Spatial Heterogeneity and Characteristic Scales of Species–Habitat Relationships

ROBERT L. SCHOOLEY

Determining the characteristic spatial scales at which species respond most strongly to the amount of available habitat is crucial for developing cross-species scaling rules and for predicting species' responses to landscape modification. A Biologist's Toolbox article by Holland and colleagues ("Determining the Spatial Scale of Species' Response to Habitat," BioScience *54: 227–233) presents a multiscale approach and computer program (Focus) for detecting characteristic scales using a resampling procedure, species–habitat regression models, and nonoverlapping sampling sites. Holland and colleagues refer to these nonoverlapping areas as "spatially independent sites," when in fact spatial independence includes additional concerns not addressed by the Focus approach. Here I discuss issues of spatial heterogeneity—spatial autocorrelation and spatial dependence—as they relate to measuring the spatial scaling of organism–environment relationships with regression models. I present an empirical example with cactus bugs (*Chelinidea vittiger*), demonstrating how spatial heterogeneity complicates the task of determining characteristic scales of species–habitat relationships. Finally, I provide some cautions and suggestions for researchers who are considering using Focus to examine scaling patterns with existing data sets.*

Keywords: autocorrelation, landscape heterogeneity, patch context, spatial scale, spatial regression

Our perception of how species interact with their habitat depends on spatial scale, which can be characterized by extent and grain (Wiens 1989). Extent, the overall area covered by a study, sets the upper bound for generalizations. Grain, the size of sampling units, sets the lower limit for the scale of detectable patterns. The variability that researchers view in spatial data results from how the scale of measurement filters reality (Atkinson and Tate 2000).

Because ecological patterns and processes are scale dependent, an important undertaking is to identify particular spatial scales at which species respond most strongly to the amount and structure of habitat. These characteristic scales of response may differ among species and may be related to mobility and other natural-history traits (Wiens 1989, Levin 1992). Determining characteristic scales of species–habitat relationships should allow researchers to identify probable scales of critical processes, and to determine whether cross-species generalizations and scaling rules can be developed (Wiens 1989, Levin 1992) to facilitate predictions of species responses to landscape change. One approach for evaluating the spatial scaling of associations between organisms and their environment is to conduct repeated regressions while changing the scale of measurement (typically grain size). For instance, such procedures have been used to investigate the effects of grain size on the relationship between vegetation composition or biomass and environmental variables (Bian and Walsh 1993, Reed et al. 1993), between aphid densities and the percentage of arable land (Thies et al. 2005), between bird species richness and a suite of environmental correlates (van Rensburg et al. 2002), and between beetle abundance and forest cover (Holland et al. 2005).

In a Biologist's Toolbox article, Holland and colleagues (2004) presented a multiscale approach and an associated computer program (Focus) for detecting characteristic scales at which species respond to habitat amount. The Focus approach is applicable to studies in which the effects of patch context on species abundance are evaluated by measuring habitat amount around sampling sites with variable-sized circular buffers. A key aspect of the Focus approach is the use of species–habitat regressions with a resampling procedure in

Robert L. Schooley (e-mail: schooley@uiuc.edu) is an assistant professor in the Department of Natural Resources and Environmental Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801.

which only sites with nonoverlapping buffers are selected at each scale. Holland and colleagues (2004, 2005) consistently refer to these nonoverlapping measurement areas as "spatially independent sites"; however, nonoverlap of measurement areas does not ensure spatial independence, which includes a broader range of considerations than the one aspect that Holland and colleagues emphasized.

My objectives are to highlight issues involving spatial heterogeneity as they relate to determining characteristic scales of species–habitat relations, and to suggest that the Focus approach often will require some extensions to remedy the important problem of spatial nonindependence. My purpose is not to critique directly the research of Holland and colleagues (2004, 2005) on cerambycid beetles and forest structure. However, I am concerned that other researchers, especially graduate students, might apply the Focus approach blindly because of the availability of free software and the strong encouragement from Holland and colleagues (2004) to use Focus for mining existing data sets.

The Focus approach in brief

Holland and colleagues (2004) used simple linear regression to model the relationship between species abundance and habitat amount for a range of scales, and defined the "characteristic scale" as the scale with the strongest correlation coefficient (*r*). The Focus program tests for nonlinear relationships by comparing models with and without polynomial terms of the habitat predictor variable, but Holland and colleagues (2004, 2005) report results only for linear models. Species abundances (counts of beetles caught in funnel traps) were measured at multiple sample points, and habitat amount (forest cover) was measured at each sampling point within a circular area centered on the point. The spatial scale of measurement was altered by using circles with different radii (19 sizes between 20 and 2000 meters [m]) to measure habitat amount. Note that Holland and colleagues (2004) refer to the circle radius as the extent of the predictor variable, but in landscape ecology the circle radius normally would be termed the measurement grain of the study (Wiens 1989, Dungan et al. 2002, Wu et al. 2002). The problem underscored by Holland and colleagues (2004) is that the circular measurement areas can overlap, especially for larger radii, resulting in data points that may be spatially dependent. Nonindependence of data in regression analyses reduces the true degrees of freedom, makes significance tests too liberal, and inflates measures of correlation strength (Legendre et al. 2002, Lichstein et al. 2002, Haining 2003, Fortin and Dale 2005, Wagner and Fortin 2005). The Focus approach attempts to address this problem of spatial nonindependence by conducting regressions for each measurement scale using a random selection of sites in which the circular areas do not overlap. The selection and regression procedure is repeated a number of times for each scale, and the mean value of *r* is used to assess association strength.

Spatial autocorrelation, dependence, and characteristic scales

Spatial heterogeneity is the spatial structuring of a variable due to two main causes: spatial autocorrelation and spatial dependence (Legendre et al. 2002, Wagner and Fortin 2005). Spatial autocorrelation includes endogenous, biotic processes such as dispersal and conspecific attraction that produce patchiness in organism distributions (Wagner and Fortin 2005). When spatial autocorrelation is present, the value of a response variable depends on values of that variable at other surrounding sites (Legendre et al. 2002). Spatial dependence is an exogenous process caused by species responding to environmental conditions that are spatially structured by their own physical generating processes (Wagner and Fortin 2005). These two types of spatial heterogeneity are difficult to distinguish in practice (Lichstein et al. 2002, Wagner and Fortin 2005), and both can affect inferences about bivariate relationships (Legendre et al. 2002).

Data sets that might be used with the Focus approach could include spatial heterogeneity in species abundances, in habitat predictors, or both. The use of nonoverlapping measurement areas for the predictor variable should reduce the degree of autocorrelation, but geographic separation of sampled areas does not guarantee statistical independence (Lichstein et al. 2002). Some range of spatial structure probably exists in habitat variables such as forest cover, and nonoverlapping areas still can be related by shared structure. Environmental gradients in soil texture, soil moisture, topography, and aspect can generate spatial heterogeneity in composition and abundance of plant species at distances that exceed the separation distance of sampling areas. Moreover, the Focus approach as currently applied (Holland et al. 2004, 2005) does not explicitly consider positive autocorrelation of the response variable. Holland and colleagues (2004) note that constraints on site selection could be imposed so that sites need not be significantly autocorrelated to be included in regressions. This brief mention of autocorrelation understates a central issue regarding spatial independence of samples in landscape studies.

For correlation analysis that assesses the association between two response variables (Wagner and Fortin 2005), a modified *t* test for correlation coefficients is available that accounts for levels of autocorrelation in each variable (Dutilleul 1993). A modified test is necessary when there is autocorrelation in both variables (Lennon 2000, Legendre et al. 2002). The presence of broadscale environmental structure can reduce the power of the modified test, and good estimates of the spatial autocorrelation are required so that large sample sizes ($n \geq 100$) may be needed for strongly autocorrelated data (Legendre et al. 2002).

For models that analyze directional relationships, such as species abundance–habitat regression models, the assumption that must be tested is that model residuals are not autocorrelated (Lichstein et al. 2002, Haining 2003, Wagner and Fortin 2005). Spatial autocorrelation of residuals from regression models indicates autocorrelation of the response variable, or presence of an important, unmeasured environmental variable that is spatially structured (Keitt et al. 2002, Wagner and Fortin 2005). Haining (2003) emphasized the latter cause when he stated, "If unmeasured predictors are spatially correlated this property will be inherited by the errors" (p. 312).

Several approaches exist for analyzing data with positive autocorrelation via regression models, at least for certain response distributions, but the details of these methods are beyond the scope of this paper. In general, broadscale spatial trends initially can be removed by using linear or polynomial trend surfaces (Legendre et al. 2002, Lichstein et al. 2002). Finer-scale autocorrelation, as expressed by model residuals, can be accounted for by using autoregressive models, regression models with spatially lagged predictors, or models with spatially correlated errors (Keitt et al. 2002, Haining 2003, Fortin and Dale 2005). Ideally, a model should be based on assumptions about generating processes (Wagner and Fortin 2005).

Empirical example: Cactus bugs and habitat scaling

Here I illustrate how spatial heterogeneity can complicate the detection of characteristic scales of association between organisms and their environment. My data are from a study of cactus bugs (*Chelinidea vittiger*) and their host plant, the plains pricklypear (*Opuntia polyacantha*), conducted at Central Plains Experimental Range in Colorado during October 1999. Cactus bugs feed only on *O. polyacantha* at the site, and they reproduce within *O. polyacantha* patches, so definition of suitable habitat is straightforward. Adult cactus bugs normally walk between patches of host plants and have limited movement rates of no more than 2.5 m per day (Schooley and Wiens 2004). I present data for nonmating adults, which have the strongest association with *O. polyacantha* (hereafter "cactus") among the three life stages examined (Schooley and Wiens 2005).

I established a transect (700 X 1 m) and then sampled cactus and *C. vittiger* within 1400 contiguous quadrats. Each quadrat was 0.5 m^2 and represented my smallest grain of measurement. I partitioned each quadrat into eight subquadrats (0.25 X 0.25 m each) and then measured the frequency of cactus per quadrat by recording the number of subquadrats containing cactus (0 through 8). For quadrats with cactus, I searched for and counted the number of *C. vittiger.* I summed data on frequency of cactus and abundance of *C. vittiger* from the 0.5-m^2 quadrats to form data sets for eight larger grain sizes (1, 2, 4, 6, 8, 12, 16, and 24 m^2). Sampling frames without cactus after data aggregation were considered "structural zeros" and removed before analysis, so I evaluated the association between *C. vittiger* and cactus only where suitable habitat was present.

I developed a simple, nonspatial regression model (Holland et al. 2004) to analyze the relationship between abundance of *C. vittiger* and cactus frequency separately for each of the nine grain sizes. I used 95th-quantile regression (Cade et al. 1999) to model the association, because previous research indicated that *C. vittiger* and cactus exhibit a limiting-factor relationship (Schooley and Wiens 2005). All associations between *C. vittiger* abundance and cactus frequency were positive. I report R^2 for quantile regressions as a measure of association strength (McKean and Sievers 1987, Schooley and Wiens 2005). I evaluated residuals from regression models for spatial autocorrelation using a Moran's *I* coefficient (Fortin 1999, Haining 2003, Wagner and Fortin 2005) for the first lag distance (Qi and Wu 1996). Moran's *I* is a measure of autocorrelation that generally ranges from –1 (negative correlation) to 1 (positive correlation) and is near 0 ($E[I] = -[n - 1]^{-1}$) for spatially uncorrelated raw data (Qi and Wu 1996, Fortin 1999). Because residuals sum to zero, they have some level of correlation even if they are spatially independent (Lichstein et al. 2002). My purpose was to examine the relative autocorrelation strength of residuals across grain sizes, not to test for significance.

The amount of variation explained by regression models depended on grain size (figure 1). The degree of spatial autocorrelation in model residuals also changed substantially with grain size (figure 1). Most important, the pattern of autocorrelation intensity across grain sizes closely mirrored that observed for R^2 values. Similar results were obtained when trend surface terms were included in regression models. A naive interpretation of the simple species–habitat regressions might lead one to conclude that the characteristic scale for the association between cactus bugs and *Opuntia* habitat was 24 m^2, with a secondary peak at 12 m^2. However, this pattern partly reflects concurrent change in the intensity of spatial autocorrelation in model residuals and its ability to inflate R^2 values. The problem of spatially correlated errors

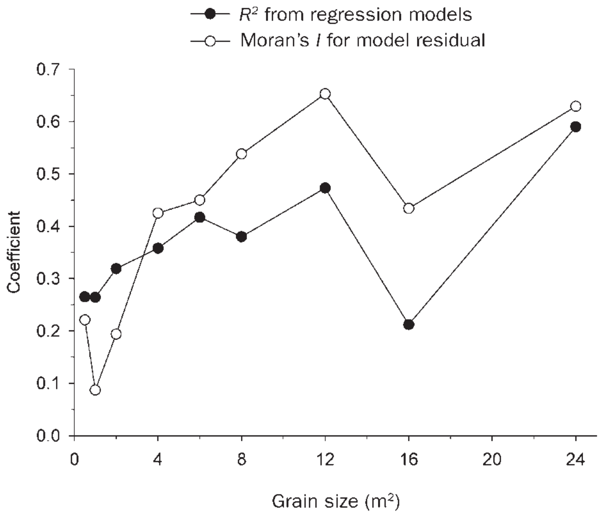

Figure 1. Effect of spatial scale (grain size of measurement) on correlation strength (R²) and autocorrelation of model residuals (Moran's I) for separate quantile regression models relating the abundance of cactus bugs to the frequency of habitat. After each regression model was fit, residuals were evaluated for strength of spatial autocorrelation at the first lag distance (neighboring quadrats).

depends on grain size. This outcome exemplifies one dilemma of data aggregation within what geographers call the "modifiable areal unit problem" (Dungan et al. 2002, Wu et al. 2002, Haining 2003). My results indicate spatial autocorrelation of cactus bug abundances due to movement behavior and conspecific attraction, or to the presence of an unmeasured environmental variable such as plant quality that operates on its own characteristic spatial scales. When spatial grain filters reality, it filters both the measured and the unmeasured components.

My approach and that of Holland and colleagues (2004) share many salient features: Both involve multigrain regressions between species abundance and habitat amount, conducted using potentially adjacent, nonoverlapping sampling areas. However, I changed the grain size of measurement for the predictor variable and for the response variable when altering spatial scale, whereas Holland and colleagues (2004) changed only the grain size of the predictor variable. Although simultaneously altering both response and predictor grain sizes might create additional problems of spatial heterogeneity, I think that the general issue applies to both designs. With Focus, the level of autocorrelation in the measured response variable could be altered by changing buffer sizes because of constraints imposed on the selection of sampling sites. The use of larger buffers forces the selection of sampling sites that are farther apart and less susceptible to strong positive autocorrelation in the response variable. Furthermore, altering buffer sizes with Focus can directly change the degree of spatial dependence of measured or unmeasured habitat variables. The resulting spatial structuring in the error terms of regression models should depend on grain sizes.

Additional issues

Counts of species abundance at sampling sites often include many counts of zero even when suitable habitat is present. Linear regression that assumes a normal distribution often is inappropriate for such count data. Generalized linear models such as zero-inflated Poisson regression or negative binomial regression might be better approximations for abundance data. However, autoregressive models for Poisson and negative binomial response distributions can model only negative spatial dependence (Haining 2003, but see Kaiser and Cressie 1997), which restricts their usefulness. Limiting-factor approaches using quantile regression could be effective for some data sets (Cade et al. 1999, Schooley and Wiens 2005), but quantile models for spatially correlated errors are not well developed. I concur with Holland and colleagues (2004) that researchers will need to "adapt the Focus program to suit their particular needs and applications." This advice pertains to statistical model formulation.

The Focus approach involves extensive data thinning, because only sample sites with nonoverlapping measurement areas are included in regressions. For instance, Holland and colleagues (2004) limited all regressions to a sample size of 16, because that was the number of nonoverlapping sample sites at their largest grain size. Although the data set was used more fully because of the resampling procedure, each individual regression was based on a relatively small sample size, even though 190 sites were sampled. Researchers should consider the trade-offs between such a heavy data reduction procedure and one that uses more of the data while explicitly modeling the spatial structure inherent in ecological patterns and processes (Keitt et al. 2002, Legendre et al. 2002, Lichstein et al. 2002, Haining 2003, Fortin and Dale 2005, Wagner and Fortin 2005).

Multigrain correlations of species abundances and habitat, like all correlative methods, describe patterns in nature. Insights about actual scaling processes are most likely to be gained from correlative studies in which a proposed set of plausible mechanisms is defined a priori on the basis of natural-history information. Scaling relationships often will involve multiple explanatory variables describing different aspects of habitat amount, composition, and configuration, with different predictors expressed most strongly within particular scale domains (Wiens 1989). Furthermore, changing the spatial extent of a study also can greatly affect how spatial heterogeneity is viewed (Wu et al. 2002). Such complications should be kept in mind by anyone contemplating a data-mining endeavor using the Focus program.

Conclusions

The Focus approach is a useful tool for selecting nonoverlapping sites, which is a problem for many landscape-level studies. However, multigrain regression is a tricky business, and nonoverlap of measurement areas does not ensure spatial independence; thus Focus should not be the only tool in a biologist's toolbox for multiscale analysis. Whether or not a data-thinning approach like Focus is used, researchers should evaluate residuals of species–habitat regression models for spatial autocorrelation. If residual autocorrelation is present, it should be dealt with before inferences are made about characteristic spatial scales of response.

Acknowledgments

I thank Brandon Bestelmeyer, Lyn Branch, and Helene Wagner for their valuable comments on the manuscript. The example of habitat scaling of cactus bugs was extracted from research funded by a fellowship from the Graduate School of Colorado State University to R. L. S. and by a US National Science Foundation grant to John Wiens. Brian Cade provided helpful advice regarding quantile regression.

References cited

Atkinson PM, Tate NJ. 2000. Spatial scale problems and geostatistical solutions: A review. Professional Geographer 52: 607–623.

Bian L, Walsh SJ. 1993. Scale dependencies of vegetation and topography in a mountainous environment of Montana. Professional Geographer 45: 1–11.

Cade BS, Terrell JW, Schroeder RL. 1999. Estimating effects of limiting factors with regression quantiles. Ecology 80: 311–323.

Dungan JL, Perry JN, Dale MRT, Legendre P, Citron-Pousty S, Fortin M-J, Jakomulska A, Miriti M, Rosenberg MS. 2002. A balanced view of scale in spatial statistical analysis. Ecography 25: 626–640.

Dutilleul P. 1993. Modifying the *t* test for assessing the correlation between two spatial processes. Biometrics 49: 305–314.

Fortin M-J. 1999. Effects of sampling unit resolution on the estimation of spatial autocorrelation. Ecoscience 6: 636–641.

Fortin M-J, Dale M. 2005. Spatial Analysis: A Guide for Ecologists. Cambridge (United Kingdom): Cambridge University Press.

Haining R. 2003. Spatial Data Analysis: Theory and Practice. Cambridge (United Kingdom): Cambridge University Press.

Holland JD, Bert DG, Fahrig L. 2004. Determining the spatial scale of species' response to habitat. BioScience 54: 227–233.

Holland JD, Fahrig L, Cappuccino N. 2005. Body size affects the spatial scale of habitat–beetle interactions. Oikos 110: 101–108.

Kaiser MS, Cressie N. 1997. Modeling Poisson variables with positive spatial dependence. Statistics and Probability Letters 35: 423–432.

Keitt TH, Bjørnstad ON, Dixon PM, Citron-Pousty S. 2002. Accounting for spatial pattern when modeling organism–environment interactions. Ecography 25: 616–625.

Legendre P, Dale MRT, 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.

Lennon JJ. 2000. Red-shifts and red herrings in geographical ecology. Ecography 23: 101–113.

Levin SA. 1992. The problem of pattern and scale in ecology. Ecology 73: 1943–1967.

Lichstein JW, Simons TR, Shriner SA, Franzreb KE. 2002. Spatial autocorrelation and autoregressive models in ecology. Ecological Monographs 72: 445–463.

McKean JW, Sievers GL. 1987. Coefficients of determination for least absolute deviation analysis. Statistics and Probability Letters 5: 49–54.

Qi Y, Wu J. 1996. Effects of changing spatial resolution on the results of landscape pattern analysis using spatial autocorrelation indices. Landscape Ecology 11: 39–49.

Reed RA, Peet RK, Palmer MW, White PS. 1993. Scale dependence of vegetation–environment correlations: A case study of a North Carolina piedmont woodland. Journal of Vegetation Science 4: 329–340.

Schooley RL, Wiens JA. 2004. Movements of cactus bugs: Patch transfers, matrix resistance, and edge permeability. Landscape Ecology 19: 801–810.

———. 2005. Spatial ecology of cactus bugs: Area constraints and patch connectivity. Ecology 86: 1627–1639.

Thies C, Roschewitz I, Tscharntke T. 2005. The landscape context of cereal aphid–parasitoid interactions. Proceedings of the Royal Society of London B: Biological Sciences 272: 203–210.

van Rensburg BJ, Chown SL, Gaston KJ. 2002. Species richness, environmental correlates, and spatial scale: A test using South African birds. American Naturalist 159: 566–577.

Wagner HH, Fortin M-J. 2005. Spatial analysis of landscapes: Concepts and statistics. Ecology 86: 1975–1987.

Wiens JA. 1989. Spatial scaling in ecology. Functional Ecology 3: 385–397.

Wu J, Shen W, Weizhong S, Tueller PT. 2002. Empirical patterns of the effects of changing scale on landscape metrics. Landscape Ecology 17: 761–782.