WD**) information (Table 1, Fig. 3). This leads to the violation of one of the assumptions of regression analyses (i.e., zero covariance between the explanatory variables and error term), and therefore to overestimation of the parameter of interest and henceforth faulty conclusions (Ebbes et al. 2005, Luskin 2008). Note that OLS residuals (but not the errors) are, by definition, uncorrelated to the predictors; this violation is thus difficult to diagnose. Moreover, the magnitude of the regressor-error correlation is dependent on the scale of landscape heterogeneity, as well as on the scale of the species' response (Fig. 3).

While several authors have argued that RSA may well bias coefficient estimation (e.g., Lennon 2000, Dormann

2007, Kühn 2007), other studies found nonspatial OLS models to be robust and unbiased (Diniz-Filho et al. 2003, Beale et al. 2007, Hawkins et al. 2007). Our findings support the latter view, since we found unbiased coefficient estimates under RSA for predictors uncorrelated to the error term. This is consistent with statistical literature (Cressie 1991, Legendre et al. 2002, Schabenberger and Gotway 2005, Hawkins et al. 2007, Tiefelsdorf and Griffith 2007). Note, however, that there is no doubt that RSA inflates the chance of Type I errors, so that coefficients obtained by OLS are not minimum variance estimators (Fortin and Dale 2005, Hawkins et al. 2007).

#### *Dealing with RSA resulting from a scale mismatch*

Since RSA due to a scale mismatch is an issue of data analysis, constrained by the resolution and extent of the data, analyzing the data at appropriate scales suffices as remedial action against RSA and the error–regressor dependency. This is where our simulations diverged from those of other studies on the influence of SAC on regression analyses (e.g., Dormann et al. 2007, Kissling and Carl 2008). These studies typically focus on endogenous SAC by adding an aggregation mechanism to the error term such that it is uncorrelated with the predictor variables (Dormann 2009), and conclude that making a correction for RSA through applying one of the spatial regression methods is important since these show good type I error control and precise parameter estimation (e.g., Keitt et al. 2002, Dormann et al. 2007, Kissling and Carl 2008).

However, although all our tested spatial methods yielded a better model fit than nonspatial OLS and decreased RSA to insignificant levels, parameter estimation turned out to be problematic in case of a scale mismatch. All tested methods underestimated the influence of  $D$ , both in terms of regression coefficient as well as significance. Why our analyses yielded biased coefficient estimates has a dual explanation. First, GLS and SEVM, like OLS, suffer from space–environment confounding when the error term is correlated with a regressor (Ebbes et al. 2005, Griffith and Peres-Neto 2006, Ayinde 2007, Hawkins et al. 2007, Kissling and Carl 2008, Luskin 2008, Betts et al. 2009). Second, the spatial methods test for marginal effects of environmental predictors after controlling for SAC due to an unknown spatial process, thereby leading to a reduction in the strength of environmental effects if the response is controlled by exogenous predictors (Segurado et al. 2006, Currie 2007).

Our simulations thus emphasize what several authors (Lennon 2000, Haining 2003, Wagner 2004, Anselin 2006, Van Teeffelen and Ovaskainen 2007, Dormann 2009) warned against: using “rough and ready” methods to improve model fit can yield rough and ready answers that may be quite wrong. Moreover, even the most advanced and computer-intensive statistical procedures are no guarantee for improving our understanding of

ecological responses, as such methods often do not give straightforward information about the underlying processes (Borcard et al. 2004, Dormann et al. 2007). Thus, spatial regression methods should not be used as a quick fix for modeling spatial data: as the species–environment relationships are scale dependent and this is not incorporated in the analysis, this may be a more severe threat to the interpretation of RSA than an inflated Type I error (Haining 2003, Wagner 2004, Van Teeffelen and Ovaskainen 2007). Discounting the results of nonspatial OLS in favor of that of spatial models when the coefficients differ is thus not tenable (Bini et al. 2009).

#### *Differentiating the sources of SAC*

Instead, researchers should focus on the sources of RSA and hence on what causes the differences between spatial and nonspatial methods. Although little emphasis has been put on disentangling the sources of RSA (Van Teeffelen and Ovaskainen 2007), understanding species–environment relationships can only be achieved by distinguishing between endogenous and exogenous SAC (Wagner and Fortin 2005, Fortin and Dale 2009). This is mostly done by partitioning the spatial pattern into a pure environmental component, a pure spatial component, a partition shared by environmental and spatial influence, and an unexplained portion (Borcard et al. 1992, 2004, Legendre et al. 2005, 2009, Peres-Neto et al. 2006, Laliberté et al. 2009). Scale influences the amount of variation explained by environmental predictors (Laliberté et al. 2009, Legendre et al. 2009), where erroneously ignoring landscape context leads to a downward-biased contribution of environmental predictors and an upward-biased contribution of pure spatial influences (Appendix). Consequently, conclusions regarding the contributions of exogenous and endogenous processes based on such analyses are thus scale-dependent: a scale mismatch leaves room for the inference that endogenous processes are at play even when only exogenous factors are driving the response variable. Much weight thus rests upon rationales used in the selection of appropriate spatial scales; yet they are often selected based on data availability, convenience, or our human perception of the system (Mayer and Cameron 2003, Dormann 2007, Wheatley and Johnson 2009).

#### *Generality and caveats*

Although we have framed our analysis in terms of the relationships between species distributions and landscape predictors, our arguments are not limited to such analyses only, but apply to a large array of ecological investigations where ecological responses are regressed against landscape predictors. Our analyses were kept simple for the purpose of demonstration, yet more complex analyses can be conducted within the framework outlined in this paper, e.g., by including distance–decay relationships in the specification of the spatial weight matrix  $\mathbf{W}$ . However, we do not claim that

explicitly considering scale effects necessarily increases our understanding of ecological phenomena, since ecological analyses are mostly correlative and are thus not strict inferential tests of causality: they can only suggest potential explanatory factors (Diniz-Filho et al. 2003). Moreover, cross-scale correlations can confound the interpretation of species–environment relationships, since predictors at one scale might be correlated with predictors at other scales (Mayor et al. 2007).

### Conclusion

We conclude that a scale mismatch decreases the portion of spatial variation explained by landscape predictors, while elevating the level of RSA and thereby obstructing inferential tests. Moreover, it leads to error–regressor correlations resulting in problematic parameter estimation. Because accurately describing the strength and significance of species–environment relationships is central to understanding the functioning of ecosystems, conserving biodiversity and managing ecosystems, we argue that explicitly considering scaling issues should be part of the formal framework within which proper estimation and inference are carried out. Doing so could contribute to a better understanding of spatial patterns in ecological phenomena by avoiding false inferences. The simplicity, power and many important potential applications make the approach outlined in this paper a useful addition to ecologists' toolbox.

### ACKNOWLEDGMENTS

We thank Carsten Dormann for sharing the data set employed in this paper as well as valuable insights. We also thank Mike Peel, Edward Kohi, Sip van Wieren, Raymond Florax, and two anonymous referees for stimulating and constructive comments on draft manuscripts. This study was financially supported by The Netherlands Foundation for the Advancement of Tropical Research (WOTRO) of The Netherlands Organization for Scientific Research (NWO).

### LITERATURE CITED

- Anselin, L. 2002. Under the hood. Issues in the specification and interpretation of spatial regression models. *Agricultural Economics* 27:247–267.
- Anselin, L. 2006. How (not) to lie with spatial statistics. *American Journal of Preventive Medicine* 30:S3–S6.
- Anselin, L. 2009. Spatial regression. Pages 255–275 in A. S. Fotheringham and P. A. Rogerson, editors. *The SAGE handbook of spatial analysis*. SAGE Publications, Los Angeles, California, USA.
- Anselin, L., R. J. G. M. Florax, and S. J. Rey. 2004. *Advances in spatial econometrics: methodology, tools and applications*. Springer-Verlag, Berlin, Germany.
- Ayinde, K. 2007. A comparative study of the performances of the OLS and some GLS estimators when stochastic regressors are both collinear and correlated with error terms. *Journal of Mathematics and Statistics* 3:196–200.
- Beale, C. M., J. J. Lennon, D. A. Elston, M. J. Brewer, and J. M. Yearsley. 2007. Red herrings remain in geographical ecology: a reply to Hawkins et al. (2007). *Ecography* 30:845–847.
- Betts, M. G., L. M. Ganio, M. M. P. Huso, N. A. Som, F. Huettman, J. Bowman, and B. A. Wintle. 2009. The ecological importance of space in species distribution models: a comment on Dormann et al. *Ecography* 32:374–378.
- Bini, L. M., J. A. F. Diniz-Filho, T. F. L. V. B. Rangel, and T. S. B. Akre, et al. 2009. Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. *Ecography* 32:193–204.
- Bivand, R. 2009. *spdep: spatial dependence: weighting schemes, statistics and models*. R package version 0.4-34. (<http://cran.r-project.org/web/packages/spdep/>)
- Bjornstad, O. N. 2009. *ncf: spatial nonparametric covariance functions*. R package version 1.1-2. (<http://cran.r-project.org/web/packages/ncf/>)
- Borcard, D., and P. Legendre. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* 153:51–68.
- Borcard, D., P. Legendre, C. Avois-Jacquet, and H. Tuomisto. 2004. Dissecting the spatial structure of ecological data at multiple scales. *Ecology* 85:1826–1832.
- Borcard, D., P. Legendre, and P. Draqueau. 1992. Partialling out the spatial component of ecological variation. *Ecology* 73:1045–1055.
- Brennan, J. M., D. J. Bender, T. A. Contreras, and L. Fahrig. 2002. Focal patch landscape studies for wildlife management: optimizing sampling effort across scales. Pages 67–91 in J. Liu and W. W. Taylor, editors. *Integrating landscape ecology into natural resource management*. Cambridge University Press, Cambridge, UK.
- Cliff, A. D., and J. K. Ord. 1981. *Spatial processes: models and applications*. Pion, London, UK.
- Clifford, P., S. Richardson, and D. Hemon. 1989. Assessing the significance of the correlation between two spatial processes. *Biometrics* 45:123–134.
- Cressie, N. A. C. 1991. *Statistics for spatial data*. John Wiley and Sons, New York, New York, USA.
- Currie, D. J. 2007. Disentangling the roles of environment and space in ecology. *Journal of Biogeography* 34:2009–2011.
- de Knegt, H. J., T. A. Groen, C. A. D. M. Van de Vijver, H. H. T. Prins, and F. Van Langevelde. 2008. Herbivores as architects of savannas: inducing and modifying spatial vegetation patterning. *Oikos* 117:543–554.
- de Knegt, H. J., G. M. Hengeveld, F. Van Langevelde, W. F. De Boer, and K. P. Kirkman. 2007. Patch density determines movement patterns and foraging efficiency of large herbivores. *Behavioral Ecology* 18:1065–1072.
- Diniz-Filho, J. A. F., and L. M. Bini. 2005. Modelling geographical patterns in species richness using eigenvector-based spatial filters. *Global Ecology and Biogeography* 14:177–185.
- Diniz-Filho, J. A. F., L. M. Bini, and B. A. Hawkins. 2003. Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography* 12:53–64.
- Diniz-Filho, J. A. F., A. Hawkins, L. M. Bini, P. De Marco, and M. Blackburn. 2007. Are spatial regression methods a panacea or a Pandora's box? A reply to Beale et al. (2007). *Ecography* 30:848–851.
- Dormann, C. F. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography* 16:129–138.
- Dormann, C. F. 2009. Response to Comment on “Methods to account for spatial autocorrelation in the analysis of species distributional data: a review.” *Ecography* 32:379–381.
- Dormann, C., et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30:609–628.
- Dormann, C. F., and R. Seppelt. 2007. Do organisms use landscape at certain spatial scales? A null model for diversity pattern in relation to the spatial extend of landscapes. Pages 35–42 in B. Shchröder, H. I. Reuter, and B. Reineker, editors. *Multiple scales in ecology*. Peter Lang, Internationaler Verlag der Wissenschaften, Frankfurt am Main, Germany.
- Dutilleul, P., P. Clifford, S. Richardson, and D. Hemon. 1993. Modifying the *t* test for assessing the correlation between two spatial processes. *Biometrics* 49:305–314.

- Ebbes, P., M. Wedel, U. Böckenholt, and T. Steerneman. 2005. Solving and testing for regressor-error (in)dependence when no instrumental variables are available: with new evidence for the effect of education on income. *Quantitative Marketing and Economics* 3:365–392.
- Florax, R. J. G. M., and H. Folmer. 1992. Specification and estimation of spatial linear regression models: Monte Carlo evaluation of pre-test estimators. *Regional Science and Urban Economics* 22:405–432.
- Fortin, M. J., and M. R. T. Dale. 2005. *Spatial analysis: a guide for ecologists*. Cambridge University Press, Cambridge, UK.
- Fortin, M. J., and M. R. T. Dale. 2009. Spatial autocorrelation. Pages 89–103 in A. S. Fotheringham and P. A. Rogerson, editors. *The SAGE handbook of spatial analysis*. SAGE Publications, Los Angeles, California, USA.
- Gotway, C. A., and L. J. Young. 2002. Combining incompatible spatial data. *Journal of the American Statistical Association* 97:632–648.
- Graham, M. H. 2003. Confronting multicollinearity in ecological multiple regression. *Ecology* 84:2809–2815.
- Griffith, D. A., and P. R. Peres-Neto. 2006. Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology* 87:2603–2613.
- Haining, R. P. 2003. *Spatial data analysis: theory and practice*. Cambridge University Press, Cambridge, UK.
- Hawkins, B. A., J. A. Diniz-Filho, L. M. Bini, P. De Marco, and T. M. Blackburn. 2007. Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology. *Ecography* 30:375–384.
- Holland, J. D., D. G. Bert, and L. Fahrig. 2004. Determining the spatial scale of species' response to habitat. *BioScience* 54:227–233.
- Keitt, T. H., O. N. Bjornstad, P. M. Dixon, and S. Citron-Pousty. 2002. Accounting for spatial pattern when modeling organism–environment interactions. *Ecography* 25:616–625.
- Kissling, W. D., and G. Carl. 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography* 17:59–71.
- Kühn, I. 2007. Incorporating spatial autocorrelation may invert observed patterns. *Diversity and Distributions* 13:66–69.
- Laliberté, E., A. Paquette, P. Legendre, and A. Bouchard. 2009. Assessing the scale-specific importance of niches and other spatial processes on beta diversity: a case study from a temperate forest. *Oecologia* 159:377–388.
- Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74:1659–1673.
- Legendre, P., D. Borcard, and P. R. Peres-Neto. 2005. Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs* 75: 435–450.
- Legendre, P., M. R. T. Dale, M. J. Fortin, J. Gurevitch, M. Hohn, and D. Myers. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography* 25:601–615.
- Legendre, P., X. Mi, H. Ren, K. Ma, M. Yu, I. F. Sun, and F. He. 2009. Partitioning beta diversity in a subtropical broad-leaved forest of China. *Ecology* 90:663–674.
- Lennon, J. J. 2000. Red-shifts and red herrings in geographical ecology. *Ecography* 23:101–113.
- Levin, S. A. 1992. The problem of pattern and scale in ecology: the Robert H. MacArthur Award lecture. *Ecology* 73:1943–1967.
- Luskin, R. C. 2008. Wouldn't it be nice ...? The automatic unbiasedness of OLS (and GLS). *Political Analysis* 16:345–349.
- Mayer, A. L., and G. N. Cameron. 2003. Consideration of grain and extent in landscape studies of terrestrial vertebrate ecology. *Landscape and Urban Planning* 65:201–217.
- Mayor, S. J., J. A. Schaefer, D. C. Schneider, and S. P. Mahoney. 2007. Spectrum of selection: new approaches to detecting the scale-dependent response to habitat. *Ecology* 88:1634–1640.
- Mayor, S. J., J. A. Schaefer, D. C. Schneider, and S. P. Mahoney. 2009. The spatial structure of habitat selection: a caribou's-eye-view. *Acta Oecologica* 35:253–260.
- McIntire, E. J. B., and A. Fajardo. 2009. Beyond description: the active and effective way to infer processes from spatial patterns. *Ecology* 90:46–56.
- Peres-Neto, P. R., P. Legendre, S. Dray, and D. Borcard. 2006. Variance partitioning of species data matrices: estimation and comparison of fractions. *Ecology* 87:2614–2625.
- Pinheiro, J. C., D. Bates, S. DebRoy, and D. Sarkar. 2008. nlme: Linear and nonlinear mixed effects models. R package version 3.1-90. (<http://cran.r-project.org/web/packages/nlme/>)
- R Development Core Team. 2009. R: a language and environment for statistical computing, version 2.8.1. R Foundation for Statistical Computing, Vienna, Austria. ([www.R-project.org](http://www.R-project.org))
- Schabenberger, O., and C. A. Gotway. 2005. *Statistical methods for spatial data analysis*. Chapman and Hall, Boca Raton, Florida, USA.
- Schlather, M. 2008. RandomFields: simulation and analysis of random fields, R package version 1.3.35. (<http://cran.r-project.org/web/packages/RandomFields/index.html>)
- Segurado, P., M. B. Araújo, and W. E. Kunin. 2006. Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology* 43:433–444.
- Steffan-Dewenter, I., U. Munzenberg, C. Burger, C. Thies, and T. Tscharntke. 2002. Scale-dependent effects of landscape context on three pollinator guilds. *Ecology* 83:1421–1432.
- Tiefelsdorf, M., and D. A. Griffith. 2007. Semiparametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning A* 39:1193–1221.
- Van Langevelde, F. 2000. Scale of habitat connectivity and colonization in fragmented Nuthatch populations. *Ecography* 23:614–622.
- Van Teeffelen, A. J. A., and O. Ovaskainen. 2007. Can the cause of aggregation be inferred from species distributions? *Oikos* 116:4–16.
- Wagner, H. H. 2004. Direct multi-scale ordination with canonical correspondence analysis. *Ecology* 85:342–351.
- Wagner, H. H., and M. J. Fortin. 2005. Spatial analysis of landscapes: concepts and statistics. *Ecology* 86:1975–1987.
- Wheatley, M., and C. Johnson. 2009. Factors limiting our understanding of ecological scale. *Ecological Complexity* 6: 150–159.
- Wong, D. 2009. The modifiable areal unit problem (MAUP). Pages 105–123 in A. S. Fotheringham and P. A. Rogerson, editors. *The SAGE handbook of spatial analysis*. SAGE Publications, Los Angeles, California, USA.
- Wu, J. 2007. Scale and scaling: a cross-disciplinary perspective. Pages 115–136 in J. Wu and R. J. Hobbs, editors. *Key topics in landscape ecology*. Cambridge University Press, Cambridge, UK.

## APPENDIX

The influence on omitting landscape context when partitioning the variance of our virtual species into environmental and spatial contributions (*Ecological Archives* E091-177-A1).