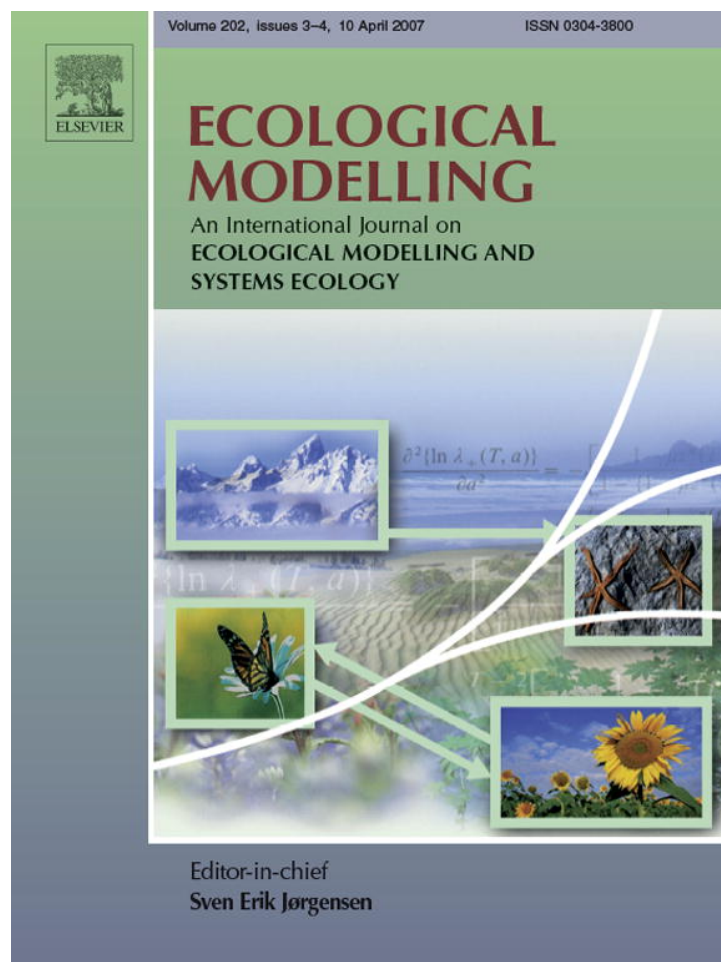


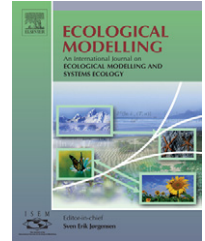
Provided for non-commercial research and educational use only.
Not for reproduction or distribution or commercial use.



This article was originally published in a journal published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues that you know, and providing a copy to your institution's administrator.

All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

<http://www.elsevier.com/locate/permissionusematerial>

available at www.sciencedirect.comjournal homepage: www.elsevier.com/locate/ecolmodel

Review

Incorporating spatial dependence in predictive vegetation models

Jennifer Miller^{a,*}, Janet Franklin^{b,c}, Richard Aspinall^d

^a Department of Geology and Geography, West Virginia University, Morgantown, WV 26506-6300, USA

^b Department of Biology, San Diego State University, San Diego, CA 92182, USA

^c Department of Geography, San Diego State University, San Diego, CA 92182, USA

^d The Macaulay Institute, Aberdeen, Scotland, UK

ARTICLE INFO

Article history:

Received 20 April 2005

Received in revised form

4 December 2006

Accepted 11 December 2006

Published on line 12 February 2007

Keywords:

Species distribution models

Habitat distribution models

Spatial dependence

GIS

ABSTRACT

Predictive vegetation modeling (PVM), is defined as predicting the distribution of vegetation across a landscape based on the relationship between the spatial distribution of vegetation and environmental variables. PVM requires digital maps of the environmental variables, as well as spatial information on the vegetation attribute of interest (e.g., species, type, abundance), usually from a sample of locations. Often these predictive models are developed using traditional statistical methods and are based on the implicit assumption that the distribution of vegetation is random and, therefore, each observation is independent. This approach violates one of the basic tenets of geography, the direct relationship between distance and likeness, as well as of ecological theory, that elements of an ecosystem close to one another are more likely to be influenced by the same generating process and will therefore be similar. Some of the spatial structure can be explained by the predictor variables used in the model. Environmental variables such as precipitation, temperature and elevation exhibit spatial dependence, some of which is responsible for spatial clustering in vegetation distribution, but remaining spatial dependence can result from either unmeasured environmental variables or biotic processes that cause spatial clustering. Spatial dependence in biogeographical data has been recently identified as an important area of future PVM research, and many studies have begun to explore ways to incorporate spatial dependence in predictive models. Here we review the different approaches to incorporating spatial dependence into predictive vegetation models focusing on four statistical methods: autoregressive models, geostatistics, geographically weighted regression, and parameter estimation models. Autoregressive models may be more capable of describing the fine-scaled spatial dependence that results from local biotic factors, such as disturbance, competition, or dispersal, while geostatistical methods may be more suitable for modeling broad-scale spatial dependence. The other methods focus on global and local parameter estimation in the presence of spatially structured or nonstationary data. While this review focuses on incorporating spatial dependence into statistical models for predictive purposes, explicitly including spatial dependence in models can also aid in clarifying the effect of different explanatory variables, thereby improving inferences.

© 2006 Elsevier B.V. All rights reserved.

* Corresponding author.

E-mail address: Jennifer.miller@mail.wvu.edu (J. Miller).

0304-3800/\$ – see front matter © 2006 Elsevier B.V. All rights reserved.

doi:10.1016/j.ecolmodel.2006.12.012

Contents

1. Introduction	226
1.1. Predictive vegetation modeling	227
2. Spatial dependence in biogeographical data	227
2.1. Nature of spatial dependence	227
2.2. Scale	228
2.3. Consequences of spatial dependence in traditional models	229
2.4. Measurement of spatial dependence	230
3. Methods that explicitly incorporate spatial dependence	230
3.1. Autoregressive methods	231
3.2. Applications of autoregressive methods	234
3.3. Geostatistical methods	235
3.4. Applications of geostatistical methods	236
3.5. Parameter estimation methods	236
3.6. Geographically weighted regression	236
3.7. Other methods	237
4. Summary	237
Acknowledgments	239
References	239

1. Introduction

Analysis of the geographic distribution of biological phenomena has always been an important issue in vegetation science, and is now the main focus of other sub-disciplines such as biogeography and landscape ecology. One of the most important research areas in vegetation science examines the relationship between environmental gradients and vegetation distribution. The mutual influence of pattern and process is manifested in the spatial distribution of vegetation depicted at different scales in vegetation maps. This information on the distribution of plant species over large areas has long been the main component of large-scale biodiversity assessment and ecosystem management (Scott et al., 1993). Environment–vegetation relationships have also become increasingly important in the context of anthropogenically driven climate change, and the ability to quantify these relationships provides a basis for predicting vegetation distribution as a result of changing environmental conditions.

Maps of vegetation composition have traditionally been produced by some combination of field survey and photointerpretation, but these methods are costly and inefficient. Recent developments (in the past 25 years) in remote sensing and geographic information science have produced more efficient alternatives for mapping vegetation. One of the most frequently used methods, predictive vegetation modeling (PVM), can be defined as predicting the distribution of vegetation across a landscape based on the relationship between the spatial distribution of vegetation and relevant environmental variables (Franklin, 1995). PVM requires digital maps of environmental variables, as well as spatial information on vegetation attributes, usually from a sample of locations. The environment–vegetation relationship can be based on observed correlation or on the theoretical or experimental physiological limitations of different plant species. The result is a vegetation map that is stored in a geographic informa-

tion system (GIS), which provides an organized structure for manipulation, analysis and display of the data.

Along with such issues as biotic interactions and spatially explicit uncertainty assessment, seminal reviews of PVM have identified spatial dependence as an important area of future research (Franklin, 1995; Guisan and Zimmermann, 2000). This involves a shift in assumptions made during model specification—traditional statistical methods used to analyze environmental–vegetation relationships are based on the assumption that the distribution of vegetation is random and, therefore, each observation is independent. This assumption violates one of the basic tenets of geography, the direct relationship between distance and similarity [Tobler's 'first law of Geography'¹ (Tobler, 1979, also see Sui, 2004)], as well as basic ecological theory. Ecosystem elements close to one another are more likely to be influenced by the same generating process and will therefore be similar (Legendre and Fortin, 1989).

The majority of previous PVM studies ignore spatial dependence altogether, and those that do acknowledge it usually consider it a nuisance and attempt to manipulate the sampling scheme to avoid autocorrelated observations. More recent research has focused on appropriate sampling schemes, scale, quantification of spatial pattern and statistical methods for use with spatial data in ecology (Miller et al., 2004, Rushton et al., 2004; special issue of *Ecography* 2002, v.25). However, the potential predictive ability of spatial dependence is only recently being explored in PVM and similar research.

Haining (1989) describes variation in spatial data as being derived from three sources: a deterministic structured element (first order effect), a stochastic structured element

¹ Tobler's law is given as "everything is related to everything else, but near things are more related than distant things", which Sui (2004) and Goodchild (2004) contend denotes both spatial dependence and spatial heterogeneity, respectively.

(second order effect), and a random error element (noise). Regression and other traditional statistical techniques focus on the first element, typically expressed as a function of a (potentially quite flexible) mathematical relationship between the response and predictor variables, while the second and third elements tend to be combined in the model error. We use this three part separation of variation in spatial data to review the past, present, and potential future role and contribution of spatial properties of environmental and vegetation data in PVM.

In this review we will briefly outline the ecological concepts on which PVM is based, as well as geographical concepts related to its application, such as scale. We describe sources of spatial dependence in biogeographic data, then discuss how spatial dependence has been addressed in models of biospatial response variables (of which PVM is a specific case), and focus specifically on four statistical methods that include spatial dependence: autoregressive models, geostatistics, geographically weighted regression, and parameter estimation models.

1.1. Predictive vegetation modeling

As noted in the introduction, much attention has been placed on PVM recently in the context of global change, where perturbations in climatic factors can have long-lasting consequences for vegetation distribution. Climate has been linked with vegetation distribution from at least the early 1800s when von Humboldt wrote about the relationship between latitude and vegetation type (Legendre and Fortin, 1989; Jongman et al., 1995). Early global plant distribution maps based on climatic factors alone were surprisingly accurate (Holdridge, 1967) and the relationship between climate and vegetation remains very important in vegetation modeling. Bioclimatic indices such as potential solar radiation, mean relative humidity and potential evapotranspiration may be more directly related to plant distribution than average precipitation and temperature and are becoming more common in vegetation models (Franklin, 1998; Leathwick, 1998; Cairns, 2001; Meentemeyer et al., 2001).

The models are considered to be ‘static’ and assume that vegetation distribution is in (temporary) equilibrium with the environment and that the relationships between vegetation distribution and environmental variables detected in a sample of observed distributions are consistent throughout the study area (Franklin, 1995; Guisan and Zimmermann, 2000). Although static models do not usually consider dynamic ecological processes such as competition, predation and disturbance, or feedbacks between vegetation and changing environmental conditions as more complex dynamic vegetation models do, they are a necessary ‘first step’ in developing dynamic models, and are a commonly used method for predicting vegetation response to environmental change at regional to global scales.

What is now referred to as PVM evolved from research methods that used gradient analysis to explore the relationship between plant species composition and distribution and environmental gradients (Whittaker, 1973; Kessell, 1979). Species distributions along gradients can be characterized in terms of changing abundance related to changing values along the gradient, and this information used to determine niche or

habitat breadth. Whether the type of gradient is direct, indirect, or resource (see Austin, 1980) determines how predictive the resulting model is likely to be.

Gradient analysis describes the distribution of vegetation as it changes along gradients in *environmental* space, which does not necessarily translate into *geographic* space (Austin, 1985; Austin and Smith, 1989). Environmental space is defined by the range of values measured for an environmental variable (e.g., temperature, precipitation), while geographic space is defined by the range of values of geographic coordinates (e.g., latitude/longitude). This separation between environmental and geographic space during the development of PVM may be responsible for the general avoidance of any explicit representation of geographic space as a correlate of vegetation distribution.

2. Spatial dependence in biogeographical data

Predictive vegetation models are often developed without considering the spatial pattern that exists in biogeographical data. Some spatial pattern can be explained by the predictor variables used in models. Environmental variables such as precipitation, temperature and elevation exhibit positive spatial dependence, some of which is responsible for spatial dependence in vegetation distribution. Remaining spatial dependence can result from either unmeasured environmental variables or biotic processes that cause spatial clustering and therefore can be an important component in predictive vegetation models.

2.1. Nature of spatial dependence

Spatial dependence is defined by Legendre and Legendre (1998) as “the property of random variables taking values, at pairs of locations a certain distance apart, that are more similar (positive autocorrelation) or less similar (negative autocorrelation) than expected for randomly associated pairs of observations” (p. 8). Spatial dependence, long considered important in geostatistics and econometrics, is becoming more widely recognized in other fields. Spatial dependence is observed in vegetation data, where heterogeneous spatial structure is largely a function of patchiness in environmental conditions or community processes (Borcard et al., 1992).

Potential sources/underlying causes of spatial dependence in biogeographic data (plant and animals) include (K. Van Niel, personal communication, 2004):

- dispersal/growth;
- disturbance;
- predation/herbivory;
- environmental/physical barriers;
- historical biogeography;
- interaction/competition/mutualism;
- metapopulation dynamics;
- learned behaviors (foraging, fear);
- spatial dependence in predictor variables;
- missing/mis-specified predictor variables.

In addition, any of these factors and processes may interact with each other or other factors to produce more complex spatial patterns (see Berlow, 1999; Schenk et al., 2003; Schwarz et al., 2003; Vaughan and Ormerod, 2003; Lortie et al., 2004).

The manifestation of these factors and processes in biogeographical data can be differentiated (albeit somewhat superficially) by the scale at which their influence is observed. Broad-scale spatial dependence, Haining's first order effect, refers to a predictable pattern or spatial trend that is deterministic and usually related to an unmeasured environmental variable (also known as drift in the geostatistical literature). Legendre and Legendre (1998) refer to this as a true gradient, where gradient implies a gradual change in values in geographic space, as would be observed in an environmental variable with a strong spatial structure. In socio-economic models, Anselin (1992) describes the same phenomenon as 'nuisance' spatial dependence, as it does not characterize an observation that is truly jointly determined by its neighbors. This type of spatial dependence results in spatially correlated errors and would therefore be specified in the model by a lagged error term, and the mean of the response variable would not be affected (Anselin, 1992).

The second order effect described by Haining (1989) is fine-scaled spatial dependence and occurs when the process(es) that produce the distribution of the response variable are inherently spatial, reflecting biotic interaction between sites. Legendre and Legendre (1998) call this same effect a false gradient. The spatial dependence that results from these complex spatial/behavioral/physiological processes comes closer to representing what Anselin (1992) terms 'substantive' spatial dependence, and would be specified in a model with a lagged response variable as an explanatory variable. Spatially, this type of pattern would generally be observed to be more patchy than gradual, as the processes that generate it are more stochastic than deterministic. Fewer previous studies have attempted to incorporate fine-scaled spatial dependence in PVM (Lichstein et al., 2002).

While Lagrange multiplier tests have been developed to distinguish between 'substantive' and 'nuisance' spatial dependence in econometric applications (Anselin, 1992), Legendre and Legendre (1998) note that distinguishing between false gradient and true gradient in ecological data is a "moot point" (p. 725). If the process resulting in the pattern is known, usually a determination as to whether it is a false gradient or true gradient can be made, but the pattern itself is not enough to enable that distinction.

Both types of spatial dependence can have potential predictive ability, but as they also both represent combinations of factors that are location-specific (possible environmental factors for the true gradient and biotic factors such as competition, dispersal, and disturbance for the false gradient), they should be considered indirect gradients. It should be noted that Legendre and Legendre (1998) further differentiate true gradient as 'spatial dependence' and false gradient as 'spatial autocorrelation,' but here we use the terms to represent the same general spatial phenomena. However, we suggest that what is discussed here as spatial dependence refers to Haining's second order effect (Haining, 1989), and the true gradient discussed in Legendre and Legendre (1998) should be considered a spatial trend, which occurs when

an environmental gradient is coincident with a geographic gradient.

2.2. Scale

Scale has been described as the fundamental conceptual problem in ecology (Levin, 1992) and "perhaps the most important topic in GIScience" (Goodchild, 2001). The evolution of the concept of scale in GIS from cartographic scale to pixel size has been addressed by Goodchild (2001), while Dungan et al. (2002) distinguish between cartographic scale and two-dimensional descriptions of scale as *support*, the term used in geostatistics (see also, Goodchild and Proctor, 1997; Quattrochi and Goodchild, 1997; Atkinson and Tate, 2000; Tate and Atkinson, 2001; Lilburne et al., 2004). In GIS analysis, the concept of a single scale to report has become irrelevant, and instead separate aspects of scale are often reported. Fig. 1 illustrates some of the different aspects of scale relevant to PVM applications in general, and measuring and modeling spatial dependence in particular. Fig. 1a represents a typical plot that would be used to record information on vegetation distribution in the field (e.g., there is an implicit coarsening of measurement scale as one converts measurements of species abundance to vegetation type or species presence/absence). The area of the plot is determined by some combination of factors involving the level and type of measurement (e.g., if the vegetation variable is at the community level, the unit would have to be large enough to describe different components of the community), and the natural scale of vegetation distribution (e.g., a plot size of 4 m² might be used for sampling herbaceous communities and 0.1–1 ha for sampling forests). Large plot sizes reduce the ability to discriminate finer spatial variation, while small plot sizes may not represent the area adequately (Bellehumeur and Legendre, 1998). The plot size also affects the ability to measure spatial dependence. If a biotic process that results in clustering occurs at a scale similar to or less than the plot, it will not be detectable. Fig. 1b shows how this information would be represented in a raster GIS. Here the important scale issue is the determination of an appropriate resolution (grid cell size). Resolution of predictor variables, including explicit representation of spatial dependence, should ideally be a function of the ecological scale at which they are associated with the response data, but compromises are often made with respect to data availability or computational limits.

Finally, Fig. 1c shows how the data are typically used for analysis. The plots can be point or grid cell representations of the sample plot information from Fig. 1a. Either way, the area covered by the plot is often much greater than the corresponding map area used to represent the plot. The study area represents the entire spatial extent of the area used in the analysis (e.g., where predictions will be generated). As Dungan et al. (2002) point out, this is often not the same thing as the spatial extent of the observations. The sampling interval or lag expresses the average distance between sample points/cells. A too-short sampling interval may include too many observations influenced by the same spatial process; a too-long sampling interval may select observations with no detectable spatial dependence.

The ability to use spatial dependence as a predictive tool depends upon whether: (1) there is spatial dependence in the

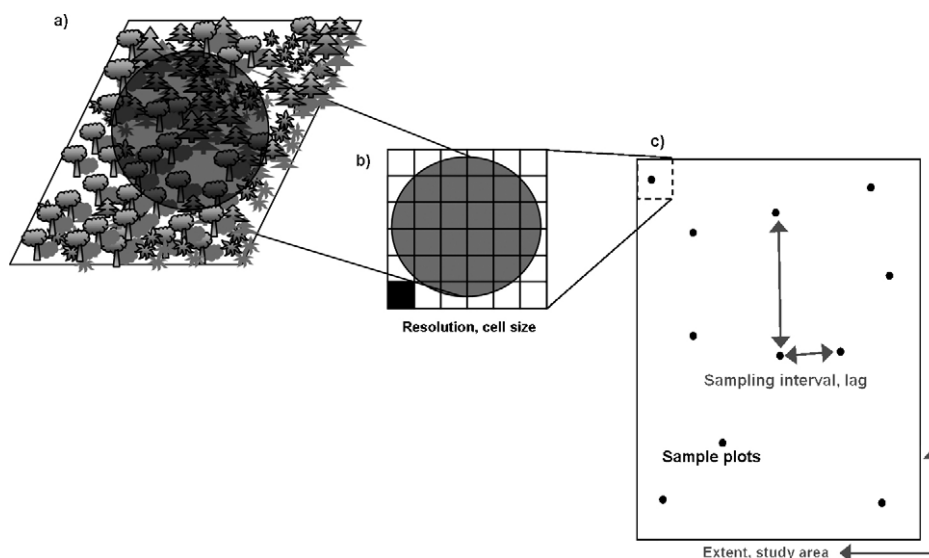


Fig. 1 – Different aspects of scale that are used in PVM: (a) a sample plot in the real world phenomena of interest; (b) the GIS data model representation of that observation; (c) the data analysis representation.

vegetation distribution (e.g., that is observed in Fig. 1a); (2) the resolution (shown in Fig. 1b) is adequate for representing the continuous nature of all of the processes associated with the vegetation variable (including spatial dependence); (3) the sampling interval and intensity (Fig. 1c) are appropriate for describing the spatial dependence. Even if there is spatial dependence in the vegetation distribution, and the resolution is appropriate, the sampling interval may exceed the scale at which spatial dependence is observed, rendering it unresolvable. When sampling dynamic populations in space and time, Tobin (2004) found that estimates of spatial dependence are degraded with reduced sample sizes, while estimates of the range (distance beyond which dependence is negligible) were more stable. Typical compiled or opportunistic datasets used in PVM result from some combination of sampling schemes, a by-product of which may be an inability to measure spatial dependence among the observations.

2.3. Consequences of spatial dependence in traditional models

The analytical framework for PVM is most generally based on multiple regression, where the vegetation response variable is related to a suite of environmental predictor variables. This traditional statistical framework assumes that observations are independent of one another, a condition unlikely to occur with ecological phenomena. When untreated, spatial dependence in ecological data violates the assumption of independence of observations required by most commonly used statistical methods (e.g., ordinary least squares, OLS). Traditional hypothesis testing measures can produce underestimated standard errors, resulting in increased Type I errors (Legendre, 1993; Legendre and Legendre, 1998; Legendre et al., 2002). In regression, this can result in erroneous assessment of the significance of relationships between explanatory and response variables, and can lead to, among other things, incorrectly rejecting null hypotheses of no relationship. It should be

noted that Vaughan and Ormerod (2003) suggest that for management applications, spatial dependence is more likely to be treated as a nuisance, as it represents landscape and contextual information that is beyond the control of managers.

Beyond this, Lennon (2000) found that variable selection in multiple regression analysis was biased towards autocorrelated predictor variables, and suggested that previous ecological analyses based on these traditional methods are flawed. Diniz-Filho et al. (2003) refute this and point out that, while methods that address spatial dependence (e.g., generalized least squares as opposed to OLS) tend to de-emphasize autocorrelated predictor variables, the issue is related more to the different spatial scales at which predictor variables influence a response variable (broad-scale influences are perceived as having greater spatial dependence). Legendre et al. (2002) found that broad-scale spatial structure had the same effect on statistical hypothesis testing (inflated significance) as spatial autocorrelation.

Some studies have attempted to avoid problems associated with spatial dependence by manipulating the sampling strategy to avoid autocorrelated samples (e.g., Sokal and Oden, 1978b; Legendre and Fortin, 1989; Davis and Goetz, 1990; Borcard et al., 1992; Legendre, 1993). This focuses on lag/sampling intensity as one element of scale and a single source of spatial dependence. Borcard et al. (1992) were able to separate the spatial component that was related to vegetation pattern from the environmental component in a correspondence analysis. One problem caused by spatial dependence is that each observation contributes less information and the degrees of freedom used in analyses are exaggerated. Thomson et al. (1996) used a method to modify the degrees of freedom based on spatial dependence in order to proceed with analysis. They found that previously statistically significant relationships that ignored autocorrelation in the data were spurious.

However, as the paradigm in vegetation analysis shifts from one in which data are considered to be independent, to one in

which the data are assumed to have some spatial structure, the methods that address spatial dependence will also have to evolve. Developing and using methods that incorporate this spatial dependence, rather than focusing on strategies that identify, avoid, lessen, or remove it, will become more important.

2.4. Measurement of spatial dependence

Spatial dependence in point and area data has been studied more extensively in socioeconomic and medical geography applications (Cliff and Ord, 1981; Haining, 1990; Cressie, 1991). A suite of statistics has been developed to test for the presence and magnitude of spatial dependence in data. These include global distance – or contiguity – based measures such as Moran's *I*, and Geary's *c* (see Cressie, 1991 for discussion), as well as more recent local statistics that can characterize spatial clustering in a portion or geographic subset of the dataset (Getis and Ord, 1996). These spatial dependence statistics can also be used as a more objective means of quantifying landscape pattern in terms of patchiness (Aspinall, 1999). While some ecological studies have used these statistics to test for spatial dependence (Fortin, 1999; Dale et al., 2002; Perry et al., 2002), their use has typically preceded removal or justified avoidance of spatial dependence in a model, rather than explicitly incorporate spatial dependence (Klute et al., 2002; Lichstein et al., 2002).

Another way in which spatial dependence has been quantified follows work done by Borcard et al. (1992) and Legendre (1993), where the variation of the response variable is separated into environmental (e.g., climate, topography) and spatial (represented by a trend surface of geographic coordinates) components (see also Lobo et al., 2002, 2004; Nogués-Bravo and Martínez-Rica, 2004; Titeux et al., 2004). Partial regression analysis (Legendre and Legendre, 1998) using either GLMs or GAMs is used to decompose the explanatory effects of a group of variables into single variation fractions. Fig. 2 shows a conceptualization of this model: the outer circle represents the variation in the response variable: (a) represents unexplained variation (model error); (b) represents the (nonspatial) effect of an environmental variable; (c) represents the overlapping effects of environment and spatial dependence (the coincident environmental and geographic gradients); (d) represents the effect of spatial dependence. However, what is described as the 'pure effect of spatial variation' (Lobo et al., 2002, 2004), and the 'pure spatial fraction' (Nogués-Bravo and Martínez-Rica, 2004) represented by (d) in Fig. 2 should be interpreted with caution (Legendre and Legendre, 1998). As the spatial variables are usually represented by a trend surface of the geographic coordinates, Fig. 2d describes some broad-scale spatial trend that does not covary with the environmental variable(s), and is most likely related to a spatially varying explanatory variable that is not included in the model. Unless specified to be extraordinarily complex, the polynomial equations defined in the trend surface analysis are usually not capable of describing fine-scaled spatial variation, as would result from patchy biotic processes.

This method is particularly useful for differentiating between the broad-scale correlation between environmental variables and the response variable (Fig. 2c), and the

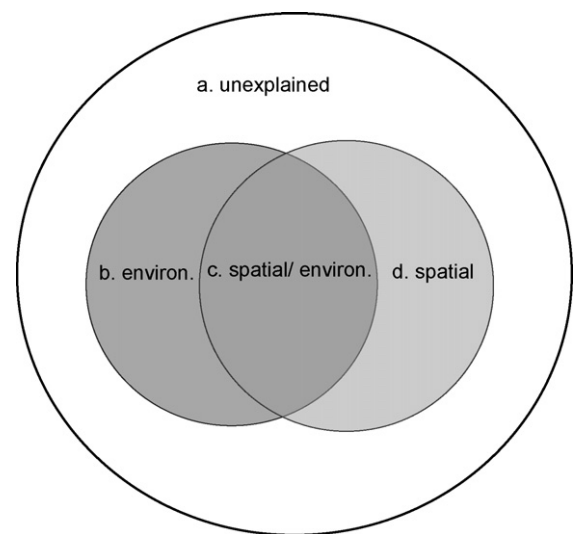


Fig. 2 – A conceptualization of the variation in vegetation distribution; the outer circle represents the variation in the vegetation distribution: (a) represents unexplained variation (model error); (b) represents the variation explained by an environmental variable; (c) represents the variation explained by spatially structured environment; (d) represents the variation explained by spatial structure alone.

nonspatial usually finer-scaled correlation between the environmental variables and response variable (Fig. 2b) (Legendre and Legendre, 1998). The importance of distinguishing between the spatial structure of an environmental variable and some true relationship between environment and the response variable is often overlooked and can result in determination of statistically significant relationships that are really based on false correlations. The ability to separate the spatial structure from the environmental variable can be an important first step in developing hypotheses related to causation versus correlation (Legendre and Legendre, 1998; MacNally, 2002; McMillen, 2003).

The methods and techniques described above are important for identifying and characterizing spatial dependence, and their burgeoning use in vegetation analysis is encouraging. However, the potential predictive ability of spatial dependence has not been fully explored in PVM research.

3. Methods that explicitly incorporate spatial dependence

In general, kriging and autoregressive models estimate the values at unsampled locations by a weighted averaging of nearby samples. With kriging, the correlations among neighboring values are modeled as a function of the geographic distance between the points across the study area, defined by a variogram. With autoregressive methods, the correlations are modeled as a function of an *a priori* determined neighborhood, typically the number of grid cells surrounding each grid cell used in the analysis.

Although kriging provides a mechanism for combining global and local information in predictions, the ability of the variogram to describe spatial dependence is directly a function of the quantity and quality of the sample data. A true local cluster may not be detected if the number of sample pairs affected by it is not adequate. For this reason, kriging and other variogram-based prediction methods may be more suitable for modeling the broad-scale spatial dependence described as the ‘true gradient’ above. Conversely, neighborhood-based prediction methods, such as autoregressive models, may be more capable of describing the fine-scaled spatial dependence that results from local biotic factors, such as disturbance, competition, or dispersal.

Some studies have used the geographic coordinates of the response data as explanatory variables. *Pereira and Itami (1991)* fit a trend surface to the geographic coordinates and combined this information with a regression model using environmental predictors (*Le Duc et al., 1992; Lichstein et al., 2002*). In comparing classification tree models using environmental data to “spatial” models in which latitude and longitude were included as predictors, *Matsui et al. (2004)* unsurprisingly found that the geographic coordinates mimicked the effect of climate variables. Geographic coordinates have been included as predictor variables in GAMs, with the resulting nonparametric smooth surface describing the effect of location as a proxy for unobserved factors on the response variable (*Preisler et al., 1997; Franklin, 1998*). While including geographic coordinates as predictor variables can improve model accuracy efficiently (coordinate information is readily available), this effect should not be confused with spatial dependence (and would perhaps be more appropriately referred to as ‘geographic dependence’). Geographic coordinates represent absolute locations and, when correlated with spatially varying environmental variables, may describe broad-scale trends in the pattern of the response variable, but contribute little to the potential clarification of vegetation–environment relationships or to inclusion of spatial dependence in models.

While this paper focuses on methods of incorporating spatial dependence specifically in PVM, we broaden the context of the review to include models of all biospatial response variables (e.g., species/habitat distribution models, predictive soil modeling) in the following section. Unlike, for example, econometrics and geostatistics where spatial dependence has long been explicitly incorporated in models, PVM applications have been slower to recognize its importance. We consider the representation of spatial dependence in models of biospatial variables in general to be similar enough with PVM applications specifically to be discussed here, and examples are given in *Table 1*, along with information on the model used and the means by which spatial dependence is incorporated. For reasons discussed above, we do not include studies that incorporated geographic or spatial coordinates as predictor variables, unless they are used in concert with other spatial dependence methods (see *Augustin et al., 1996a,b, 1998; Gumpertz et al., 2000; Weir and Pettitt, 2000; Ferrier et al., 2002*). Similarly, from the extensive literature on geostatistical applications (e.g., soil mapping), only those methods which show particular promise and applicability for PVM (e.g., regression kriging: *Knotters et al., 1995;*

Odeh et al., 1995; Bishop and McBratney, 2001) are included in *Table 1*.

When spatial dependence is explicitly incorporated in models with an emphasis on prediction, it is generally characterized using one of two methods. Variogram-based methods (e.g., kriging) describe the correlation among neighboring values as a function of the geographic distance between the points. Up to a certain distance, beyond which correlation is not apparent, the farther apart two points are the less likely their values will be similar. Neighborhood-based methods (e.g., autoregressive models) assume that only immediately proximal cells influence each other and beyond this *a priori* defined neighborhood, the influence is not modeled.

3.1. Autoregressive methods

While geostatistical models are traditionally associated with point data, autoregressive models have long been used in econometric research associated with area data. GIS analysis using raster data models increases flexibility for treating point data as area data. *Anselin (1993)* defines the autoregressive (AR) model as

$$y = \alpha + \rho Wy + \varepsilon \quad (1)$$

where α is the constant term, ρ the spatial autoregressive coefficient, Wy the spatial lag for variable y (W describes the neighborhood, which can be based distance or other topological relations), and ε is the error term. The response variable is predicted using only information on its spatial neighbors, similar to ordinary kriging. The AR model can be generalized (and its predictive ability presumably increased) with the addition of other predictor variables, formally called a mixed-regressive–spatial-autoregressive model (*Anselin, 1993*), although widely referred to as the autoregressive model (including here). This model is expressed as

$$y = \alpha + \rho Wy + X\beta + \varepsilon \quad (2)$$

where $X\beta$ are the other predictor variables and coefficients. While the neighborhood (W) can be based on Euclidean distance, it is most often defined for raster data in terms of the number of cells surrounding each grid cell (*Fig. 3*). Similar to AR models, conditional autoregressive (CAR) and simultaneous autoregressive (SAR) modify the spatial lag of y to accommodate the difference between expected and observed values based on predictor variables (residual variation) and differ by the type of neighborhood they can express (*Keitt et al., 2002*). Although the neighborhoods are usually defined in the context of contiguity, distance-based measures are also used. A special revision of the logistic regression model to incorporate spatial dependence with binary response data was first suggested by *Besag (1972, 1974)*. This autologistic model (ALM) uses an autocovariate to represent neighboring values (see *Fig. 3*, for examples) used to predict a binary response. The original autologistic regression model has been modified to incorporate additional predictor covariates (*Augustin et al., 1996b; Gumpertz et al., 1997; Wu and Huffer,*

Table 1 – Survey of studies with biospatial response variables that explicitly incorporate spatial dependence

Study	Modeling method(s)	Response variable	Spatial dependence incorporated as	Additional variables used
Autoregressive methods				
Augustin et al. (1996a)	GLM, GAM	Deer presence/absence, number of eggs	Autocovariate	Temperature, salinity, landcover
Augustin et al. (1996b)	GLM	Deer presence/absence	Autocovariate	Habitat-related
Augustin et al. (1998)	GLM, GAM	Deer presence/absence	Autocovariate	Altitude, landcover
Betts et al. (2006)	GLM	Bird presence/absence	Autocovariate	Topography, age, cover class, landscape structure
Bullock and Burkhart (2005)	SAR	Tree stem diameter	Distance-based weights matrix	Stem height
Dark (2004)	SAR	Number of invasive plant species	Spatial weights matrix	Elevation, native species richness, road density, population density
Dennis et al. (2002)	GLM	Butterfly presence/absence	Autocovariate	Area, altitude
Ferrier et al. (2002)	GLM	Plant presence/absence	Modified trend surface, modified autocovariate	Topography, climate, soil
Gumpertz et al. (1997)	GLM	Disease presence/absence	Autocovariate	Water content, pathogen density, leaf disk assay
He et al. (2003)	GLM	Plant presence/absence	Autocovariate	Climate
Heikkinen and Hogmander (1994)	GLM	Toad range	Autocovariate	Search intensity
Hoeting et al. (2000)	GLM	Simulated data	Autocovariate	Simulated data, search intensity
Hubbell et al. (2001)	GLM	Odds of tree species survival	Autocovariate	Density, plant size, species richness
Huffer and Wu (1998)	GLM	Plant presence/absence	Autocovariate	Climate
Jeganathan et al. (2004)	GLM	Bird track presence/absence	Autocovariate	Principal components analysis of habitat factors
Keitt et al. (2002)	AR, CAR, SAR	Abundance of voles	Correlogram of residuals, residual variogram	Vegetation
Klute et al. (2002)	GLM	Bird presence/absence	Autocovariate	Habitat type
Knapp et al. (2003)	GLM, GAM	Frog larvae presence/absence	Autocovariate	Topography, substrate, stream quality
Lichstein et al. (2002)	CAR, GLM	Bird abundance	Trend surface, correlogram of residuals	Habitat, landscape
Luoto et al. (2002)	GLM	Butterfly presence/absence	Autocovariate	Habitat, topography
Mattson and Merrill (2002)	GLM	Bear presence/absence	Autocovariate	Ecoregions, food range, human factors
Milsom et al. (2000)	GLM	Bird presence/absence	Autocovariate	Habitat factors
Osborne et al. (2001)	GLM	Bird presence/absence	Autocovariate	Cultural features, NDVI, rivers, terrain
Overmars et al. (2003)	SAR	Land use	Distance-based weights matrix	Soil, altitude, population, distance to roads
Preisler (1993)	GLM	Probability of tree infestation	Autocovariate	DBH, age
Segurado and Araújo (2004)	GLM, GAM	Amphibian and reptile presence/absence	Contagion predictor variable	Climate, land use, human population density, soil
Silva et al. (2002)	GLM	Probability of species colonization	Autocovariate	Altitude, solar radiation
Smith (1994)	GLM	Plant presence/absence	Autocovariate	Altitude, climate, slope
Tognelli and Kelt (2004)	CAR, SAR	Mammalian species richness	Distance-based weights matrix	Climate, NDVI, elevation
Weir and Pettitt (2000)	Modified CAR	Toad presence/absence	Bayesian neighborhood; grid coordinates	
Wu and Huffer (1997)	GLM	Plant presence/absence	Autocovariate	Climate
Geostatistical methods				
Bishop and McBratney (2001)	GAM, DT, GLM, RK	Soil cation exchange capacity	Variogram, residual variogram	Terrain, TM imagery, crop yield, conductivity
Bolstad et al. (1998)	CK	Basal area of tree species	Variogram	Elevation, terrain shape

Table 1 (Continued)

Study	Modeling method(s)	Response variable	Spatial dependence incorporated as	Additional variables used
Carroll and Pearson (1998)	UK	Butterfly species richness	Variogram	Beetle species richness, geographic coordinates
Chong et al. (2001)	GLM, RK	Number and presence of plant species	Variogram, residual variogram	Location, elevation, slope, aspect, TM imagery
Gotway and Hartford (1996)	Extension of UK, CK	Nitrate concentration	Variogram, residual variogram	Grain yield
Kleinschmidt et al. (2001)	GLMM, RK	Malaria counts	Residual variogram	Climate, distance to water
Knotters et al. (1995)	CK, RK, MLR	Soft layers depth	Variogram, residual variogram	Conductivity
Miller (2005)	RK	Vegetation type presence/absence	Residual variogram	Climate, topography, geology
Odeh et al. (1995)	UK, CK, GLM, RK	Soil properties	Variogram, residual variogram	Landform
Pfeffer et al. (2003)	UK	Detrended correspondence analysis of vegetation type	Residual variogram	Elevation, slope, solar radiation, wetness index, profile curvature
Ver Hoef (1993)	UK	Log-transformed plant species cover	Residual variogram	Area of woody plant species
Geographically weighted regression				
Foody (2004a)	GWR	Bird species richness	Spatially varying coefficients	Climate, NDVI
Shi et al. (2006)	GWR	Deer density	Spatially varying coefficients	Climate, landcover
Wang et al. (2005)	GWR	Net primary production	Spatially varying coefficients	Climate, elevation, NDVI
Zhang et al. (2004)	GWR	Tree height	Spatially varying coefficients	Tree diameter at breast height
Parameter estimation methods				
Gotway and Stroup (1997)	GEE	Presence of plant damage, groundwater above sea level	Variogram, residual variogram	-
Gumpertz et al. (2000)	GEE	Proportion of years of beetle outbreaks	Variogram	Timber volume, topography, climate
Hooten et al. (2003)	GLMM	Plant presence/absence	Exponential correlation function based on distance	Aspect, elevation, land type, variable depth soil
Mugglestone et al. (2002)	GEE	Deer presence/absence	'Oddsrationgram'	Altitude, habitat
Other methods				
Dirnböck and Dullinger (2004)	(Ordinal) GLM	Plant abundance	Dummy variable indicating neighborhood presence	Climate, geomorphology
Leathwick (1998)	GAM	Tree species presence/absence	Dummy variable indicating neighborhood presence	Climate, geology, plot size
Miller and Franklin (2002)	DT, GLM	Vegetation type presence/absence	Variogram, autologistic term	Climate, topography
Olivier and Wotherspoon (2005)	DT, GLM	Bird (nest) presence/absence	Variables representing distance to neighboring nests and average number of nests in adjacent unit	Topography, substrate

Model abbreviations are—(C/S)AR: (conditional/simultaneous) autoregressive model; CK: co-kriging; DT: decision tree; GAM: generalized additive model; GEE: generalized estimating equation; GLM: generalized linear model; GLMM: generalized linear mixed model; GS: Gaussian simulation; GWR: geographically weighted regression; MLR: maximum likelihood regression; RK: regression kriging; UK: universal kriging.

1997) as

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta_1 \text{cov}_1 + \dots + \beta_n \text{cov}_n + \beta_{n+1} \text{autocov}_i \quad (3)$$

where an autocovariate at site *i* is defined as a weighted sum of observations in neighboring sites in a neighborhood defined

by N_i (analogous to W_y in Eq. (2)):

$$\text{autocov}_i = \sum_{j \in N_i} w_{ij} y_j \quad (4)$$

If the coefficient for the autocovariate (β_{n+1}) is equal to zero, indicating that spatial dependence is not important to

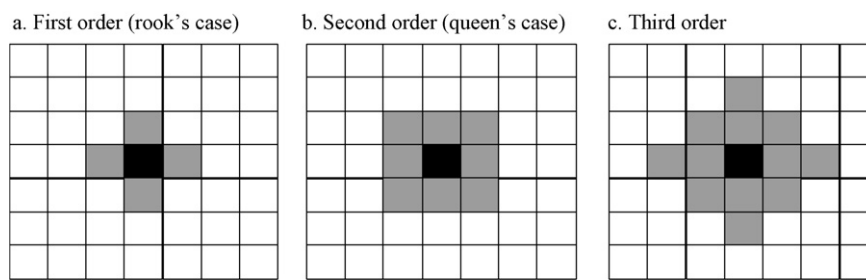


Fig. 3 – Neighborhood definitions. The gray cells show the cells considered to influence the center (black) cell for three of the most commonly used neighborhood definitions.

the responses, the model reduces to the ordinary logistic regression model. If the coefficients for all of the covariates (β_1, \dots, β_n) equal zero, the model follows Besag's (1974) autologistic model where only the spatial effects are used. The resulting model is more accurate because it includes the inherent spatial dependence (Besag, 1974). A logistic regression model that ignores spatial dependence may need more predictor variables to explain the variation, whereas an autologistic model includes a variable to describe the dependence and other variables that were used as proxies to capture the spatial correlation presumably become insignificant (Wu and Huffer, 1997).

However, tests used to indicate spatial dependence in the data, such as spatially patterned residuals, can also result from misspecification of the functional form of the model (McMillen, 2003). Austin (2002) points out that several papers (Wu and Huffer, 1997; Gumpertz et al., 1997; Weir and Pettitt, 2000) that demonstrate the advantage of incorporating spatial dependence with an autocovariate used linear relationships between the species and the environment and it was quite probably model misspecification that led to evidence of spatial dependence.

Autoregressive models allow for more flexible relationships to describe the deterministic element (e.g., logistic link), as well as for spatial relationships that are not consistent throughout the study area and can be defined by covariance matrices based on distance or contiguity. Wall (2004) points out that the spatial structures implied by SAR and CAR models are often not indicative of intuitive or meaningful spatial patterns and in fact the specification of neighborhood influence is generally calibrated by the spatial weights matrix and may not be supported by the data. Unless used mainly to clarify the importance of other predictor variables, more attention should be paid to understanding the spatial patterns represented in these models (Wall, 2004).

When the autocovariate term is based on response values that are predicted rather than observed, autologistic model fitting requires iteration. This has been accomplished in most of the literature using Gibbs sampler and Markov chain methods (Augustin et al., 1996b, 1998) and Markov chain Monte Carlo methods (MCMC; Gumpertz et al., 1997; Wu and Huffer, 1997). Basically this technique involves running iterations with simulations that were based on observed values until the output converges. When there is strong intrinsic spatial dependence, the MCMC estimation methods can become numerically unstable (Wu and Huffer, 1997). Further, there are

two issues that impede more extensive application of autoregressive methods in PVM studies: the procedures for fitting the full model are quite computationally intensive and software to do this is not readily available; and selection of an appropriate neighborhood for calculating the autocovariate is highly subjective and is extremely affected by the original sampling scheme.

3.2. Applications of autoregressive methods

Using an ALM to model butterfly species presence, Dennis et al. (2002) found that neighborhood models were more successful than models based on geographic coordinates (latitude and longitude). They also found that a simple contiguity-based neighborhood measure was slightly more important in the models than a more complex distance-weighted neighborhood measure (Dennis et al., 2002). Using an autologistic model, Ferrier et al. (2002) generalized the autocovariate term to include information on 'effective distance' between the focal and surrounding cells, which increases when dissimilar habitat occurs between them.

Gumpertz et al. (1997), Wu and Huffer (1997), and Hubbell et al. (2001) used ALM to obtain more accurate parameter estimates and to clarify the true effects of other covariates, while Hoeting et al. (2000) focused on the accuracy of the image restoration using simulated data with known 'true' populations. Keitt et al. (2002) also focused on parameter estimation and compared the performance of three autoregressive models (AR, CAR, SAR) with vole species abundance as the response variable and several vegetation descriptions as the predictor variables. They determined that the methods of incorporating spatial dependence produced similar results, and that it was more important to include spatial dependence than to quibble about which way to include it.

Segurado and Araújo (2004) predicted the probability of reptile and amphibian occurrence using a measure of contagion for each cell, defined as the weighted average number of occupied grid cells within a neighborhood. They found that this spatial term improved model accuracy more often for generalist species, as it reduced false positive predictions.

In a study based on Bayesian image restoration, Heikkinen and Hogmander (1994) also extended the autologistic model to include information on 'search intensity', which distinguishes between observed absence and unobserved sites (Hoeting et al., 2000; Wintle, 2003). Silva et al. (2002) modeled range expansion of an alien bird species in Portugal by specifying an

autocovariate term that included both spatial and temporal information.

Lichstein et al. (2002) investigated the effects of broad-scale (with a trend surface) and fine-scale (with a CAR model) spatial patterns on the relative importance of habitat predictor variables in a regression model of bird counts. They found that when spatial dependence is incorporated explicitly, the importance of environmental variables decreases, for both fine- and broad-scale spatial dependence, implying that the influence of highly spatially structured predictor variables will tend to show weaker effects when spatial dependence is incorporated (Lichstein et al., 2002; also see Overmars et al., 2003 for similar results in land use models). In a study that related mammalian species richness in South America to environmental variables, Tognelli and Kelt (2004) compared OLS regression to CAR and SAR models and found that, in addition to achieving better fit with the AR models, the relative importance of the explanatory variables shifted.

In a logistic model to predict presence of ground-nesting birds based on habitat factors, Milsom et al. (2000) sought to incorporate land-holding information, based on the assumption that marshes within a land-holding were more similar than marshes on different land-holdings. They found that an autologistic term representing spatial correlation among observations in nearby marshes was highly significant, unless a factor variable representing land-holding differences was also used, indicating that the ‘spatial autocorrelation’ was negligible beyond that which was described by land-holding (Milsom et al., 2000). Luoto et al. (2002) developed a logistic model to predict butterfly presence as a function of habitat and topographical variables, then added an autologistic term to assess the importance of spatial dependence. They found that the autologistic term was statistically not significant, although presumably the other predictor variables (grassland connectivity, deciduous forest cover, and grassland cover) could have made it superfluous (Luoto et al., 2002).

3.3. Geostatistical methods

Geostatistical interpolation methods, originally developed for mining applications, are concerned with modeling spatial dependence explicitly and they consider spatial dependence to be the most important, and often only, determinant of the predicted values. Kriging, one of the most widely used interpolation methods, attempts to optimize interpolation by dividing spatial variation into three components: deterministic variation, spatial autocorrelation (described by a variogram), and noise (Burrough and McDonnell, 1998).

Many of the steps involved in variogram fitting are subjective—selection of lag distance, model shape, sill, nugget and range values. The representativeness of a variogram also depends on the sample data—data that are too sparse will not adequately describe fine-scaled autocorrelation (lag distances are selected based on, among other criteria, a minimum number of data pairs that occur within them). For spatial prediction, if an unsampled cell is separated from an observation by a distance greater than the range, the observation makes no contribution to the predicted value of the unsampled cell and the unsampled cell is estimated to be either the mean value of all of the data values or null. Kriging is intended to describe a

spatial stochastic process that is stationary, with a mean and variance that are both constant and independent of location (Bailey and Gatrell, 1995). Differences in values between observations should be a function only of the distance between them (relative location), rather than their absolute location.

Kriging and other geostatistical techniques, specifically developed for spatial prediction, are relatively untested in vegetation mapping (likely because of the difficulty involved in including environmental predictors). The concept of the regionalized variable on which geostatistical prediction is based, describes continuous phenomena such as soil properties, elevation, atmospheric pressure, etc. As vegetation distribution is somewhat different being related to environmental gradients, that have a distinct spatial pattern resulting from a set of underlying spatial generating processes, use of geostatistical methods requires a slight theoretical adjustment. Geostatistical interpolation methods are based on a set of models of underlying spatial processes—not ecological processes.

Geostatistical prediction methods assume a constant mean value for the study area and focus on the second element, describing the spatial structure explicitly as a function of distance with a variogram. The deterministic element is typically removed (in a ‘detrended’ model), or its effects are modeled in combination with the second element (e.g., co-kriging, universal kriging). Variograms are considered a global spatial statistic, as they represent simultaneous measurement from many locations (Cliff and Ord, 1981). Increased attention has been focused on local spatial statistics that can measure spatial dependence in only a portion of the study area, but they require very large datasets with intensive sampling (Getis and Ord, 1996). Although most commonly used geostatistical prediction methods do not accommodate environmental predictors, a few types and modifications of kriging do, and it is these we discuss here.

While all kriging methods consider distance from observed values to be the most important predictor, co-kriging and universal kriging are the only kriging methods that allow ancillary (e.g., environmental) variables to be included in the calculation of predictions (but see the hybrid methods described below as further exceptions). Co-kriging (CK) incorporates information on the joint spatial co-variation between two variables in order to model one of them. Usually one of the variables is more easily measured and can be used to help explain fluctuations from the mean in the variable of interest (Bailey and Gatrell, 1995). The co-variogram used in co-kriging must be described by a variogram model that appears in both of the individual variograms, but it does not improve on ordinary kriging when both variograms are too similar (Bailey and Gatrell, 1995). Universal kriging (UK) extends ordinary kriging to accommodate a global trend, however the estimation of the residual variogram is problematic and probably explains why this is not a more commonly used method (Bailey and Gatrell, 1995).

Geostatistical methods have been widely used in predictive soil modeling (McBratney et al., 2003; Scull et al., 2003), and one relatively new hybrid method (regression kriging) was introduced in that context. Regression kriging (RK) assumes that the model residuals have a spatial structure, resulting from ‘model’ factors such as mis-specified or inadequate explanatory variables, or ‘real’ factors such as biotic processes that

cause spatial patterns. RK combines predictions from a regression model along with the resulting kriged residuals. Although similar in theory to universal kriging, RK methods allow for more flexible regression models to be used (e.g., GAMs, GLMs, even decision trees), and a wider variety of predictor variables to be included.

Regression-kriging methods focus on residuals from a deterministic model, which can comprise both the second and third elements. These methods combine information on the vegetation–environmental relationships through the deterministic model used as well as local and error components, and therefore show great potential in PVM research (Chong et al., 2001; Miller, 2005).

3.4. Applications of geostatistical methods

Variogram-based methods have a more extensive history in mining, geological, and soil applications, therefore we focus here on either novel geostatistical methods that have potential use in PVM applications, or geostatistical studies involving vegetation distribution. Pfeffer et al. (2003) used universal kriging, where the trend was a linear function of topographic variables, to interpolate detrended correspondence analysis (DCA) scores of alpine vegetation. Bishop and McBratney (2001) compared standard model results to regression kriging results for several different models (GLM, GAM, and DT) to model soil properties (regression kriging was only performed when there was spatial autocorrelation in the residuals). Using root mean square error (RMSE) measures for comparison, they found that the performance of all of the regression methods was improved when model residuals were incorporated by regression-kriging. Using only one ancillary variable, soil electrical conductivity, Knotters et al. (1995) found that maximum likelihood regression with kriged residuals performed better than co-kriging for predicting soil horizon depth. In comparisons with co-kriging, regression kriging with maximum likelihood estimation (Knotters et al., 1995) and multiple linear regression with least squares estimation (Odeh et al., 1995) resulted in higher prediction accuracies.

Using a terrain-based model to predict vegetation composition in the Southern Appalachian Mountains (North Carolina, USA), Bolstad et al. (1998) found that regression with log transformations resulted in higher map accuracy than kriging and co-kriging (but co-kriging only used one environmental variable, elevation or terrain shape, at a time). They concluded that kriging and co-kriging were not appropriate in this study because the scale/resolution of the sample data was probably too broad (30 and 80 m resolution) (Bolstad et al., 1998). Chong et al. (2001) found that adding kriged residuals to regression model predictions of plant species abundance in Colorado, USA improved accuracy, particularly for native species with a much larger range than for exotic species with a patchy distribution and smaller range. Although they used a multi-scale sampling approach, they still had difficulty predicting spatial variability within 1000 m² plots (Chong et al., 2001).

Although the relationship between tree species distribution and topographic and other environmental factors and land-use histories was noted, Hershey (2000) used only spatial information to create maps with indicator kriging and sequential Gaussian conditional simulation in a study in Penn-

sylvania, USA. She found that simulation resulted in higher accuracy with rare species, while kriging tended to ‘smooth out’ the distribution of rare species. Although these methods could be improved by including ancillary data in the analyses, they represented an improvement from previous maps based on county summaries.

3.5. Parameter estimation methods

The two methods discussed above, variogram- and neighborhood-based methods, focus on incorporating spatial dependence to increase the predictive ability of a model. The two following methods are discussed briefly, as they do involve explicit description of spatial dependence, but their focus is on generating unbiased parameter estimates. These methods are traditionally associated with longitudinal data analysis, where the correlation occurs across time or through different treatments in the same group, but have been used more recently with spatially correlated data.

Generalized linear mixed models (GLMM) extend a GLM by introducing a spatial random component in addition to the fixed components. GLMMs allow for correlation among observations as well as among the random effects and the resulting model can provide unbiased parameter estimates (Diggle et al., 1994; Stralberg and Bao, 1999; Kleinschmidt et al., 2001; Hooten et al., 2003). After fitting a generalized linear mixed model (GLMM) using malaria counts as the response variable and environmental factors as the predictor variables, Kleinschmidt et al. (2001) found that spatial dependence remained in the residuals. Adding a kriged residual surface to the model predictions improved overall map accuracy.

Generalized estimating equations (GEE) are a class of marginal models, where ‘marginal’ refers to modeling the mean response to the predictor variables, rather than the joint responses of all sites simultaneously, as in conditional models (Cressie, 1991). To build a GEE model, spatial dependence is explicitly incorporated in a GLM using an exponential covariance function (Albert and McShane, 1995; Gotway and Stroup, 1997; Gumpertz et al., 2000; Muggleston et al., 2002). In these types of models, spatial dependence and its use for prediction is of secondary interest, included only to obtain unbiased parameter estimates and their spatial error.

3.6. Geographically weighted regression

Although the methods described above recognize the local nature of spatial data (i.e., by explicitly modeling the covariance structure of the error terms, rather than assuming they are independent), the results are ultimately equations with global parameter estimates, in that the relationships they describe, e.g., between temperature and vegetation distribution, are consistent throughout the region of interest. However, in reality the relationships between certain environmental gradients and biogeographical phenomena are more complex than that, and may actually vary according to where they are located in geographical space, resulting in spatial nonstationarity. For example, the relationship between temperature and vegetation distribution may change from positive to negative depending on precipitation. When nonstationarity is evident, data partitioning or local methods are

more appropriate (Osborne and Suárez-Seoane, 2002; Foody, 2004b). The final method we describe, geographically weighted regression (GWR) allows for locational differences in relationships by estimating regression parameters and models that vary across space (see Fotheringham et al., 2002). All observations are used to fit the regression parameters, but the observations that are closest to a point are weighted more heavily than observations farther away. GWR extends a global regression model such as

$$y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i \quad (5)$$

by allowing parameter estimates to vary locally:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \quad (6)$$

where (u_i, v_i) are the coordinates of the i th point in space.

The geographical weighting is determined by a spatial kernel, whose geometric shape (circle, square), bandwidth (the rate at which the weight of increasingly distant observations is counted), and functional form (e.g., Gaussian, exponential) is determined by the user (Fotheringham et al., 2002). Adaptive spatial kernels with varying bandwidth can also be used to accommodate local regressions based on sparse data (Fotheringham et al., 2002).

Although GWR “incorporates local spatial relationships into the regression framework in an intuitive and explicit manner” (Fotheringham et al., 2002, p. 27), the manner requires a slight conceptual adjustment compared to the manners in which spatial relationships are incorporated with the previously summarized methods. Rather than representing spatial dependence with an additional term (autoregressive models) or with a variogram describing the average relationship between distance and similarity, GWR incorporates spatial dependence in each of the n local regression equations.

Although developed primarily for human geography applications (Jones and Cassetti, 1992) where the effect of location is more intrinsic and less based on physical processes (and presumably less ‘model-able’), there have been a few recent applications in physical geography (Atkinson et al., 2003; Foody, 2004a). Foody (2004a) investigated the relationship between temperature, precipitation, and NDVI with avian species richness, and found that global parameters describing the relationships were insufficient and inadequate. Wang et al. (2005) found that GWR performed better than both OLS and a spatial lag model in a study relating net primary production to environmental variables in Chinese forest ecosystems.

Zhang et al. (2004) were interested in exploring the spatial heterogeneity in the relationship between tree diameter and total height in a forest stand. They found that GWR produced significantly smaller residuals than OLS. In a model relating deer distribution to climate and land cover patch metrics in Michigan’s Upper Peninsula, Shi et al. (2006) compared both local and global kernel bandwidth GWR models to OLS and found that OLS performed the worst and the local bandwidth GWR performed the best.

Fotheringham et al. (2002) also suggest an interesting modification to the general GWR equation where an autoregressive

term is added, the result of which can be used to produce a surface of local estimates of the spatial autocorrelation in the response variable. A method that extends GWR by incorporating variogram functions to account for spatial dependence, and may be more appropriate for prediction, is also being addressed (Páez et al., in preparation).

However, Jetz et al. (2005) suggest that GWR should not be used *instead* of global regression techniques, but rather as a supplement to global regression. They suggest that many of the local relationships captured in GWR can also be specified by including interaction terms in global methods (although this requires *a priori* knowledge of important interaction effects). Generally, GWR does not address the issue of spatially correlated residuals, although such extensions are in development (Fotheringham et al., 2002). Further, Austin (2007) notes that incorrectly specified (linear) models could be producing what appears to be nonstationarity, and a correctly specified model should be used for more ecologically appropriate predictions.

3.7. Other methods

Although he did not use an autologistic model, Leathwick (1998) incorporated neighborhood information informally in a study to determine environmental and spatial factors in tree distribution in New Zealand. He included a dummy variable in a binomial GAM that indicated whether a species was present in an adjacent plot (grid cell) within a specified radius. In comparing the prediction results from the environmental models and environmental and spatial models, he found that environmental factors alone did not adequately explain the distribution. Dirnböck and Dullinger (2004) used a similar method to explore the environmental and spatial influences of alpine plant species abundance in a high mountain region of Austria. They found significant spatial patterning independent of the environmental predictors, and concluded it was most likely a function of disturbance.

In a study investigating distribution of snow petrels in East Antarctica, Olivier and Wotherspoon (2005) expected the colonial nature of snow petrels and conspecific attraction to play an important role in their spatial clustering. They used two different variables, average distance to a neighboring nest, and average number of neighboring nests to describe these effects. They found that the spatial dependence variables greatly improved the model fit, but were less useful for prediction with new data (Olivier and Wotherspoon, 2005).

4. Summary

PVM seeks to quantify ecological relationships that have been observed for hundreds of years and has as its foundation ecological theories that are now widely accepted. Positioned at the confluence of ecology, statistical analysis and geographic information science, PVM has benefited from the broad expertise of its practitioners but there is still room for improvement. Austin (2002) points out that statisticians may formulate models that are ecologically inappropriate and ecologists may not take advantage of more sophisticated modeling techniques. Spatial dependence is an important issue in this context, as

it requires some understanding of ecological processes in its manifestation, and its incorporation in models obviates the use of many traditional statistical methods.

Much attention has been placed on models in ecology, particularly in the context of changing environmental conditions associated with global change. Increased computing ability and available technology have rapidly advanced model capabilities, but important objectives of reducing uncertainty and ambiguity in predictions remain (Clark et al., 2001). In addition to potentially improving model accuracy, incorporating spatial dependence in PVM as an element of variation in spatial data can help to clarify the influence of predictor variables and improve accuracies of prediction for response variables. A variety of approaches and methods for this have been described above and they are summarized below along with their respective advantages and disadvantages.

Autoregressive methods can be more appropriate when the spatial dependence is intrinsic and causes local clusters in the data. Specifying an autocovariate term can help to clarify the influence of the other predictor variables, as well as add to the explanatory value of the model. If the autocovariate term is not significant, the model is equivalent to a nonspatial regression model. However, Austin (2002) suggests that even a significant autocovariate term could be indicative of a mis-specified model (e.g., linear relationship where it should be nonlinear) rather than spatial dependence. Wu and Huffer (1997) note that the MCMC estimation methods can be numerically unstable in the presence of strong intrinsic spatial dependence. In addition, the iterative procedures for fitting the full model are computationally intensive and software routines for these purposes are not readily available. Finally, the ability to calculate an appropriate autocovariate term is a function of the sample data, which is often not available at appropriate spatial lags and densities.

The most common geostatistical method, ordinary kriging, is generally not appropriate for predictive vegetation modeling, as it does not accommodate environmental predictors. Co-kriging can include information on one environmental predictor by modeling the response variable using information on the joint spatial co-variation between response and single predictor variable. However, it does not improve on ordinary kriging when both variograms are too similar. Universal kriging extends ordinary kriging to accommodate a global trend (which would be appropriate for extrinsic spatial dependence), but estimating the residual variogram remains a challenge to implementation.

In biogeographical applications, geographically weighted regression has so far been used mainly as a data exploration technique rather than a predictive method. The resulting map of spatially varying model coefficients can then be used to suggest patterns or potential missing variables. Austin (2007) has suggested that the standard linear relationships used in GWR models could be erroneously producing what appears to be nonstationarity, and a correctly specified model should be used instead. Parameter estimation models include spatial dependence not for predictive purposes, but to obtain unbiased parameter estimates and their spatial error.

As discussed above, the utility of models in which spatial dependence can be incorporated explicitly differs based on the goals of the research and characteristics of the data used. In

addition to these model differences, some important issues still remain with respect to incorporating spatial dependence in PVM:

- *Model data*—Rempel and Kushneriuk (2003) discuss two crucial and related uncertainties for spatial mapping as determining the best sampling scheme to estimate spatial dependence, and the best interpolation technique to map the resulting pattern. Compiled datasets, typically based on sampling strategies that emphasized independence among the observations, are generally not available at an adequate (spatial) sampling density amenable to analysis of spatial dependence. When the importance of incorporating spatial dependence becomes more widely accepted, sampling schemes can be designed that achieve that objective (possibly a two-stage process involving the variogram or other measures of spatial dependence to determine the appropriate scale of spatial dependence).
- *The causes of observed spatial dependence*—Is spatial dependence a result of a fine-scale stochastic biotic process, or is it, more likely, the result of a broad-scale deterministic gradient? Bailey and Gatrell (1995) note that these effects are confounded in observed data, and any separation between them is ultimately arbitrary. Both produce similar spatial patterns and demonstrate similar diagnostic results (i.e., autocorrelated residuals). As it stands, incorporating spatial dependence is largely of use for prediction purposes rather than inference purposes at present.
- *Characterizing spatial dependence*—Should spatial dependence be characterized as a function of contiguity, a discrete neighborhood or distance (which is also often based on selection of a discrete neighborhood size (lag) in the case of variogram estimation)? Each of these approaches produces a different representation of spatial dependence. Ideally, spatial dependence will be characterized by an understanding of the ecological processes that produce it.
- *Approach for incorporating spatial dependence in a model*—A variety of approaches have been discussed. Should spatial dependence be used as an explanatory variable, in addition to, or instead of environmental variables? Should a 'global' model be fit that is presumed to be representative of the average relationship between response and predictors, and any local spatial variation be contained in the residuals? Or should the entire relationship be treated as being explicitly spatial and thus locally varying? Should the spatial and environmental components of a model be separated, and is this possible or necessary?

While this paper has focused on incorporating spatial dependence in PVM for predictive purposes, explicitly including spatial dependence in models can also aid in clarifying the effect of different explanatory variables, thereby improving inferences that can be drawn (Wu and Huffer, 1997; Legendre and Legendre, 1998; Hoeting et al., 2006). The ability to identify whether an environmental variable has a physical influence on vegetation distribution, or merely covaries spatially with it will allow development of more process-based models (providing greater predictive power) and added insight in vegetation–environment relationships.

Acknowledgments

This research was supported by the WVU Geology and Geography Department, SDSU Department of Geography, and the National Science Foundation (award #0451486). In addition to comments and suggestions by the editor and anonymous reviewers, discussions with M. Goodchild, J. Michaelsen, S. Rey, J. Rogan, K. van Niel and all of the participants of the Riederalp Workshop (Switzerland, 2001 and 2004) are gratefully acknowledged.

REFERENCES

- Albert, P.S., McShane, L.M., 1995. A generalized estimating equations approach for spatially correlated binary data: applications to the analysis of neuroimaging data. *Biometrics* 51, 627–638.
- Anselin, L., 1992. *Spatial Data Analysis with GIS: An Introduction to Application in the Social Sciences*. NCGIA, Santa Barbara, CA, pp. 10–92.
- Anselin, L., 1993. Discrete space autoregressive models. In: Goodchild, M., Parks, B., Steyaert, L. (Eds.), *Environmental Modeling with GIS*. Oxford University Press, Oxford, pp. 454–469.
- Aspinall, R.J., 1999. GIS for landscape conservation. In: Longley, P., Maguire, D.M., Goodchild, M.F., Rhind, D.R. (Eds.), *GIS: Principles, Techniques, Applications and Management*. Wiley, New York, pp. 967–980.
- Atkinson, P.M., German, S.E., Sear, D.A., Clark, M.J., 2003. Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression. *Geograph. Anal.* 35, 58–82.
- Atkinson, P.M., Tate, N.J., 2000. Spatial scale problems and geostatistical solutions: a review. *Professional Geograph* 52, 607–623.
- Augustin, N., Borchers, D., Muggleston, M., Buckland, S., 1996a. Regression methods with spatially referenced data. *Asp. Appl. Biol.* 46, 67–74.
- Augustin, N., Muggleston, M., Buckland, S., 1996b. An autologistic model for the spatial distribution of wildlife. *J. Appl. Ecol.* 33, 339–347.
- Augustin, N., Muggleston, M., Buckland, S., 1998. The role of simulation in modelling spatially correlated data. *Environmetrics* 9, 175–196.
- Austin, M.P., 1980. Searching for a model for use in vegetation analysis. *Vegetatio* 42, 11–21.
- Austin, M.P., 1985. Continuum concept, ordination methods, and niche theory. *Annu. Rev. Ecol. Systemat.* 16, 39–61.
- Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecol. Model.* 157, 101–118.
- Austin, M.P., 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecol. Model.* 200, 1–19.
- Austin, M.P., Smith, T.M., 1989. A new model for the continuum concept. *Vegetatio* 83, 35–47.
- Bailey, T.C., Gatrell, A.C., 1995. *Interactive Spatial Data Analysis*. Longman, Essex.
- Bellehumeur, C., Legendre, P., 1998. Multiscale sources of variation in ecological variables: modeling spatial dispersion, elaborating sampling designs. *Landsc. Ecol.* 13, 15–25.
- Berlow, E., 1999. Strong effects of weak interactions in ecological communities. *Nature* 398, 330–334.
- Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Stat. Soc., Ser. B* 36, 192–236.
- Besag, J., 1972. Nearest-neighbour systems and the autologistic model for binary data. *J. Roy. Stat. Soc. B* 34, 75–83.
- Betts, M.A., Diamond, G., Forbes, M.-A., Villard, Gunn, J., 2006. The importance of spatial autocorrelation, extent and resolution in predicting forest bird occurrence. *Ecol. Model.* 191, 197–224.
- Bishop, T.F.A., McBratney, A.B., 2001. A comparison of prediction methods for the creation of field-extent soil property maps. *Geoderma* 103, 149–160.
- Bolstad, P.V., Swank, W., Vose, J., 1998. Predicting Southern Appalachian overstory vegetation with digital terrain data. *Landsc. Ecol.* 13, 271–283.
- Borcard, D., Legendre, P., Drapeau, P., 1992. Partialling out the spatial component of ecological variation. *Ecology* 73, 1045–1055.
- Bullock, B.P., Burkhart, H.E., 2005. An evaluation of spatial dependency in juvenile loblolly pine stands using stem diameter. *For. Sci.* 51 (2), 102–108.
- Burrough, P., McDonnell, R., 1998. *Principles of Geographical Information Systems*. Oxford University Press, Oxford.
- Cairns, D.M., 2001. A comparison of methods for predicting vegetation type. *Plant Ecol.* 156, 3–18.
- Carroll, S.S., Pearson, D.L., 1998. Spatial modeling of butterfly species richness using tiger beetles (Cicindelidae) as a bioindicator taxon. *Ecol. Appl.* 8, 531–543.
- Chong, G.W., Reich, R.M., Kalkhan, M.A., Stohlgren, T.J., 2001. New approaches for sampling and modeling native and exotic plant species richness. *Western N. Am. Natur.* 61, 328–335.
- Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dobson, A., Foley, J.A., Lodge, D.M., Pascual, M., Pielke, R., Pizer, W., Pringle, C., Reid, W.V., Rose, K.A., Sala, O., Schlesinger, W., Wall, D.H., Wear, D., 2001. Ecological forecasts: an emerging imperative. *Science* 293, 657–660.
- Cliff, A., Ord, J., 1981. *Spatial Processes: Models and Applications*. Pion Limited, London.
- Cressie, N., 1991. *Statistics for Spatial Data*. John Wiley and Sons, New York.
- Dale, M.R.T., Dixon, P., Fortin, M.-J., Legendre, P., Myers, D.E., Rosenberg, M.S., 2002. Conceptual and mathematical relationships among methods for spatial analysis. *Ecography* 25, 558–577.
- Dark, S.J., 2004. The biogeography of invasive alien plants in California: an application of GIS and spatial regression analysis. *Diver. Distribut.* 10, 1–9.
- Davis, F., Goetz, S., 1990. Modeling vegetation pattern using digital terrain data. *Landsc. Ecol.* 4, 69–80.
- Dennis, R.L.H., Shreeve, T.G., Sparks, T.H., Lhonore, J.E., 2002. A comparison of geographical and neighbourhood models for improving atlas databases. The case of the French butterfly atlas. *Biol. Conserv.*, 108.
- Diggle, P.J., Liang, K.-Y., Zeger, S.L., 1994. *Analysis of Longitudinal Data*. Oxford University Press, Oxford.
- Diniz-Filho, J.A.F., Bini, L.M., Hawkins, B.A., 2003. Spatial autocorrelation and red herrings in geographical ecology. *Global Ecol. Biogeogr.* 12, 53–64.
- Dirnböck, T., Dullinger, S., 2004. Habitat distribution models, spatial autocorrelation, functional traits and dispersal capacity of alpine plant species. *J. Veg. Sci.* 15, 77–84.
- Dungan, J.L., Perry, J.N., Dale, M.R.T., Legendre, P., Citron-Pousty, S., Fortin, M.-J., Jakomulska, A., Miriti, M., Rosenberg, M.S., 2002. A balanced view of scale in spatial statistical analysis. *Ecography* 25, 626–640.
- Ferrier, S., Watson, G., Pearce, J., Drielsma, M., 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in north-east New South Wales. I. Species level modelling. *Biodiver. Conserv.* 11, 2275–2307.

- Foody, G.M., 2004a. Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. *Global Ecol. Biogeogr.* 13, 315–320.
- Foody, G.M., 2004b. GIS: stressing the geographical. *Progr. Phys. Geogr.* 28 (1), 152–215.
- Fortin, M.-J., 1999. Spatial statistics in landscape ecology. In: Klopatek, J.M., Gardner, R.H. (Eds.), *Landscape Ecological Analysis: Issues and Applications*. Springer-Verlag, New York.
- Fotheringham, A.S., Brunsdon, C., Charlton, M., 2002. *Geographically Weighted Regression: The analysis of spatially varying relationships*. Wiley & Sons, Chichester.
- Franklin, J., 1995. Predictive vegetation mapping: geographic modeling of biospatial patterns in relation to environmental gradients. *Progr. Phys. Geogr.* 19, 474–499.
- Franklin, J., 1998. Predicting the distributions of shrub species in California chaparral and coastal sage communities from climate and terrain-derived variables. *J. Veg. Sci.* 9, 733–748.
- Getis, A., Ord, J.K., 1996. Local spatial statistics: an overview. In: Longley, P., Batty, M. (Eds.), *Spatial Analysis: Modelling in a GIS Environment*. John Wiley & Sons, New York, pp. 261–277.
- Goodchild, M.F., 2001. Models of scale and scales of modelling. In: Tate, N.J., Atkinson, P.M. (Eds.), *Modelling Scale in Geographical Information Science*. John Wiley & Sons, Chichester.
- Goodchild, M.F., 2004. The validity and usefulness of laws in geographic information science and geography. *Ann. Assoc. Am. Geograph.* 94 (2), 300–303.
- Goodchild, M.F., Proctor, J., 1997. Scale in a digital geographic world. *Geograph. Environ. Model.* 1, 5–23.
- Gotway, C., Stroup, W., 1997. A generalized linear model approach to spatial data analysis and prediction. *J. Agric. Biol. Environ. Stat.* 2, 152–178.
- Gotway, C.A., Hartford, A.H., 1996. Geostatistical methods for incorporating auxiliary information in the prediction of spatial variables. *J. Agric. Biol. Environ. Stat.* 1, 17–39.
- Guisan, A., Zimmermann, N., 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135, 147–186.
- Gumpertz, M., Graham, J., Ristaino, J., 1997. Autologistic model of spatial pattern of *Phytophthora* epidemic in bell pepper: effects of soil variation on disease presence. *J. Agric. Biol. Environ. Stat.* 2, 131–156.
- Gumpertz, M., Wu, C., Pye, H., 2000. Logistic regression for southern pine beetle outbreaks with spatial and temporal autocorrelation. *For. Sci.* 46, 95–107.
- Haining, R., 1989. *Geography and spatial statistics: current positions, future developments*. In: Macmillan, B. (Ed.), *Remodelling Geography*. Basil Blackwell, Cambridge, MA, pp. 191–203.
- Haining, R., 1990. *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press, Cambridge, UK.
- He, F., Zhou, J., Zhu, H., 2003. Autologistic regression model for the distribution of vegetation. *J. Agric. Biol. Environ. Stat.* 8, 205–222.
- Heikkinen, J., Hogmander, H., 1994. Fully Bayesian approach to image restoration with an application to biogeography. *Appl. Stat.* 43, 569–582.
- Hershey, R.R., 2000. Modeling the spatial distribution of ten tree species in Pennsylvania. In: Mowrer, H.T., Congalton, R.G. (Eds.), *Quantifying Spatial Uncertainty in Natural Resources: Theory and Applications for GIS and Remote Sensing*. Ann Arbor Press, Chelsea, MI, pp. 119–135.
- Hoeting, J., Leecaster, M., Bowden, D., 2000. An improved model for spatially correlated binary responses. *J. Agric. Biol. Environ. Stat.* 5, 102–114.
- Hoeting, J.A., Davis, R.A., Merton, A.A., Thompson, S.E., 2006. Model selection for geostatistical models. *Ecol. Appl.* 16, 87–98.
- Holdridge, L., 1967. *Life Zone Ecology*. Tropical Science Center, San Jose, Costa Rica.
- Hooten, M.B., Larsen, D.R., Wikle, C.K., 2003. Predicting the spatial distribution of ground flora on large domains using a hierarchical Bayesian model. *Landsc. Ecol.* 18, 487–502.
- Hubbell, S.P., Ahumada, J.A., Condit, R., Foster, R.B., 2001. Local neighborhood effects on long-term survival of individual trees in a neotropical forest. *Ecol. Res.* 16, 859–875.
- Huffer, F., Wu, H., 1998. Markov chain Monte Carlo for autologistic regression models with application to the distribution of plant species. *Biometrics* 54, 70–85.
- Jeganathan, P., Green, R.E., Norris, K., Vogiatzakis, I.N., Bartsch, A., Wotton, S.R., Bowden, C.G.R., Griffiths, G.H., Pain, D., Rahmani, A.R., 2004. Modelling habitat selection and distribution of the critically endangered Jerdon's courser *Rhinoptilus bitorquatus* in scrub jungle: an application of a new tracking method. *J. Appl. Ecol.* 41, 224–237.
- Jetz, W., Rahbek, C., Lichstein, J.W., 2005. Local and global approaches to spatial data analysis in ecology. *Global Ecol. Biogeogr.* 14, 97–98.
- Jones, J.P., Cassetti, E., 1992. *Applications of the Expansion Method*. Routledge, London.
- Jongman, R., Ter Braak, C., Van Tongeren, O. (Eds.), 1995. *Data Analysis in Community and Landscape Ecology*. Cambridge University Press, Cambridge.
- Keitt, T.H., Bjornstad, O.N., Dixon, P.M., Citron-Pousty, S., 2002. Accounting for spatial pattern when modeling organism–environment interactions. *Ecography* 25, 616–625.
- Kessell, S., 1979. *Gradient Modeling: Resource and Fire Management*. Springer-Verlag, New York.
- Kleinschmidt, I., Sharp, B.L., Clarke, G.P.Y., Curtis, B., Fraser, C., 2001. Use of generalized linear mixed models in the spatial analysis of small-area malaria incidence rates in KwaZulu Natal South Africa. *Am. J. Epidemiol.* 153, 1213–1221.
- Klute, D.S., Lovallo, M.J., Tzilkowski, W.M., 2002. Autologistic regression modeling of American woodcock habitat use with spatially dependent data. In: Scott, J., Heglund, P., Morrison, M., Haufler, J., Raphael, M., Wall, W., Samson, F. (Eds.), *Predicting Species Occurrences; Issues of Accuracy and Scale*. Island Press, Washington, pp. 335–344.
- Knapp, R., Matthews, K., Preisler, H., Jellison, R., 2003. Developing probabilistic models to predict amphibian site occupancy in a patchy landscape. *Ecol. Appl.* 13, 1069–1082.
- Knotters, M., Brus, D.J., Oude Voshaar, J.H., 1995. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. *Geoderma* 67, 227–246.
- Leathwick, J., 1998. Are New Zealand's *Nothofagus* species in equilibrium with their environment? *J. Veg. Sci.* 9, 719–732.
- Le Duc, M., Hill, M., Sparks, T., 1992. A method for predicting the probability of species occurrence using data from systematic surveys. *Watsonia* 19, 97–105.
- Legendre, P., 1993. Spatial autocorrelation: problem or new paradigm? *Ecology* 74, 1659–1673.
- Legendre, P., Dale, M.R.T., Fortin, M.-J., Gurevitch, J., Hohn, M., Myers, D., 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography* 25, 601–615.
- Legendre, P., Fortin, M.J., 1989. Spatial pattern and ecological analysis. *Vegetatio* 80, 107–138.
- Legendre, P., Legendre, L., 1998. *Numerical Ecology*, 2nd English ed. Elsevier, Amsterdam.
- 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. *Ecology* 73, 1943–1967.

- Lichstein, J.W., Simons, T.R., Shriner, S.A., Franzreb, K.E., 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecol. Monogr.* 72, 445–463.
- Lilburne, L.R., Webb, T.H., Benwell, G.L., 2004. The Scale Matcher: a procedure for assessing scale compatibility of spatial data and models. *Int. J. Geograph. Inform. Sci.* 18, 257–279.
- Lobo, J.M., Jay-Robert, P., Lumaret, J.-P., 2004. Modelling the species richness distribution for French Aphodiidae (Coleoptera, Scarabaeoidea). *Ecography* 27, 145–156.
- Lobo, J.M., Lumaret, J.-P., Jay-Robert, P., 2002. Modelling the species richness of French dung beetles (Coleoptera, Scarabaeidae) and delimiting the predictive capacity of different groups of explanatory variables. *Global Ecol. Biogeogr.* 11, 265–277.
- Lortie, C., Brooker, R., Choler, P., Kikvidze, Z., Michalet, R., Pugnaire, F., Callaway, R., 2004. Rethinking plant community theory. *Oikos*, 102.
- Luoto, M., Kuussaari, M., Toivonen, T., 2002. Modelling butterfly distribution based on remote sensing data. *J. Biogeogr.* 29, 1027–1037.
- MacNally, R., 2002. Multiple regression and inference in ecology and conservation biology: further comments on identifying important predictor variables. *Biodiver. Conserv.* 11, 1397–1401.
- Matsui, T., Yagihashi, T., Nakaya, T., Tanaka, N., Taoda, H., 2004. Climatic controlled on distribution of *Fagus crenata* forests in Japan. *J. Veg. Sci.* 15, 57–66.
- Mattson, D.J., Merrill, T., 2002. Extirpations of grizzly bears in the contiguous United States, 1850–2000. *Conserv. Biol.* 16, 1123–1136.
- McBratney, A.B., Mendonca Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117, 3–52.
- McMillen, D.P., 2003. Spatial autocorrelation or model misspecification? *Int. Regional Sci. Rev.* 26, 208–217.
- Meentemeyer, R., Moody, A., Franklin, J., 2001. Landscape-scale patterns of shrub-species abundance in California chaparral: the role of topographically mediated resource gradients. *Plant Ecol.* 156, 19–41.
- Miller, J., 2005. Incorporating spatial dependence in predictive vegetation models: residual interpolation methods. *Professional Geograph.* 57 (2), 169–184.
- Miller, J., Franklin, J., 2002. Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecol. Model.* 157, 227–247.
- Miller, J.R., Turner, M.G., Smithwick, E.A.H., Dent, C.L., Stanley, E.H., 2004. Spatial extrapolation: the science of predicting ecological patterns and processes. *BioScience* 54, 310–320.
- Milsom, T.P., Langton, S.D., Parkin, W.K., Peel, S., Bishop, J.D., Hart, J.D., Moore, N.P., 2000. Habitat models of bird species' distribution: an aid to the management of coastal grazing marshes. *J. Appl. Ecol.* 37, 706–727.
- Mugglestone, M.A., Kenward, M.G., Clark, S.J., 2002. Generalized estimating equations for spatially referenced binary data. In: Gregori, D., Carmeci, G., Friedl, H., Ferligoj, A., Wedlin, A. (Eds.), *Correlated Data Modeling*. FrancoAngeli S.R.L., Trieste, pp. 121–130.
- Nogués-Bravo, D., Martínez-Rica, J.P., 2004. Factors controlling the spatial species richness pattern of four groups of terrestrial vertebrates in an area between two different biogeographic regions in northern Spain. *J. Biogeogr.* 31, 629–640.
- Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1995. Further results on prediction of soil properties from terrain attributes: heterotropic cokriging and regression-kriging. *Geoderma* 67, 215–226.
- Olivier, F., Wotherspoon, S.J., 2005. GIS-based application of resource selection functions to the prediction of snow petrel distribution and abundance in East Antarctica: comparing models at multiple scales. *Ecol. Model.* 189, 105–129.
- Osborne, P.E., Alonso, J.C., Bryant, R.G., 2001. Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. *J. Appl. Ecol.* 38, 458–471.
- Osborne, P., Suárez-Seoane, S., 2002. Should data be partitioned spatially before building large-scale distribution models? *Ecol. Model.* 157, 249–259.
- Overmars, K.P., de Koning, G.H.J., Veldkamp, A., 2003. Spatial autocorrelation in multi-scale land use models. *Ecol. Model.* 164, 257–270.
- Páez, A., Kawai, K., Vichiensan, V., Miyamoto, K., in preparation. Nonstationarity in regression-based spatial interpolation models. Centre for Spatial Analysis, McMaster University, Hamilton, Ontario.
- Pereira, J., Itami, R., 1991. GIS-based habitat modeling using logistic multiple regression: a study of the Mt. Graham Red Squirrel. *Photogr. Eng. Remote Sensing* 57, 1475–1486.
- Perry, J.N., Liebhold, A.M., Rosenberg, M.S., Dungan, J., Miriti, M., Jakomulska, A., Citron-Pousty, S., 2002. Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. *Ecography* 25, 578–600.
- Pfeffer, K., Pebesma, E.J., Burrough, P.A., 2003. Mapping alpine vegetation using vegetation observations and topographic attributes. *Landsc. Ecol.* 18, 759–776.
- Preisler, H.K., 1993. Modelling spatial patterns of trees attacked by bark beetles. *Appl. Stat.* 42, 501–514.
- Preisler, H.K., Rappaport, N.G., Wood, D.L., 1997. Regression methods for spatially correlated data: an example using beetle attacks in a seed orchard. *For. Sci.* 43, 71–77.
- Quattrochi, D.A., Goodchild, M.F. (Eds.), 1997. *Scale in Remote Sensing and GIS*. CRC Press, New York.
- Rempel, R.S., Kushneriuk, R.S., 2003. The influence of sampling scheme and interpolation method on the power to detect spatial effects of forest birds in Ontario (Canada). *Landsc. Ecol.* 18, 741–757.
- Rushton, S.P., Ormerod, S.J., Kerby, G., 2004. New paradigms for modelling species distributions? *Journal of Applied Ecology* 41, 193–200.
- Schenk, H., Holzapfel, C., Hamilton, J., Mahall, B., 2003. Spatial ecology of a small desert shrub on adjacent geological substrates. *J. Ecol.* 91, 383–395.
- Schwarz, P.A., Fahey, T.J., McCulloch, C.E., 2003. Factors controlling spatial variation of tree species abundance in a forested landscape. *Ecology* 84 (7), 1862–1878.
- Scott, J., Davis, F., Csuti, B., Noss, R., Butterfield, B., Groves, C., Anderson, H., Caicco, S., D'Erchia, F., Edwards, T., Ulliman, J., Wright, R., 1993. Gap analysis: a geographic approach to protection of biological diversity. *Wildl. Monogr.* 123, 1–41.
- Scull, P., Franklin, J., Chadwick, O.A., McArthur, D., 2003. Predictive soil mapping: a review. *Progr. Phys. Geogr.* 27, 171–197.
- Segurado, P., Araújo, M., 2004. An evaluation of methods for modelling species distributions. *J. Biogeogr.* 31, 1555–1568.
- Shi, H., Laurent, E., LeBouton, J., Racevskis, L., Hall, K., Donovan, M., Doepker, R., Walters, M., Lupi, F., Liu, J., 2006. Local spatial modeling of white-tailed deer distribution. *Ecol. Model.* 190, 171–189.
- Silva, T., Reino, L.M., Borralho, R., 2002. A model for range expansion of an introduced species: the common waxbill *Estrilda astrild* in Portugal. *Diver. Distribut.* 8, 319–326.
- Smith, P.A., 1994. Autocorrelation in logistic regression modeling of species' distributions. *Global Ecol. Biogeogr. Lett.* 4, 47–61.
- Sokal, R.R., Oden, N.L., 1978b. Spatial autocorrelation in biology. 2. Some biological implications and four applications of evolutionary and ecological interest. *Biol. J. Linn. Soc.* 10, 229–249.
- Stralberg, D., Bao, S., 1999. Identifying the spatial structure in error terms with spatial covariance models: a case study on

- urbanization influence in chaparral bird species. *Geogr. Inform. Sci.* 5, 106–120.
- Sui, D., 2004. Tobler's first law of geography: a big idea for a small world? *Ann. Assoc. Am. Geograph.* 94, 269–277.
- Tate, N.J., Atkinson, P.M. (Eds.), 2001. *Modelling Scale in Geographical Information Science*. John Wiley & Sons, Chichester.
- Thomson, J., Weiblen, G., Thomson, B., Alfaro, S., Legendre, P., 1996. Untangling multiple factors in spatial distributions: lilies, gophers and rocks. *Ecology* 77, 1698–1715.
- Titeux, N., Dufrene, M., et al., 2004. Multivariate analysis of a fine-scale breeding bird atlas using a geographical information system and partial canonical correspondence analysis: environmental and spatial effects. *J. Biogeogr.* 31, 1841–1856.
- Tobin, P.C., 2004. Estimation of the spatial autocorrelation function: consequences of sampling dynamic populations in space and time. *Ecography* 27, 767–775.
- Tobler, W., 1979. Cellular geography. In: Gale, G.O.S. (Ed.), *Philosophy in Geography*. Reidel, Dordrecht, pp. 379–389.
- Tognelli, M.F., Kelt, D.A., 2004. Analysis of determinants of mammalian species richness in South America using spatial autoregressive models. *Ecography* 27, 427–436.
- Vaughan, I.P., Ormerod, S.J., 2003. Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. *Conserv. Biol.* 17 (6), 1601–1611.
- Ver Hoef, J., 1993. Universal kriging for ecological data. In: Goodchild, M., Parks, B., Steyaert, L. (Eds.), *Environmental Modeling with GIS*. Oxford University Press, Oxford, pp. 447–453.
- Wall, M.M., 2004. A close look at the spatial structure implied by the CAR and SAR models. *J. Stat. Plan. Inference* 121, 311–324.
- Wang, Q., Ni, J., Tenhunen, J., 2005. Application of geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems. *Global Ecol. Biogeogr.* 14, 379–393.
- Weir, I.S., Pettitt, A.N., 2000. Binary probability maps using a hidden conditional autoregressive Gaussian process with an application to Finnish common toad data. *Appl. Stat.* 49, 473–484.
- Whittaker, R., 1973. Direct gradient analysis. In: Whittaker, R. (Ed.), *Ordination and Classification of Communities*. Dr. W. Junk, The Hague, pp. 9–45.
- Wintle, B.A., 2003. *Dealing with uncertainty in wildlife habitat models*. Ph.D. Dissertation. University of Melbourne, Melbourne.
- Wu, H., Huffer, F.W., 1997. Modelling the distribution of plant species using the autologistic regression model. *Environ. Ecol. Stat.* 4, 49–64.
- Zhang, L., Bi, H., Cheng, P., Davis, C., 2004. Modeling spatial variation in tree diameter–height relationships. *For. Ecol. Manage.* 189, 317–329.