

## Accounting for spatial pattern when modeling organism-environment interactions

Timothy H. Keitt, Ottar N. Bjørnstad, Philip M. Dixon and Steve Citron-Pousty

Keitt, T. H., Bjørnstad, O. N., Dixon, P. M. and Citron-Pousty, S. 2002. Accounting for spatial pattern when modeling organism-environment interactions. – *Ecography* 25: 616–625.

Statistical models of environment-abundance relationships may be influenced by spatial autocorrelation in abundance, environmental variables, or both. Failure to account for spatial autocorrelation can lead to incorrect conclusions regarding both the absolute and relative importance of environmental variables as determinants of abundance. We consider several classes of statistical models that are appropriate for modeling environment-abundance relationships in the presence of spatial autocorrelation, and apply these to three case studies: 1) abundance of voles in relation to habitat characteristics; 2) a plant competition experiment; and 3) abundance of Oribatid mites along environmental gradients. We find that when spatial pattern is accounted for in the modeling process, conclusions about environmental control over abundance can change dramatically. We conclude with five lessons: 1) spatial models are easy to calculate with several of the most common statistical packages; 2) results from spatially-structured models may point to conclusions radically different from those suggested by a spatially independent model; 3) not all spatial autocorrelation in abundances results from spatial population dynamics; it may also result from abundance associations with environmental variables not included in the model; 4) the different spatial models do have different mechanistic interpretations in terms of ecological processes – thus ecological model selection should take primacy over statistical model selection; 5) the conclusions of the different spatial models are typically fairly similar – making any correction is more important than quibbling about which correction to make.

*T. H. Keitt (timothy.keitt@stonybrook.edu), National Center for Ecological Analysis and Synthesis, Univ. of California Santa Barbara, Santa Barbara, CA 93101, USA. – O. N. Bjørnstad, Dept of Entomology Biology, Pennsylvania State Univ., University Park, PA 16802, USA. – P. M. Dixon, Dept of Statistics, Iowa State Univ., Ames, IA 50011-1210 USA. – S. Citron-Pousty, Dept of Ecology and Evolutionary Biology, Univ. of Connecticut, Storrs, CT 06269-3043, USA.*

The environment-abundance relationship has been featured as a central topic throughout the history of ecology. Classically, it was a main topic in order to understand limits to species' distributions, and species' responses to environmental gradients. More recently, understanding how qualitative and quantitative aspects of environment affects abundance and distribution has taken on a more acute practical dimension, because of the conservation implications of anthropogenic habitat loss and climate change. The critical link is that in

order to predict the effects of anthropogenic or other changes, we need reliable measures of the association between abundance and environmental variables. The estimation of these associations is the topic of this paper – the statistical complication is that classical methods used to quantify environment-abundance associations assume independence of observations. However, the distribution or abundance of a species is typically spatially autocorrelated due to locomotory constraints (e.g., Orians and Pearson 1979, Abrahams

Accepted 23 January 2002

Copyright © ECOGRAPHY 2002  
ISSN 0906-7590

1986), social organization (e.g., Stamps 1988, Morris et al. 1992), or aggregative responses to cues from conspecifics (Turchin and Kareiva 1989, Turchin and Thoeny 1993). At the same time the environment is usually also spatially autocorrelated (e.g., Manly 1986, Leduc et al. 1992, Legendre 1993). Careless regression (or correlation) of autocorrelated variables across an explicit or implicit spatial surface may, as a consequence, highlight spurious associations (Lennon 2000). Thus, models of abundance-environment relationships that ignore autocorrelated spatial pattern may place undue emphasis on environmental factors that in truth have little or no bearing on a species' distribution and abundance. Perhaps more importantly, models that ignore spatial autocorrelation may fail to place sufficient emphasis on true abundance-environment relationships, and thus lead to omission of important variables during model selection.

A variety of methods have been applied to quantify species-habitat associations. These methods fall generally into two classes: regression methods and multivariate ordination techniques. Regression methods include simple linear regression on log or square root transformed count data, Poisson regression on raw count data and logistic regression on presence-absence data (e.g., Guisan et al. 1999). Additive regression models (Hastie and Tibshirani 1987) have also been used to model complex, non-parametric relationships between abundance and environmental predictors (Yee and Mitchell 1991). Multivariate techniques include canonical correlation (Gittens 1980) and Canonical Correspondence Analysis (Ter Braak 1987, Palmer 1993). These models seek linear combinations of environmental factors that are correlated to linear combinations of species' abundances. Both ordination and regression methods can be used to map abundance and distribution in space (see Guisan and Zimmermann 2000 for a comprehensive review with additional methods for distribution mapping). Whereas these techniques represent significant progress in modeling environment-abundance associations, in their standard form they fail to take into account the possibility that either the environment or the abundance may be spatially autocorrelated, and therefore these models may fail to provide an accurate representation of the environmental factors governing abundance.

Fortunately, there are a growing number of statistical methods that can incorporate spatial autocorrelation into environment-abundance models. Many of these techniques have received exhaustive review and technical discussion elsewhere (e.g., Haining 1990, Cressie 1993, Griffith and Layne 1999) and it is not our aim to repeat these general efforts. Rather we wish to focus the discussion on models for spatial autocorrelation in abundance and how they might improve modeling environment-abundance relationships. In particular, we emphasize the effect of spatial autocorrelation on

parameter estimates, and the interpretation of parameter values for understanding population level processes, as opposed to the effect of autocorrelation on significance tests, which have typically received more attention in the literature (see also other papers, this issue).

We illustrate the various statistical methods using three case studies: 1) habitat selection in bank voles *Clethrionomys glareolus*, 2) an experimental study of competition between a C<sub>3</sub> and C<sub>4</sub> grass, and 3) habitat associations for three species of Oribatid mites. Finally, we review possible extensions of our models to include scenarios more complex than the relatively simple regression models presented in the case studies.

## Methods

Since we are primarily interested in parameter estimation, we restrict our discussion to a specific set of models for which interpretation of estimated parameters is relatively straightforward, specifically linear regression models for log-abundance data. Within the class of linear log-abundance models we focus our discussion on four classes of models (Table 1). These are 1) standard regression models that ignore spatial dependence, 2) spatial autoregression models, 3) geostatistical correlation models, and 4) random block models for designed experiments.

The independence model is the familiar regression model whose errors are distributed independently and identically across all observations. The independence model is of course the standard choice when predictor or response variables are not spatially patterned. The independence model is usually a good place to start when first exploring relationships. If there are strong spatial signatures in the data, then one should assess the degree of spatial dependence in the residuals. Generally, plotting a correlogram of the residuals is sufficient to detect important patterns.

When violations of independence do occur, alternative models that account for dependence in the residuals should be used. This is true despite the fact that ordinary least squares estimates of the independence model parameters remain asymptotically unbiased in the presence of spatial autocorrelation. In any real situation, however, sample sizes will be finite; hence, appeals to asymptotic behavior are not terribly useful. Alternative models that explicitly account for spatial autocorrelation can reduce bias, as well as increase precision, relative to the independence model with sample sizes typical of ecological studies (see Krämer and Donninger 1987). An added dividend is the explicit characterization of spatial patterns, opening the opportunity for ecological interpretation.

Alternative models include autoregressive models (AR, CAR, SAR; see Table 1), which augment the

standard linear model with an additional term that accounts for patterns in abundance that are not predicted by local habitat variables, but are instead related to abundance in neighboring locations (or, more often, the “residual” in neighboring locations). In the AR model, the autoregressive term is independent of the environmental predictors, and is thus most appropriately applied when correlations in abundance result from endogenous population processes that are unrelated to environmental conditions, such as conspecific attraction or spatially contagious population growth.

The CAR and SAR models are similar to the AR model, except that abundances within the specified neighborhoods are deviations from the expected abundance given the habitat conditions (viz. the “residual variation”) at each location. The CAR model only considers first-order neighborhood effects, whereas the SAR model allows for recursive, higher order neighborhood effects. Both CAR and SAR are appropriate when residual abundance patterns depend strongly on latent environmental variables, or in situations where individuals entertain contiguous home ranges upon completing their habitat selection.

Geostatistical models (EA, GA) account for spatial pattern by modeling the correlation between errors directly as a function of distance. (We consider only two functions, exponential and Gaussian, but others may be appropriate as well.) The choice of geostatistical error-model can be based upon either biology or the nature of your data. The geostatistical models mainly have an advantage in the flexibility with which one can specify error structure, and may be more appropriate when observed patterns do not correspond to clearly defined neighborhoods.

Although not typically used for modeling environment-abundance relationships, the Random Blocks model is conceptually similar to the geostatistical models. In the Random Blocks model, two sets of errors, within-block and between-block, are computed. Between-block errors are assumed to be spatially independent, whereas within-block errors have a single pooled covariance. This model corresponds to an experimental design where treatments are randomized within larger sampling blocks.

As presented, these models assume linear responses to environmental gradients. In many cases, a non-linear, modal response may be more appropriate. In such cases, a regression of log-abundance against a second-order polynomial (modeling a Gaussian response in the raw abundances) could be used. (In the examples below, we found that the environment-abundance relationships were approximately linear, indicating that the extents of the gradients within the study area may have been insufficient to capture the entire species’ response to these variables, i.e., our models were fit over ranges of environmental variables where abundance was only either increasing or decreasing. We therefore only modeled linear responses.)

Furthermore, in deference to simplicity and illustration, all the models we consider here also assume Gaussian error distributions. However, alternative models are available that allow for other error distributions. The autologistic model (e.g., Augustin et al. 1995), for example, can be used to model presence-absence data. A more general approach estimates the parameters of a Generalized Linear Model (McCullagh and Nelder 1989) with correlated errors using the method of “Generalized Estimating Equations”. Recent studies have used this approach with binary (Albert and

Table 1. Summary of the statistical models linking the vector of log-abundance,  $Y$ , to a matrix of covariates  $X$ , and the resultant structure of dependence in the errors ( $\epsilon$ ).  $\beta$  is the vector of coefficients (to be estimated) that describes the strength of the association between the abundance and the various habitat variables.  $W$  and  $C$  are matrices that flags neighboring observations;  $d$  is the spatial distance between the observations;  $u$  represents a dummy-variable, flagging which observation belongs to which block;  $I$  represents the identity matrix;  $\rho$  is the measure of spatial dependence.

| Name   | Model  | Error covariance matrix   |
|--|--|---|
| Independent Errors<br>(no spatial dependence)                | $Y = X\beta + \epsilon$  | $\sigma^2 I$  |
| Autoregressive (AR)  | $Y = X\beta + \rho WY + \epsilon$  | $\sigma^2 [(I - \rho W)^T (I - \rho W)]^{-1}$ where $W$ is a (possibly) asymmetric neighbor connection matrix.  |
| Simultaneous Autoregressive (SAR)                            | $Y = X\beta + \rho W(Y - X\beta) + \epsilon$                                     | $\sigma^2 [(I - \rho W)^T (I - \rho W)]^{-1}$ where $W$ is a (possibly) asymmetric neighbor connection matrix.  |
| Conditional Autoregressive (CAR)                             | $Y = X\beta + \rho C(Y - X\beta) + \epsilon$                                     | $\sigma^2 (I - \rho C)^{-1}$ where $C$ is a symmetric neighbor connection matrix.   |
| Exponential Autocorrelation (EA)                             | $Y = X\beta + \epsilon$  | $\sigma^2 \Sigma(d)$ where $\Sigma_{ij}(d_{ij}) = e^{-d_{ij}/\rho}$ and $d_{ij}$ is the distance between $i$ and $j$ .  |
| Gaussian Autocorrelation (GA)                                | $Y = X\beta + \epsilon$  | $\sigma^2 \Sigma(d)$ where $\Sigma_{ij}(d_{ij}) = e^{-(d_{ij}/\rho)^2}$ and $d_{ij}$ is the distance between $i$ and $j$ .  |
| Random Blocks (for $i$ th observation<br>in $k(i)$ th block) | $Y = X\beta + u + v$<br>$\text{Var}(u) = \sigma_b^2, \text{Var}(v) = \sigma_c^2$ | $\Sigma(\sigma_b^2 + \sigma_c^2)$ where $\Sigma_{ij} = 1$ if $i = j$ , $\Sigma_{ij} = \sigma_b^2 / (\sigma_b^2 + \sigma_c^2)$ if $i$ and $j$ are in the same block, and $\Sigma_{ij} = 0$ if $i$ and $j$ are in different blocks. |

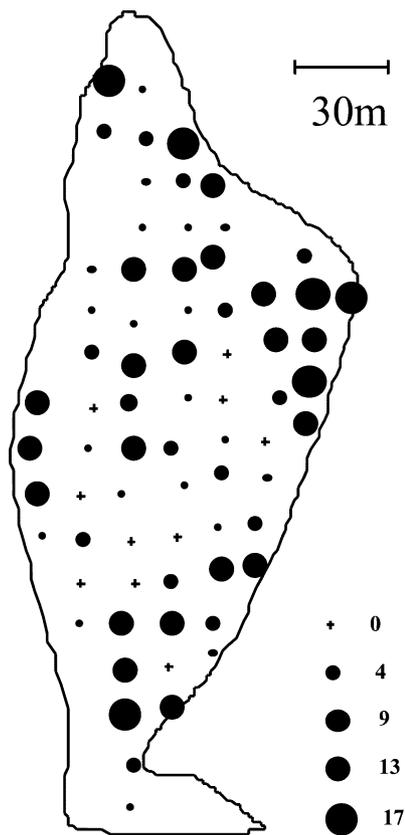


Fig. 1. The spatial map of total captures of female bank voles on the trapping grid. The minimum number of captures is zero, and the maximum is 17. The size of the symbols is proportional to the number captured.

McShane 1995), multinomial (Fahrmeir and Pritscher 1996) and Poisson data (Yasui and Lele 1997).

## Results

We present three case studies to illustrate the methodology; two that deal directly with abundance-environment relationships, and a third that examines abundance-environment relationships indirectly using data from a competition experiment.

### Example 1: spatial autocorrelation in bank vole habitat selection

Bjørnstad and Ims (unpubl.) conducted a study to examine the influence of environmental factors on bank vole abundance in central Norway (see Ims 1987 for a description of the study site). The data record the number of female bank voles captured in each of 71 pitfall live-traps. The traps were arranged in a rectangular grid with ca 15 m between each trap (Fig. 1). Vegetation and habitat variables were measured in four one-square-meter quadrats surrounding each trap. The average of these four were used for the analyses. The capture data represent eight experimental bank vole introductions between 1991 and 1994. Because female bank voles are central place foragers (Bondrup-Nielsen and Ims 1986), and have home ranges of sufficient size to encompass multiple traps (typical home range area based on data collected within the study-area is  $45 \times 45$  m; Bjørnstad and Ims unpubl.) we suspected the abundance data to be spatially autocorrelated. This is confirmed by plotting correlograms of abundance (Fig. 2a) and habitat variables (Fig. 2b, c). The spatial correlation in rodent abundance is initially high and drops off to zero around 35 m.

We used linear regression to relate log-abundance ( $\log \# \text{ captures per trap} + 1$ ) to characteristics of the local environment. After a preliminary screening, we retained five independent variables in the model: 1) vegetation height – providing shelter from predators; 2) *Cladonia* – cover of a lichen which indicates dry habitat; 3) *Empetrum* – cover of crowberry; 4) moss – cover of the dry-habitat *Pleurozium*; 5) ground structure – heterogeneity in soil and litter. *Cladonia* and “ground structure” exhibits short-range spatial dependence comparable to that seen in the rodent counts (Fig. 2b). The other environmental variables are less clearly spatially structured.

Conclusions about the importance of each environmental variable depend on the assumption of independent errors (Table 2). If spatial correlation is ignored, bank vole abundance is concluded to increase with increased vegetation height, decrease *Cladonia* cover, and decreased moss cover. The partial regression coefficients for *Empetrum* and ground structure were not

Table 2. The analysis of habitat selection in the bank vole. The parameter estimates and their standard errors are estimated assuming independence, or a variety of spatially-correlated error models. See main text for details.

| Variable              | Independence         | AR                   | SAR                  | CAR                  | EA                   |
|-----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Intercept             | $0.32 \pm 0.73$      | $-0.46 \pm 0.71$     | $0.22 \pm 0.71$      | $0.06 \pm 0.71$      | $-0.06 \pm 0.71$     |
| Vegetation height (1) | $0.06 \pm 0.02^{**}$ | $0.06 \pm 0.02^{**}$ | $0.05 \pm 0.02^{**}$ | $0.05 \pm 0.02^{**}$ | $0.06 \pm 0.02^{**}$ |
| <i>Cladonia</i> (2)   | $-0.48 \pm 0.24^*$   | $-0.44 \pm 0.23^*$   | $-0.37 \pm 0.23$     | $-0.39 \pm 0.23$     | $-0.34 \pm 0.23$     |
| <i>Empetrum</i> (3)   | $0.29 \pm 0.17$      | $0.28 \pm 0.16$      | $0.14 \pm 0.17$      | $0.16 \pm 0.17$      | $0.14 \pm 0.16$      |
| Moss (4)              | $-0.31 \pm 0.14^*$   | $-0.32 \pm 0.14^*$   | $-0.24 \pm 0.14$     | $-0.25 \pm 0.14$     | $-0.28 \pm 0.14^*$   |
| Ground structure (5)  | $0.12 \pm 0.09$      | $0.15 \pm 0.09$      | $0.19 \pm 0.09^*$    | $0.18 \pm 0.09^*$    | $0.19 \pm 0.09^*$    |
| AIC                   | 269.38               | NA§                  | 267.7                | 268.09               | 266.4                |

\*  $p \leq 0.05$ , \*\*  $p \leq 0.01$ . § We omit the AIC for this model because the likelihood was scaled differently in this case.

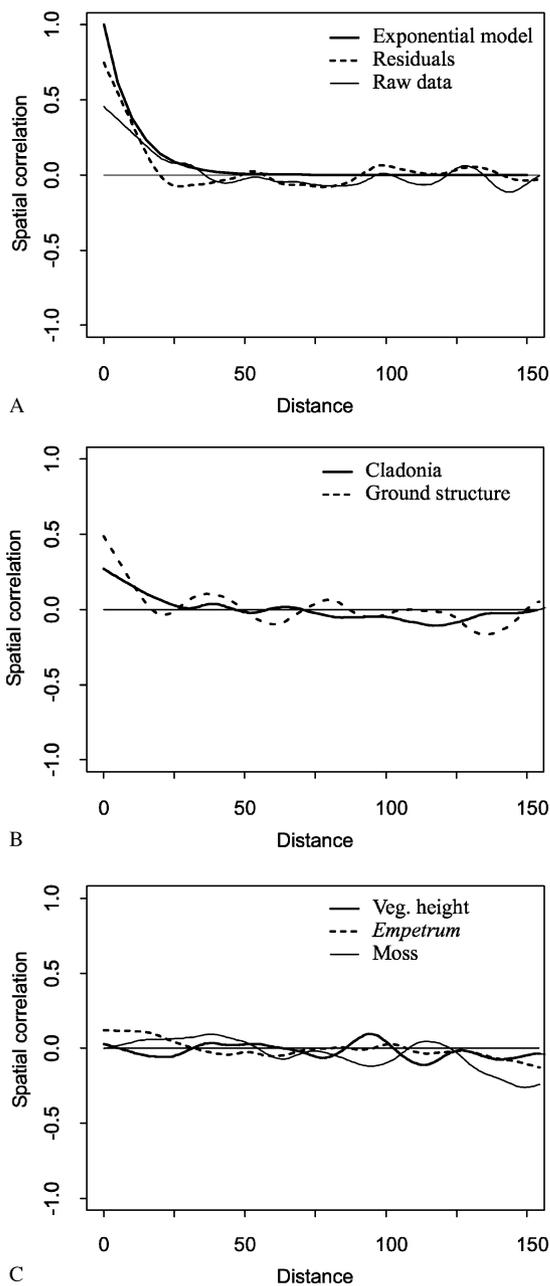


Fig. 2. The spatial correlation functions for the vole data. A) The captures and the residuals from the independence model. The exponential model is superimposed on the observed. B) and C) Spatial correlation in the habitat variables. The spatial correlation functions are estimated using spline correlograms with 15 degrees of freedom (Bjørnstad and Falck 2001).

significantly different from 0 (Table 2). However, the residuals from the independence model are strongly correlated (Fig. 2a). The model that includes spatial autocorrelation indicates a range of dependence up to ca 30 m (Fig. 2a). This corresponds roughly to the radius of female bank vole home ranges in this habitat. The estimated correlation between neighboring traps

will under this dependence range from 0.16 to 0.47 (depending on the exact distance between pairs of traps).

The results from all the spatial dependence models (with the exception of the AR model) are roughly the same. Compared to the independent errors model, explicit consideration of spatial correlations reduced the importance of two variables (2, 3), and increased the importance of another (5). Notably, the last variable, which is insignificant at the 10% level when assuming independence, is really one of the more important covariates as judged from the model accounting for the spatial autocorrelation. This is not really counter-intuitive: autocorrelation is like reading some of the data several times. The fact that estimates and levels of significance are altered when you effectively duplicate a subset of your observations stands to reason. (As an analogy, consider tossing two unbiased coins that give – according to expectation – one tail and one head. If the “tail” is inadvertently recorded twice, and the “head” is accurately recorded only once, the estimated odds are shifted by > 30% relative to the true odds.) An important take-home message from Table 2 is that the main difference is between models assuming independence and other models. Once some concession is made for interdependence, the conclusions appear to be relatively insensitive to the details of the spatial model.

### Example 2: spatial autocorrelation in a plant competition experiment

Spatial autocorrelation can influence both experimental and observational studies. To illustrate how the interpretation of spatial autocorrelation differs between experimental and observational studies, we analyzed data from a removal experiment involving the C<sub>4</sub> grass *Aristida glauca* and a competitor, the C<sub>3</sub> grass *Stipa neomexicana* (see Fortin and Gurevitch 1993). We analyzed a subset of the data comparing growth of *Stipa* in *Aristida*-removal plots to the growth in control plots. The experiment used a randomized block design involving 10 blocks. Each block was divided into eight units, of which two were randomly assigned as controls and two were assigned as *Aristida*-removal. The other 4 units were used for other treatments not considered here. The design is depicted and detailed in Fortin and Gurevitch (1993). In our analysis we investigated the plant growth rates  $\log(\text{final size}/\text{initial size})$  as a function of treatment, where size is measured by basal area in cm<sup>2</sup>. We used the data in Fortin and Gurevitch (1993): Table 15.1, except that we have deleted the outlier in Control 2 Block 5, and assigned a corrected coordinate ( $x = 23$  m,  $y = 4$  m) to Control 1 Block 6.

We use the data to illustrate how randomized field experimental data can be analyzed using independence models, spatial correlation models, or a more classical

Table 3. The experimental test of competition on *Stipa* growth in the presence vs the absence of its  $C_4$  competitor. The parameter estimates and their standard errors are estimated assuming independence, using a spatially-correlated error model, and a random block design. The optimal spatial error model uses a Gaussian structure with estimated range around 2 m.

| Effect size   | Independence | T    | p      |
|---------------|--------------|------|--------|
| Independence  | 1.05 ± 0.18  | 5.87 | <0.01  |
| Spatial error | 1.03 ± 0.13  | 7.41 | <0.01¶ |
| Block         | 1.06 ± 0.17  | 6.31 | <0.01† |

¶ Gaussian correlation (no nugget) with range: 2.01; LRT:  $\chi^2(1) = 3.58$ ,  $p = 0.06$ .

† Random block design; LRT:  $\chi^2(1) = 0.8$ ,  $p = 0.37$ .

randomized block model. As a contrast to observational studies, such as that of the rodent, the mean difference between the removal and control plots is estimated with good precision regardless of choice of correlation (or independence) model (Table 3). This robustness is, of course, at the heart of the randomization technique for field experiments, which ensures that we need only be concerned with precision, rather than dealing with block-induced biases. The geostatistical regression model does, however, reveal a strong but rather localized spatial correlation in the residuals (Fig. 3) (the range in the Gaussian correlation model is 2.01 m). As expected, explicit modeling of spatial autocorrelation results in enhanced statistical “efficiency” – we observe a 25% reduction in standard error of the estimated treatment effect using a Gaussian geostatistical model for the spatial dependence in the residuals (“spatial error” model, Table 3). The analysis using a random block design offers a surprising contrast to this as there is no apparent gain through the blocking. (Several free parameters are used, but only marginal increase in efficiency is attained.) This is despite the fact that “blocking” is the classical solution to the problem of latent variables. However, the  $8 \times 2$  m a priori selected blocks are several times larger than the distance over which observations are correlated. The blocks may

hence be seen as using up “degrees-of-freedom” (by fitting the block effects), but at the wrong scale relative to the inherent dependence in the system.

### Example 3: habitat associations in Orbatid mites

In the two examples introduced above, the spatial regression models work very well and produce efficient analyses and ecologically meaningful conclusions. In order to illustrate an important potential for confounding, we conducted an analysis of the abundance-environment relations in three mite species. The data represents the abundance of Orbatid mites in a *Sphagnum* bog at Lake Geai in Quebec (Borcard et al. 1992, Borcard and Legendre 1994). Adult numbers of 35 species of Orbatid mites were counted in 70 samples across a  $10 \times 2.6$  m transect from the floating vegetation mat to the forest edge. The dominant gradient is thus associated with humidity. A suite of environmental variables was measured at each sampling point. The relationship between the environment and the whole mite community has been analyzed and discussed in detail by Borcard and co-workers (Borcard et al. 1992, Borcard and Legendre 1994, see also other papers in this feature). We focus on the “individualistic” response of three of the more common mites: ONOV (*Oppiella nova*), a very common habitat generalist; LCIL (*Limnozetes ciliatus*) a hygrophilic species which can show very large populations under very wet conditions; and TVEL (*Tectocepheus velatus*) a “panphytophagous” habitat generalist (Borcard pers. comm.). In our analysis we include all environmental variables of Borcard and Legendre (1994), but emphasize the result for two continuous variables: humidity and substrate density.

In general, the analyses elucidate a strong association (either positive or negative) with the main humidity gradient (Table 4). The habitat generalist ONOV (and TVEL) is negatively correlated with humidity; the hygrophilic species (LCIL) is positively correlated with

Table 4. The analysis of environmental association of ONOV, LCIV and TVEL. The parameter estimates and their standard errors as estimated assuming independence, and spatially-correlated error model. The test for spatial error models are summarized below the table (including the Likelihood-ratio tests for spatial dependence). Note that the two “independence” models for TVEL are the same.

| Variable                        | Independence | T     | p     | Geostatistical | T     | p     |
|---------------------------------|--------------|-------|-------|----------------|-------|-------|
| ONOV: humidity ( $\times 100$ ) | -0.39 ± 0.08 | -4.65 | <0.01 | -0.39 ± 0.08*  | -4.65 | <0.01 |
| substrate ( $\times 100$ )      | 0.29 ± 0.09  | 0.34  | 0.74  | 0.29 ± 0.09*   | 0.34  | 0.74  |
| LCIL: humidity ( $\times 100$ ) | 0.75 ± 0.15  | 5.07  | <0.01 | 0.71 ± 0.01§   | 3.87  | <0.01 |
| substrate ( $\times 100$ )      | -1.88 ± 1.54 | -1.22 | 0.23  | -2.81 ± 1.75§  | -1.61 | 0.11  |
| TVEL: humidity ( $\times 100$ ) | -0.48 ± 0.08 | -5.44 | <0.01 | -0.08 ± 0.01¶  | -0.95 | 0.35  |
| substrate ( $\times 100$ )      | 4.89 ± 0.91  | 5.33  | <0.01 | 1.29 ± 0.78¶   | 1.65  | 0.10  |
| TVEL: humidity ( $\times 100$ ) | -0.48 ± 0.08 | -5.44 | <0.01 | -0.13 ± 0.08†  | -1.68 | 0.10  |
| Substrate ( $\times 100$ )      | 4.89 ± 0.91  | 5.33  | <0.01 | 1.70 ± 0.66†   | 2.57  | 0.01  |

\* Exponential correlation (no nugget) with range: 4.27E-18; LRT:  $\chi^2(1) = 0.14$ ,  $p = 0.71$ .

§ Gaussian correlation (nugget = 1.20) with range: 2.35; LRT:  $\chi^2(2) = 8.03$ ,  $p = 0.02$ .

¶ Exponential correlation (no nugget) with range: 6.08; LRT:  $\chi^2(1) = 37.81$ ,  $p < 0.01$ .

† Gaussian correlation (nugget = 0.91) with range: 0.81; LRT:  $\chi^2(2) = 30.51$ ,  $p < 0.01$ .

humidity. However, despite there being significant spatial autocorrelation, no clear insights result from the inclusion of spatial dependence. Indeed the analysis of TVEL testifies either to strong environmental control and localized spatial dependence, or limited environmental control and wide-ranging spatial dependence, depending on the exact spatial model assumed (Table 4).

## Discussion

There are many environmental and ecological processes that can result in spatial autocorrelation in abundance (Legendre 1993). These can include slow, large-scale processes such as uplift and erosion that form the geophysical template in which ecological interaction occur, mesoscale processes such as fire, local climate gradients and forest gap dynamics (e.g., Pickett and White 1985), and fast, fine-scale interactions between individuals such as territoriality and conspecific attraction (e.g., Stamps 1988, Ray et al. 1991) and repulsion (e.g., Bollinger et al. 1993, Lambin 1994a, b) in animal populations. Distance-limited dispersal and movement also generate autocorrelations in abundance, and these can oftentimes interact with broad-scale landscape patterns (e.g., Willson 1992, Keitt and Johnson 1995). Some of the more interesting and esoteric (albeit largely theoretical) sources of spatial autocorrelation stem from species interactions coupled with dispersal. Models of predator-prey and host-parasitoid dynamics have been shown to produce a wide range of autocorrelated spatiotemporal dynamics, including traveling waves, periodic checkerboard oscillations and spatial chaos (e.g. Bascompte and Solé 1998).

In the first two examples given above, we find biologically meaningful spatial patterns that significantly effect our interpretation of the predictive models for habitat abundance relations. In both cases, we were able to model spatial patterns by incorporating models of spatial autocorrelation into our statistical framework. In the case of bank voles, spatial correlations in abundance reflected a number of factors. Bank voles are territorial, central-place foragers. Thus we expect that there will be autocorrelation at the scale of the territory width, and this indeed appears to be the case (Fig. 2). In addition, we expect vegetation structure to influence vole abundance. Bank voles also experience significant predation, this gives rise to a dual effect of vegetation height on abundance. First, increased food abundance in the more lush vegetation leads to greater abundance, and second, taller vegetation serves as refugia from predation. Initially, when fitting the independence model, it appeared that vegetation had a mixed effect on vole abundance. Both *Cladonia* and moss cover were negatively correlated with abundance, sug-

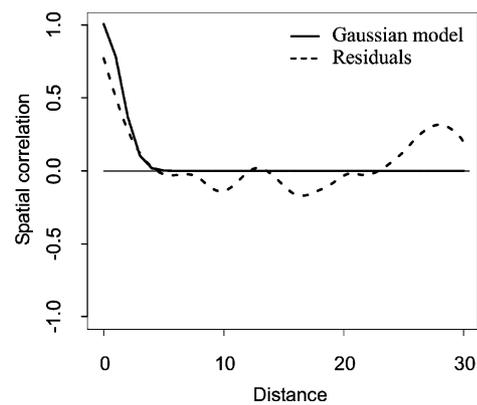


Fig. 3. The spatial correlation functions for the plant competition data. The dotted line represents the residuals from the independence model, the Gaussian model is superimposed. The spatial correlation function is estimated using a spline correlogram with 10 degrees of freedom (Bjørnstad and Falck 2001).

gesting that voles avoid dryer areas. However, when fitting the spatial autocorrelation models, neither *Cladonia* nor moss was significantly related to vole abundance (although we did observe a weak effect of moss in the EA model). Instead, it appears that ground spatial structure was more important, indicating that voles respond positively to heterogeneity in their environment.

Whereas the SAR, CAR and EA models showed significant differences in parameter estimates, the AR model showed little difference when compared to the independence model. This suggests a cautionary note. Selecting an appropriate model requires some thought about ecological sources of spatial dependence. The AR model would be a natural choice if the spatial correlations in abundance were independent of patterns in habitat variables – this could arise if offspring were highly philopatric creating a secondary aggregate of individuals. At the scale of this experiment on bank voles, however, this is clearly not the case since the dominant cause of autocorrelation is the spacing behavior of each individual (each home range encompasses multiple traps). Fortunately, all ecologically sensible spatial models result in identical conclusions.

In the case of the plant competition experiment, we found an intriguing example of a mismatch in scales of the experimental design and the extent of the underlying spatial dependence (Fig. 3). The original block design of the experiment missed the important spatial pattern, because the blocks were too large. A geostatistical model was able to detect significant patterns of spatial dependence in the residuals, but at scales considerably less than the original block size. From a biological standpoint, this indicates that growth rates probably varied according to a fine-scale spatial mosaic in moisture, competition with other plant species, and perhaps nutrient concentrations (or some combination of these).

Another possibility is herbivory which would tend to cause fine-scale patchiness. This type of mismatch between experimental design and underlying spatial dependence was in fact the catalyst for the earliest spatial regression model – the Papadakis method. Interested readers should consult Wilkinson et al. (1983).

Our analysis of Oribatid mites illustrates a final important inferential point; one that is made particularly clear by the two candidate spatial models for TVEL. In the narrow transect that parallels the dominant environmental gradient (the transect measures 10 m along the humidity gradient but only 2.5 m across it), the “effect of space” and the effect of the habitat drivers are strongly confounded. In this way it is almost impossible to distinguish between the scenario in which the spatial dependence in abundances is largely generated by dispersal and density-dependence, with little environmental control (TVEL model 1), versus the scenario in which spatial patterns in abundance largely reflect strong environmental control instead of endogenous population processes (TVEL model 2). This parallels the conclusion from the community analysis of Borcard and Legendre (1994). Their CCA variance partitioning indicated that of the 57 percent-units explicable, more than half (31 percent-units) could be explained either through spatial constraints or through environmental controls.

The potential confounding between internally generated low-frequency autocorrelation vs low-frequency environmental “forcing” (i.e., the controlling environmental factors have long-range spatial dependence) may be particularly acute in the presence of important “latent” variables. Ecological studies rarely include measures of all the important variables because we do not know all that ought to be measured. If the omitted (latent) variables are structured in space, the data themselves will be spatially autocorrelated.

When spatial autocorrelation is due to unobserved and unmodeled environmental variables, a SAR or CAR model may be the more appropriate choice of model (Griffith and Layne 1999). In such models, the value at a location depends on the residuals ( $Y_i - X_i B$ ) at nearby locations (Table 1). Although the CAR and SAR models share comparable model formulations, they represent slightly different biological mechanisms and describe different patterns of correlation between observations. As reviewed above, the CAR model is a first order spatial model where the value at a location depends on the values at other locations only if there is a non-zero connection between them, i.e. only if the appropriate value in the neighbor connection matrix,  $C$ , is non-zero. There are no spatially indirect effects. The SAR model, in contrast, is a second order spatial model that allows for propagation of effects to neighbors of neighbors, and so on. While the models are very similar, there are some subtle differences. When the data sets are large (e.g. hundreds or thousands of locations),

for instance, it is much more difficult to fit the SAR model than the CAR model. At the same time, because the SAR model includes indirect effects, it is often more appropriate when spatial dependence extends to broad scales.

The exponential and Gaussian models are geostatistical models that directly specify the correlation matrix,  $R$ , between the residuals. These models are thus empirical, not mechanistic, in their attempt to describe the spatial dependence. The exponential and Gaussian models are two of many possible correlation models. In the geostatistical models, the correlation between two locations are assumed to depend on the distance between them and not on some neighborhood connection matrix (as in the CAR or SAR). Even so, there are connections between geostatistical models and the autoregressive models. The CAR model with a “neighbor” connection matrix gives essentially the same correlation matrix as an exponential geostatistical model (Griffith and Csillag 1993). The SAR model leads to a correlation matrix that is similar to the Bessel variogram model (Griffith and Layne 1999). There is no geostatistical equivalent to the AR model.

The randomized block model assumes that the area is divided into smaller blocks, each with a different average. There are thus two sources of variability in the observations (and residuals), the variability between the block means,  $\sigma_b^2$ , and the variability between locations in a block,  $\sigma_c^2$ . In the random block model, both sources of variability are assumed to be mutually independent. The correlation between observations in the same block is  $\sigma_b^2/(\sigma_b^2 + \sigma_c^2)$ , because two observations in the same block share the same block mean. This model thus assumes that there is no other source of spatial autocorrelation. This may be appropriate if the blocks are well separated and scaled according to the dependence inherent in the underlying ecological process, but not necessarily otherwise – as is illustrated by the mite analysis.

It is important to remember that the geostatistical models describe the correlation between the model residuals. Overall, the different spatial regression models works well to correct spurious effects of autocorrelation as long as the “spatial scale of variation” (in some vague sense of the word) is not too similar between the dependent variable, the independent variable, and the “errors”. If they are similar, it is easy to get confounding between the environmental effect and the spatial effect as illustrated by the mite data where we could not distinguish between weak versus strong environmental controls on abundance.

Two decades ago, Hurlbert (1984) made an extremely important reminder about how randomization and manipulations provide the only “true path” to causal inference in ecology. Our review should in no way be read as parting with this overarching principle. However, ecology is a tricky science in that manipulation (at

the correct scale) may not always be possible. The good news – that we have reviewed and illustrated – is that recent methodological innovations provide a variety of possibilities for post-experimental (statistical) corrections to provide correct inference in the presence of autocorrelation. Our five rules of thumb are that 1) spatial models are easy to calculate with several of the most common statistical packages (we used both SAS and S-plus in our examples, however an increasingly large collection of statistical routines, including algorithms for fitting many of the models considered in this paper, is freely available in R; see <http://www.r-project.org/>); 2) results from spatially-structured models may point to conclusions radically different from those suggested by a spatially independent model; 3) not all spatial autocorrelation in abundances results from spatial population dynamics; it may also result from abundance associations with latent environmental variables; 4) the different spatial models do have different mechanistic interpretations in terms of ecological processes – thus ecological model selection should take primacy over statistical model selection; 5) the conclusions of the different spatial models are typically fairly similar – making any correction is more important than quibbling about which correction to make.

*Acknowledgements* – This work was conducted as part of the Spatial Statistics Working Group supported by the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant # DEB-0072909), the Univ. of California, and UC Santa Barbara. The authors wish to thank D. Borcard and P. Legendre for biological insights and data regarding Oribatid mites.

## References

- Abrahams, M. V. 1986. Patch choice under perceptual constraints: a cause for departure from an ideal free distribution. – *Behav. Ecol. Sociobiol.* 19: 409–415.
- Albert, P. S. and 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.
- Augustin, N. H., Muggleston, M. A. and Buckland, S. T. 1995. An autologistic model for the spatial distribution of wildlife. – *J. Appl. Ecol.* 33: 339–347.
- Bascompte, J. and Solé, R. V. (eds) 1998. *Modeling spatiotemporal dynamics in ecology*. – Springer.
- Bjørnstad, O. N. and Falck, W. 2001. Spatial covariance functions: estimation and testing. – *Environ. Ecol. Stat.* 8: 53–70.
- Bollinger, E. K., Harper, S. J. and Barrett, G. W. 1993. Inbreeding avoidance increases dispersal movements of the meadow vole. – *Ecology* 74: 1153–1156.
- Bondrup-Nielsen, S. and Ims, R. A. 1986. Reproduction and spacing behaviour of females in a peak density population of *Clethrionomys glareolus*. – *Holarct. Ecol.* 9: 109–112.
- Borcard, D. and Legendre, P. 1994. Environmental control and spatial structure in ecological communities, with an example on Oribatid mites (Acari, Orbatei). – *Environ. Ecol. Stat.* 1: 37–61.
- Borcard, D., Legendre, P. and Drapeau, P. 1992. Partialling out the spatial component of ecological variation. – *Ecology* 73: 1045–1055.
- Cressie, N. 1993. *Statistics for spatial data*. – Wiley.
- Fahrmeir, L. and Pritscher, L. 1996. Regression analysis of forest damage by marginal models for correlated ordinal responses. – *Environ. Ecol. Stat.* 3: 257–268.
- Fortin, M.-J. and Gurevitch, J. 1993. Mantel tests: spatial structures in field experiments. – In: Scheiner, S. M. and Gurevitch, J. (eds), *Design and analysis of ecological experiments*. Chapman and Hall, pp. 342–359.
- Gittens, R. 1980. *Canonical analysis: a review with applications in ecology*. – Springer.
- Griffith, D. A. and Csillag, F. 1993. Exploring relationships between semi-variogram and spatial autoregressive models. – *Pap. Reg. Sci.* 72: 283–296.
- Griffith, D. A. and Layne, L. J. 1999. *A casebook for spatial statistical analyses: a compilation of analyses of different thematic data sets*. – Oxford Univ. Press.
- Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models for ecology. – *Ecol. Model.* 135: 147–186.
- Guisan, A., Weiss, S. B. and Weiss, A. D. 1999. GLM versus CCA spatial modeling of plant species distribution. – *Plant Ecol.* 143: 107–122.
- Haining, R. 1990. *Spatial data analysis in the social and environmental sciences*. – Cambridge Univ. Press.
- Hastie, T. and Tibshirani, R. 1987. Generalized additive models: some applications. – *J. Am. Stat. Assoc.* 82: 371–386.
- Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. – *Ecol. Monogr.* 54: 187–211.
- Ims, R. A. 1987. Responses in spatial organization and behaviour to manipulations of the food resource in the vole *Clethrionomys rufocanus*. – *J. Anim. Ecol.* 56: 585–596.
- Keitt, T. H. and Johnson, A. R. 1995. Spatial heterogeneity and anomalous kinetics: emergent patterns in diffusion limited predator-prey interactions. – *J. Theor. Biol.* 172: 127–139.
- Krämer, W. and Donninger, C. 1987. Spatial autocorrelation among errors and the relative efficiency of OLS in the linear regression model. – *J. Am. Stat. Assoc.* 82: 577–579.
- Lambin, X. 1994a. Territory acquisition and social facilitation by litter-mate Townsend's voles (*Microtus townsendii*). – *Ethol. Ecol. Evol.* 6: 213–220.
- Lambin, X. 1994b. Natal philopatry, competition for resources, and inbreeding avoidance in Townsend's vole (*Microtus townsendii*). – *Ecology* 75: 224–235.
- Leduc, A. et al. 1992. Study of spatial components of forest cover using partial Mantel test and path analysis. – *J. Veg. Sci.* 3: 67–78.
- Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? – *Ecology* 74: 1659–1673.
- Lennon, J. J. 2000. Red-shifts and red herrings in geographical ecology. – *Ecography* 23: 101–113.
- Manly, B. F. J. 1986. Randomization and regression methods for testing for associations with geographical, environmental and biological distances between populations. – *Res. Popul. Ecol.* 28: 201–218.
- McCullagh, P. and Nelder, J. A. 1989. *Generalized linear models*. – Chapman and Hall.
- Morris, W. F., Wiser, S. D. and Klepetka, B. 1992. Causes and consequences of spatial aggregation in the phytophagous beetle *Altica tombacina*. – *J. Anim. Ecol.* 61: 49–58.
- Orians, G. H. and Pearson, N. E. 1979. On the theory of central place foraging. – In: Horn, D. F. (ed.), *Analysis of ecological systems*. Ohio State Univ. Press, pp. 155–177.
- Palmer, M. W. 1993. Putting things in even better order: the advantages of canonical correspondence analysis. – *Ecology* 74: 2215–2230.
- Pickett, S. T. A. and White, P. S. 1985. *The ecology of natural disturbance and patch dynamics*. – Academic Press.
- Ray, C., Gilpin, M. and Smith, A. T. 1991. The effect of conspecific attraction on metapopulation dynamics. – *Biol. J. Linn. Soc.* 42: 123–134.
- Stamps, J. A. 1988. Conspecific attraction and aggregation in territorial species. – *Am. Nat.* 131: 329–347.

- Ter Braak, C. J. 1987. CANOCO – a FORTRAN program for canonical community ordination by (partial) (detrended) canonical correspondence analysis, principal component analysis and redundancy analysis (ver. 2.1). – TNO Inst. Appl. Comp. Sci., Stat. Dept, Wageningen.
- Turchin, P. and Kareiva, P. 1989. Aggregation in *Aphis varians*: an effective strategy for reducing predation risk. – *Ecology* 70: 1008–1016.
- Turchin, P. and Thoeny, W. T. 1993. Quantifying dispersal of southern pine beetles with mark-recapture experiments and a diffusion model. – *Ecol. Appl.* 3: 187–198.
- Wilkinson, G. N. et al. 1983. Nearest neighbour (NN) analysis of field experiments (with discussion). – *J. R. Stat. Soc. B* 45: 151–211.
- Willson, M. F. 1992. The ecology of seed dispersal. – In: Fenner, M. (ed.), *The ecology of regeneration in plant communities*. CAB International, pp. 61–85.
- Yasui, Y. and Lele, S. 1997. A regression method for spatial disease rates: an estimating function approach. – *J. Am. Stat. Assoc.* 92: 21–32.
- Yee, T. W. and Mitchell, N. D. 1991. Generalized additive models in plant ecology. – *J. Veg. Sci.* 2: 587–602.