Determining the Spatial Scale of Species’ Response to Habitat

JEFFREY D. HOLLAND, DANIEL G. BERT, AND LENORE FAHRIG

Species respond to habitat at different spatial scales, yet many studies have considered this response only at relatively small scales. We developed a technique and accompanying software (Focus) that use a focal patch approach to select multiple sets of spatially independent sites. For each independent set, regressions are conducted between the habitat variable and counts of species abundance at different scales to determine the spatial scale at which species respond most strongly to an environmental or habitat variable of interest. We applied the technique to determine the spatial scales at which 12 different species of cerambycid beetles respond to forest cover. The beetles responded at different scales, from 20 to 2000 meters. We expect this technique and the accompanying software to be useful for a wide range of studies, including the analysis of existing data sets to answer questions related to the large-scale response of organisms to their environment.

Keywords: spatial scale, habitat interaction, forest cover, Cerambycidae, Focus

Ecological attributes such as population abundance and species richness depend not only on patch characteristics but also on the characteristics of the landscape surrounding the patch, known as patch context (Åberg et al. 1995, Gascon et al. 1999, Saab 1999, Szacki 1999, Fahrig 2001). Focal patch studies are one approach to studying the effects of patch context. In such studies, the data on species abundance or richness are collected in a number of patches or sites. The landscape predictor variables (e.g., habitat amount or fragmentation) are measured in areas that are centered on the patch or site locations (figure 1). Each patch (and its associated landscape) becomes a single data point in the data analysis (Brennan et al. 2002). In this way, researchers can examine the influence of habitat variables, measured at a large scale, on species abundance or richness. In this article, scale refers to the area or radius within which habitat predictor variables are measured. Therefore, we are referring only to the extent, and not the grain, of the predictor variable. Focal patch studies have been used to study the effects of road density and the amount of surrounding forest on wetland species richness (Findlay and Houllahan 1997), landscape habitat diversity on alfalfa insect richness (Jonsen and Fahrig 1997), amount of forest on raccoon density (Pedlar et al. 1997), amount of wooded border on alfalfa insect richness and density (Holland and Fahrig 2000), and amount of summer habitat and breeding pond density on leopard frog abundance (Pope et al. 2000).

In addition to these focal patch studies, there are probably hundreds of existing data sets in which researchers have studied the effects of local or patch habitat variables on population abundance or species richness in a number of patches. Can these data sets be reanalyzed to study the effects of landscape context on population abundance or species richness? Given the relative ease of obtaining remotely sensed habitat data, these data sets represent a mine of potential information on the effects of landscape context. The main problem with using these data sets, however, is that in such studies the patches or sites are often rather closely spaced. This can result in data points that are not spatially independent because the landscape areas overlap (figure 1), possibly leading to pseudoreplication. These overlapping sites may constrain the number of data points that can be used for examining the relationship between species abundance or species richness and the measures of landscape context. Nevertheless, because the collection of field data is time-consuming and expensive, it would be beneficial if there were a way to use this data to examine questions related to the larger-scale landscape context. In this article, we present a randomization method and computer program (Focus; www.carleton.ca/lands-ecol/) that permit analysis of effects of landscape context in this situation. Version 2 of this software is available as a free download to researchers and will remain so for the indefinite future.

Jeffrey D. Holland (e-mail: holland_jeff@yahoo.ca) is a postdoctoral research fellow in the Department of Biological Sciences, University of Windsor, Ontario N9B 3P4, Canada. Daniel G. Bert is a doctoral candidate, and Lenore Fahrig is a professor, in the College of Natural Sciences, Carleton University, Ottawa, Ontario K1S 5B6, Canada. © 2004 American Institute of Biological Sciences.
Multispecies, landscape-scale studies are often conducted at a single spatial scale for all the species studied (e.g., McGarigal and McComb 1995, Trzcinski et al. 1999, Holland and Fahrig 2000). However, it is likely that different species respond to their environments at different scales (e.g., Roland and Taylor 1997) and that these scales are related to the movement ranges of the organisms (Addicott et al. 1987, Wiens and Milne 1989, Wiens et al. 1993, Cale and Hobbs 1994, Vos et al. 2001, Dungan et al. 2002). Often little or nothing is known about the scales at which a species responds to structural characteristics of its environment, and this uncertainty may greatly limit the effectiveness of study designs. One method of estimating the appropriate scale is to model the relationship at a number of scales and determine which scale fits the model best (e.g., Findlay and Houlahan 1997, Elliott et al. 1998, Pope et al. 2000, Savignac et al. 2000). We applied this approach to test the hypothesis that different species of deadwood-boring beetles in the family Cerambycidae (Coleoptera) respond to their habitat at different scales. We refer to the scale at which different species respond most strongly to the amount of habitat as the characteristic scale of response to habitat amount. We use the word characteristic to imply that this scale may be an emergent property of a species, determined by the unique relationship between the species and its habitat (Mitchell et al. 2001).

The Focus program

Three matrices provide the basic inputs to the Focus computer program: a matrix of response variable measure (e.g., species abundance or richness) at each site, a matrix of between-site geometric distances, and a matrix of the predictor variable measurements at various increasing scales around the sites. At each spatial scale, Focus conducts multiple simple linear regressions of the ecological response on the landscape predictor. Each regression at each scale involves a different set of independent and randomly chosen data points.

The program selects spatially independent sites for the regressions in six steps: (1) It randomly selects a site \( n_i \) from the entire set of \( N \) sites. (2) It selects a second site \( n_j \) that satisfies the constraint of spatial independence, \( C \). For the case considered here, \( C \) is a measure of geometric distance where \( C > 2 \) (radius). In other words, areas within which the predictor variable is measured may not overlap (figure 1). (3) It continues to randomly select sites until the constraint \( C \) can no longer be met, or until a predetermined number of sites are selected. (4) It fits a regression line to the selected points. In this study, the abundance of each beetle species was regressed separately on the proportion of forest cover around the corresponding sampling location. The following regression statistics are recorded: \( R^2 \) (coefficient of determination), \( r \) (Pearson correlation coefficient), MSE (mean squared error), and \( N_p \) (number of points in each regression), among others. (5) It repeats steps 1 through 4 for \( X \) different sets of spatially independent sites to develop a distribution of regression model fit, effect size, or both. The sampling is done with replacement between sets. However, the constraint \( C \) makes it impossible to sample with replacement within a given set, since it precludes the same point being chosen more than once. The selection of points going into each set is therefore different from that of a bootstrap technique. (6) Finally, steps 1 through 5 are repeated for each spatial scale. The output of the analysis is the mean and standard error (or 95 percent confidence interval) of the regression statistics \( (R^2, r, MSE, N_p) \) at each scale. The scale of response can then be determined by examining a plot of model fit against increasing spatial scale.

In principle, the constraint \( C \) and the pattern of site selection (which in this case is random) could take many different forms. For example, instead of a geometric distance, the constraint \( C \) could be defined in such a way that sites must not be significantly spatially autocorrelated to be included in a final set for one of the regressions. The pattern of site selection could also follow a grid-based or stratified random design. We expect that researchers will adapt the program to suit their particular needs and applications.

The procedure has at least three advantages over random selection of a single set of independent sites: (1) Because it includes several regressions, site selection is not affected by the particular site that is selected first; (2) sites at different scales are not nested, because the subsets in any regression at a smaller scale may be quite different from the subsets used at larger scales; and (3) it maximizes the data available. Although the sample size of each individual regression may be much smaller at larger spatial scales, the randomization method allows for multiple estimates of the regression
(using different sets of data points), thus increasing the power of the analysis.

The Focus program also allows the user to test two assumptions that apply to the overall process of resampling the data to find the scale at which the model has the best fit: (1) functional stability at different scales and (2) representative sampling. A test for functional stability is necessary because Focus uses linear regression at all scales, whereas the relationship between the organism and the habitat variable may be nonlinear at some scales. Nonlinear relationships would cause the measures of model fit to decrease at these scales simply because a linear model is less able to represent the data at these scales, not because the actual relationship is weaker. The Focus program tests for such a nonlinear trend by using forward, stepwise multiple regression to attempt to fit second- and third-order polynomial terms (of the landscape predictor) to the model. The models are considered to be significantly improved by the addition of these higher-order terms if the fit is improved at the \( \alpha = 0.05 \) level. A test to ensure representative sampling is necessary because, if the number of regressions chosen is too small, the resulting distribution of regression statistics may not adequately represent the entire data set. If the variability in the data is not adequately represented by the sets of independent points that have been selected, then repeating the analysis could lead to very different measures of model fit reported at different scales. The representative sampling test verifies that the number of iterations used is sufficient by rerunning the analysis (including the data-point selection process) 10 times at the smallest scale and determining the mean and variance of the resulting \( \text{MSE} \) distribution as well as the mean and standard error of the regression coefficient. If these values are relatively constant across the 10 analyses, then the number of iterations used is sufficient and the results are representative of the data set. This assumption is tested at the smallest scale because, for many data sets, it is at this scale that the number of possible ways to sample the data is greatest. For the hypothetical set of points in figure 2, for example, at the smaller scale shown (figure 2b), there are 56 ways that five of the points could be chosen. At the larger scale, there are only 6 possible independent sets of points (figure 2b through 2g).

**Characteristic scales of wood-boring beetle response to habitat amount**

Most long-horned beetles have larvae that develop within the wood of living or dead trees or shrubs. Species with larvae that develop within dead wood are much less host-specific than species whose larvae develop within living or very recently dead trees (Linsley 1954), probably because the trees’ defensive chemicals are not as important in deadwood (Kletecka 1996). In a study that considered three different scales, Økland and colleagues (1996) found that long-horned beetles responded to environmental habitat variables within 1 square kilometer (km\(^2\)) of trapping sites. Schiegg (2000) found that saproxylic beetles (those that are dependent on deadwood) responded to the spatial connectedness of dead-wood pieces within 150 meters (m) in a study that considered spatial scales ranging from 50 to 200 m in 50-m increments.

We used the Focus program to determine the scales at which different species of deadwood-boring beetles respond to the amount of forest cover in the landscape. The beetles were sampled as part of a study on the response of wood-boring beetles to ice storm damage in the area surrounding Ottawa, Canada. Beetles were trapped across an area measuring approximately 80 km by 40 km, with the trapping locations clustered in 19 1-km\(^2\) areas with 10 sampling points each (figure 3). This means that the number of spatially independent points decreased as the radius of the area under consideration increased up to about 0.5 km (figure 4). At each trapping location, a Lindgren multiple-funnel trap was baited with ethanol and chipped wood. Traps were checked monthly, and all long-horned beetles were identified to species in the lab using Yanega’s field guide to Cerambycidae (1996). We used only species whose larvae develop in deadwood. We also used only species that do not appear to show a preference for any particular tree species (Linsley 1962, 1963, 1964, Linsley and Chemsak 1972, 1976, Yanega 1996) to ensure that the proportion of forest would be a fair approximation of the proportion of breeding habitat in the landscape. We refer to these species as *polyphagous*. The response variable for each species was the number of individuals caught at each site, summed over 2 years. We only used species that were caught in at least five of the sampling locations during the 2 years.

To detect the characteristic scale at which individual beetle species respond to forest habitat, we used Focus to estimate the fit of simple linear regressions of beetle abundance
on the proportion of forest in the landscapes around the beetle sampling sites. The proportion of forest was measured at multiple scales: radii of 20 to 200 m in 20-m increments, 400 to 2000 m in 200-m increments, and 3000 to 7000 m in 1000-m increments. We used digital topographical maps (National Capital Commission 1999) in ArcView to extract forest cover measures within the 24 different-sized circular areas near each site. At each scale, the regressions were repeated 200 times to generate the mean and standard error of the regression statistics. The choice of characteristic scale was made by examining a plot of a measure of model fit against the various scales at which forest proportion was measured.

We chose to conduct our analyses using the same number of data points for each regression at each scale, because the purpose of our analysis was to compare the results at different scales. This causes a trade-off between the maximum scale considered and the number of data points used in each regression. Including all scales up to 7 km would have limited the number of points in each regression to five. We found no indication in the literature that saproxylic beetles would respond to habitat at scales much larger than 1 km. However, none of the literature focused on the exact species we used in this study, so we decided to include scales up to 2 km. Using a maximum spatial scale of 2 km allowed us to increase the number of points in each regression to 16 (figure 4).

To evaluate the relationship between beetle abundance and proportion of forest cover, we used the correlation coefficient (r) as our measure of model fit, because for some species there was a range of scales over which this relationship was negative. The other possible measures of model fit (R², MSE) do not differentiate between positive and negative relationships. Using R² or MSE is also problematic as a measure of the strength of the regression, because we are averaging over many regressions. If the regressions at a given scale show both positive and negative relationships, as is likely in the case of a very weak relationship, the mean of measures such as R² and MSE will be artificially high. This occurs because the values of these statistics from individual regressions are always positive, even if some of the regressions result in negative relationships.

For each species, we examined the plot of r against spatial scale to determine the characteristic scale of response to forest cover. The scale corresponding to the best model fit was the characteristic scale of response to habitat amount.

**Results of spatial scale analysis**

We caught 13 species of polyphagous long-horned beetles associated with deadwood in at least five of the traps. Voucher specimens of all species were placed in the Carleton University entomological collection. The scales at which different species respond to forest cover varied from 20 to 2000 m (table 1). Figure 5 shows an example of the distribution of correlation coefficients from the regressions at different scales for one species, Urgleptes signatus. This figure shows that the distributions of the r values extended over both positive and negative values within a scale, justifying the use of r, rather than R², as a measure of model fit for these species.

**Hyperplatys aspersa** had r values that were negative throughout the range of scales considered, with the value at 180 m

<table>
<thead>
<tr>
<th>Species</th>
<th>Characteristic scale of response (radius in meters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bellamira scalaris (Say)</td>
<td>1000</td>
</tr>
<tr>
<td>Evodinus m. monticola (Rand.)</td>
<td>2000</td>
</tr>
<tr>
<td>Gaurotes cyanipennis (Say)</td>
<td>160</td>
</tr>
<tr>
<td>Liopinus alpha (Say)</td>
<td>20</td>
</tr>
<tr>
<td>Microgoes oculatus (LeC.)</td>
<td>60</td>
</tr>
<tr>
<td>Strictolepta c. canadensis (Oliv.)</td>
<td>1600</td>
</tr>
<tr>
<td>Strangalepta abbreviata (Germ.)</td>
<td>400</td>
</tr>
<tr>
<td>Strangalia luteicornis (F.)</td>
<td>1800</td>
</tr>
<tr>
<td>Trachysida mutabilis (Newm.)</td>
<td>1200</td>
</tr>
<tr>
<td>Trigonarthris minnesotana (Csy.)</td>
<td>200</td>
</tr>
<tr>
<td>Urgleptes signatus (LeC.)</td>
<td>180</td>
</tr>
<tr>
<td>Urographis fasciatus (DeG.)</td>
<td>180</td>
</tr>
</tbody>
</table>
approaching a zero correlation. This species did not appear to respond to forest cover in the same way as the other species and so was not included in figure 6 and table 1. Several other species had negative mean $r$ values across some range of scales, but the mean $r$ value at the scale of maximum model fit was positive for all species other than *H. aspersa*.

**Getting the spatial scale right**

It is important to use an appropriate spatial scale when considering how species respond to environmental variables. Figure 6 shows that the strength of the relationship between species abundance and proportion of forest cover varies at different spatial scales. Both the likelihood of researchers detecting this relationship and the importance ascribed to it will vary with the relationship's strength. Therefore, it is important to get the scale right when conducting such studies. In our study, different species of beetles from the family Cerambycidae responded to forest habitat at different spatial scales, showing that the scale appropriate for such research may vary even within a single family of beetles.

One interesting result is that some species showed more than one peak in model fit. It is possible that habitat variables are important at more than one scale for different reasons (Kotliar and Wiens 1990, Jonsen and Taylor 2000, Nathan 2001). For example, Kinnunen and colleagues (2001) and Rukke and Midtgård (1998) suggest that some habitat variable may cause beetles to select habitat at a fine scale, while larger-scale habitat availability may limit the areas within which a beetle species can occur.

The maximum correlation coefficient for *H. aspersa* was very close to zero, and the correlation coefficient values were negative across all scales. The abundance of this species may depend more on resources found outside the forest than on the forest cover itself. If this is the case, then the scale with the most negative peak may be characteristic of this relationship. The reason this species does not appear to respond positively to forest cover at all may be that it requires relatively open habitat. Many species of long-horned beetles are thermophilic, even in the larval stage (Barbalat 1998).

It is possible for the $r$ values of individual regressions to range over both positive and negative values within a scale (figure 5). This shows the importance of using all the data to determine the nature of the relationship between species abundance and the predictor variable, as the Focus program does. The constraint of spatial independence results in relatively few points being included in each regression. Relying on a single regression, using a single set of independent points, would not result in reproducible results. Using only a single regression could also result in incorrect conclusions about the direction of the relationship between abundance and forest cover. The repeated regressions on sets of inde-
Pendent points ensure that more information contained in the data is used to determine the relationship between abundance and forest cover.

Some of the beetle species were absent from the samples taken at many sites. This resulted in some regressions being left out of the analyses because all the data points selected for those regressions had no individuals. Therefore, the estimates of mean model fit for the less commonly caught species often had large standard errors, because the number of regressions used to estimate the fit was small. It would be possible to use Focus to resample the pool of data points for regressions that have this problem, so that the number of regressions would be constant at all scales, but this could lead to another problem: If there were relatively few points where individuals were found, it is likely that the same set of points would be regressed repeatedly using this alternative. This would result in a deceptively low standard error being reported. Both of these problems will be addressed in future versions of the software.

The sampling points in our example came from a large area, but because our study included the effects of habitat at scales of several kilometers, the number of spatially independent data points was small. However, with 190 sampling locations, we were able to use 200 regressions per scale and adequately represent the data without oversampling. With very small data sets, especially if the sampling points are close together, the issue of oversampling can be much more restrictive. When using the Focus program, researchers will have to decide on the number of regressions to use at each spatial scale. The actual number of iterations that can be used has lower and upper limits determined, respectively, by the assumption of representative sampling and by the criterion that no exact duplicate sets of points be used within a given scale.

The program we describe here should increase the use of existing data sets to answer larger-scale questions, as long as the locations where the response variables were measured are known. We did a survey of the literature to see how many studies had data appropriate for this application, using Cambridge Scientific Abstracts to search within the “Ecology Abstracts” subfile of “Biological Sciences.” Our search criteria were that the listings not contain the term model anywhere, but that they contain the term landscape or patch. This search yielded 6985 studies. We picked a random sample of 40 abstracts from these studies to check for the proportion of studies producing data sets that would lend themselves to larger-scale questions using the Focus program. To be considered appropriate, the studies had to produce numerical response data measured at multiple locations, and these locations had to be spread over an extent large enough that multiple independent data points would exist when the predictor variables were measured beyond the patch level. In other words, the sampling locations had to extend over more than one patch. Twenty of these 40 studies created data sets that met our criteria. This leads us to estimate that there are about 3500 data sets in the ecological literature that would lend themselves to answering large-scale questions of characteristic scale of response of organisms to their environment using the Focus program.

Acknowledgments
Scott Findlay provided invaluable statistical advice in the early stages of this project (however, any errors are our own). Jane Allison wrote the original Focus program (version 1.x) out of the goodness of her heart.
Lutz Tischendorf wrote the second version (version 2.x) with funding from Parks Canada. Many people in the Carleton University Landscape Ecology Lab provided helpful comments on an earlier draft, especially Naomi Cappuccino, Kathy Freemark, and Rebecca Tittler. Larissa Parriag and Moshi Kotierk provided cheerful help in the field, regardless of the conditions. The article was greatly improved by comments on an earlier draft by four anonymous reviewers. This work was funded by a National Science and Engineering Research Council (NSERC) grant to L. E., an NSERC scholarship to J. D. H., and an Ontario Graduate Scholarship to D. G. B.

References cited


