

HIERARCHICAL ANALYSIS OF SPECIES DISTRIBUTIONS AND ABUNDANCE ACROSS ENVIRONMENTAL GRADIENTS

JEFFREY M. DIEZ¹ AND H. RONALD PULLIAM

Institute of Ecology, University of Georgia, Athens, Georgia 30602 USA

Abstract. Abiotic and biotic processes operate at multiple spatial and temporal scales to shape many ecological processes, including species distributions and demography. Current debate about the relative roles of niche-based and stochastic processes in shaping species distributions and community composition reflects, in part, the challenge of understanding how these processes interact across scales. Traditional statistical models that ignore autocorrelation and spatial hierarchies can result in misidentification of important ecological covariates. Here, we demonstrate the utility of a hierarchical modeling framework for testing hypotheses about the importance of abiotic factors at different spatial scales and local spatial autocorrelation for shaping species distributions and abundances. For the two orchid species studied, understory light availability and soil moisture helped to explain patterns of presence and abundance at a microsite scale ($<4 \text{ m}^2$), while soil organic content was important at a population scale ($<400 \text{ m}^2$). The inclusion of spatial autocorrelation is shown to alter the magnitude and certainty of estimated relationships between abundance and abiotic variables, and we suggest that such analysis be used more often to explore the relationships between species life histories and distributions. The hierarchical modeling framework is shown to have great potential for elucidating ecological relationships involving abiotic and biotic processes simultaneously at multiple scales.

Key words: conditional autoregressive models; environmental gradients; *Goodyera pubescens*; hierarchical Bayesian framework; niche; *Orchidaceae*; spatial scale; species distributions; *Tipularia discolor*.

INTRODUCTION

Explaining patterns of species distributions has been a focus of ecological work for many years (Andrewartha and Birch 1954), but many important, unanswered questions remain about the relative roles of abiotic and biotic constraints and stochastic events. The interest in understanding how these processes interact spans theoretical and applied studies. Much current debate has focused on the relative importance of niche partitioning and stochastic processes in shaping community structure (Hubbell 2006, Leibold and McPeck 2006), and efforts to use niche theory and biogeographical concepts for conservation goals depend on proper quantification of abiotic drivers while accommodating stochastic effects (Whittaker et al. 2005). For both theoretical and applied efforts, an explicit consideration of spatial scale may prove critical for understanding the relative roles of these processes for shaping species distributions and community assembly (Pearson and Dawson 2003, Leibold et al. 2004, Holyoak and Loreau 2006).

Species-specific physiological constraints provide a logical first expectation for how species may be

distributed in relation to abiotic variables, such as precipitation and temperature, and form the basis for many ideas about the niche (Chase and Leibold 2003). However, there is wide recognition that numerous biotic processes, such as competition, predation, mutualisms, and disease, as well as stochastic dispersal and disturbance events, can keep species out of equilibrium with a physiologically optimal distribution. These complexities have led to the incorporation of stochasticity and nonequilibrium concepts into niche theory (Pulliam 2000, Tilman 2004).

Efforts to use observed climate–abundance relationships to build predictive models of how species may respond to future climatic scenarios (e.g., Skov and Svenning 2004, reviewed by Guisan and Thuiller 2005) have been criticized for failing to incorporate biotic interactions, dispersal limitation, and evolutionary change (Pearson and Dawson 2003, Hampe 2004). Although species occurrence and abundance can be limited measures for describing species dynamics, these are often the only feasible data to collect, necessitating approaches to inference that minimize the limitations. One improvement to these models has been the introduction of spatial autocorrelation, which can help to differentiate between responses to measurable abiotic gradients and other unmeasured (or unmeasurable) abiotic and biotic processes that generate spatial structure in species distributions (Gelfand et al. 2003, Hooten et al. 2003, Latimer et al. 2006). Even when the

Manuscript received 22 January 2007; revised 12 April 2007; accepted 23 April 2007; final version received 7 May 2007.
Corresponding Editor: J. Franklin.

¹ Present address: Centre for Advanced Bio-Protection Technologies, Lincoln University, Canterbury, New Zealand.
E-mail: diezj@lincoln.ac.nz

effects of measured covariates (e.g., temperature or precipitation) are of primary interest, including spatial structure in such models helps to improve prediction and reduce the risk of spurious correlations with abiotic variables due to spatial autocorrelation (Lichstein et al. 2002). The work thus far that incorporates spatial effects into distribution models has typically focused on large spatial scales on the order of kilometers, but such spatial dynamics may also be relevant to understanding species life histories, abiotic relationships, and stochastic events at smaller within-population scales.

Despite the recognized importance of scale and hierarchical structure for understanding many ecological processes (Levin 1992, Keitt and Urban 2005), including species' niches and distributions (Maurer and Taper 2002, Parmesan et al. 2005), there has been little use of explicitly hierarchical regression models for exploring these relationships. Having undergone more significant development in some social sciences (Raudenbush and Bryk 2002), hierarchical models are only very recently becoming more widely used by ecologists for understanding complex relationships at different scales (Wikle 2003a, b, Clark 2005, McMahan and Diez 2007).

In this study, we use hierarchical generalized linear models (HGLM), within a Bayesian framework, to test hypotheses about presence and abundance patterns of two terrestrial orchids of eastern North America, *Goodyera pubescens* and *Tipularia discolor*, across multiple spatial scales. The two study species have significant differences in key life history characteristics, including requirements for different groups of fungi in order to germinate, which lead to different expectations for patterns of recruitment. *Goodyera pubescens* and *Tipularia discolor* are terrestrial orchids distributed throughout deciduous and mixed forests of eastern North America (see Plate 1). While each may reproduce both clonally and sexually, clonal reproduction of *G. pubescens* occurs via branching growth of rhizomes at the soil surface and appears to be more prolific than that of *T. discolor*, which depends on breakage of below-ground corms (Whigham and O'Neill 1991). Like other orchids, seeds of these species lack significant nutrient reserves and are therefore dependent on colonization by the appropriate fungi for germination.

G. pubescens associates with one or two species of saprotrophic basidiomycetes from the genus *Tulasnella*, whereas *T. discolor* is found associated with a more diverse group of tulasnelloid fungi (McCormick et al. 2004, 2006). Although both species remain heavily colonized by fungi as adults, the seeds of *G. pubescens* are able to germinate with the same fungi as used by adult plants, whereas the fungi found associated with *T. discolor* seeds are distinct from those used by adults (McCormick et al. 2004). Symbiotic germination success has been found to increase sharply in close proximity (<1 m) to adult *G. pubescens* and in patches of higher soil moisture and organic content (Diez 2007), whereas the germination of *T. discolor*, and hence its fungi,

appears to be strictly associated with decaying logs (Rasmussen and Whigham 1998; R. Geffen and J. M. Diez, unpublished data). These differences in life histories suggest that *G. pubescens* is more capable of spreading laterally from established patches due to positive density-dependent sexual and clonal recruitment, and may therefore exhibit greater spatial autocorrelation.

We integrate plant surveys and intensive abiotic monitoring within hierarchical models to test the hypothesis that the influences of different key abiotic variables on the distributions and abundances of these two orchid species are scale dependent. Further, we test the hypothesis that key life history differences between the species, namely the greater clonality and differences in fungal relationships, will lead to more spatial autocorrelation for *G. pubescens* than for *T. discolor*. A more general goal of this paper is to provide an example of some of the underutilized possibilities for hierarchical models to more thoroughly and explicitly investigate multi-scale controls on ecological processes.

MATERIALS AND METHODS

Study design and sampling regime

A hierarchical study design was used to capture a range of environmental conditions across a geographic gradient of ~120 km from the Piedmont of Georgia to the southern Appalachian Mountains, USA. As part of a larger study of understory plant demography, 16 study grids were established across three landscapes, each grid between 250 m² and 480 m² in size. The size of the study grids was chosen to capture population-level processes of understory plants, and each grid was divided into 4-m² cells, within which plants were individually marked with flags. This paper uses "population scale" to refer to inference at the level of study grids and "microsite scale" for inference at the level of 2 × 2 m cells (Fig. 1). The sampling scales were chosen to reflect levels of variability in the abiotic environment in this system and the likely range of relevant plant processes and interactions (e.g., clonal growth, dispersal, and so forth). Overall, 1105 *G. pubescens* individually marked ramets on 10 of the grids and 540 *T. discolor* ramets on nine of the grids were used for analyses. Further detail on the study design can be found in Diez (2005) and in Appendix A.

Abiotic measurements

The abiotic variables considered particularly important to understory plants across this gradient were soil moisture, understory light availability, pH, soil texture, and soil organic content. Fine-scale spatial variability in soil moisture and light availability was assessed through intensive 80-point sampling of each of the grids several times throughout the year for four years, allowing assessment of spatial and temporal patterns in the abiotic environment (Diez 2005). Plant surveys and abiotic monitoring from 2003 are used in this study. The handheld Soil Water Content Measurement System CS620 from Hydrosense (Decagon, Pullman, Washing-

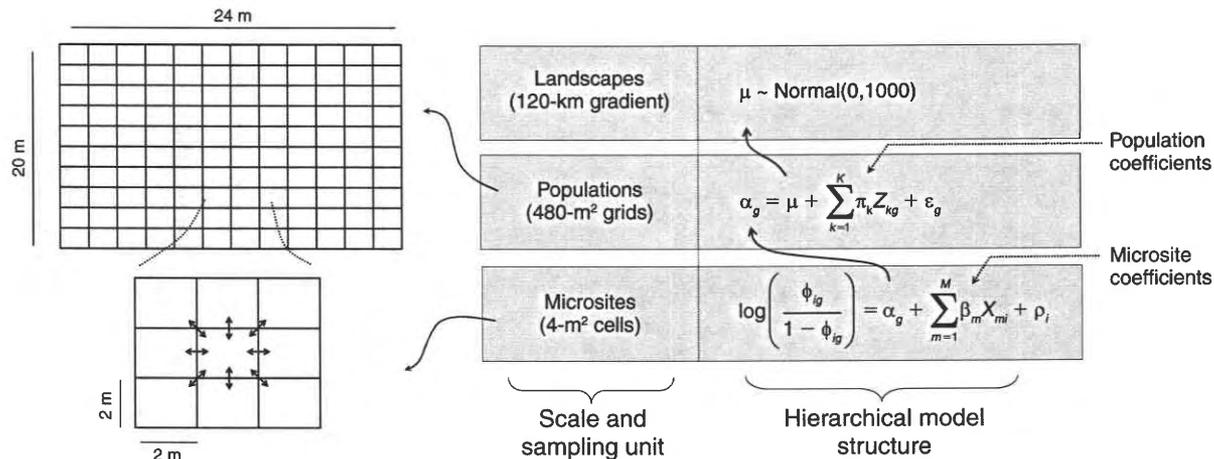


FIG. 1. A hierarchical model structure reflects the sampling design capturing species processes at different scales. Model structure allows prediction of the probability of occurrence, ϕ_{ig} in cell i and grid g , and abundance based on covariates measured at both levels, as well as the effect of neighboring cells via spatial random effects (represented as arrows in the lower left).

ton, USA) was used to measure volumetric soil moisture in the top 12 cm of soil and two AccuPAR ceptometers (Decagon, Pullman, Washington, USA) were used to measure percentage transmittance of incident canopy PAR (photosynthetically active radiation) to the forest floor, with one placed in a nearby clearing as a full sunlight reference. In order to account for differences in timing of moisture measurements in different landscapes relative to rainfall events, soil moisture readings were used as relative wetness values within a landscape and were then scaled according to total precipitation at the landscape scale over the 12 months preceding the survey (Appendix A). Geostatistical methods were used to make predictions at each of the 120 cells within each grid, using exponential variograms and linear detrending to account for directional trends in a few highly sloped grids (Rossi et al. 1992, Goovaerts 2001).

As elaborated in the statistical methods and discussion, the ability to incorporate uncertainty at different levels of a model is very important and increasingly possible. Uncertainty in abiotic measurements, however, was not included in this study (in accordance with other recent studies such as Gelfand et al. 2003, Latimer et al. 2006), but that will be an obvious and potentially important extension in future studies, particularly as the use of distributed abiotic sensors is expanded. The relative abiotic patterns on our grids were consistent at within- and among-grid scales (Diez 2005), adding confidence that a more static depiction of the abiotic environment was justified. Computational limits also restricted the incorporation of abiotic geostatistical routines directly within the distribution models, as may be preferable in the future.

Soil measurements were taken at a lower sampling intensity than moisture and light, designed to capture grid-level means and variation. Three 12-cm soil cores were taken from the center of each of 16 "intensive" cells distributed evenly in each of the 16 study grids and

bulked within cells for a total of 256 samples. A pH meter was used to determine pH of a 1:1 suspension of fresh soil to deionized water; organic content was determined by percentage loss of mass after combustion in a muffle oven; and soil texture was determined using the hydrometer method (Gee and Bauder 1986). Further detail on all sampling methods can be found in Appendix A.

Statistical analysis: hierarchical Bayesian GLMMs

In order to simultaneously consider the influence of key abiotic variables on species distributions at multiple spatial scales, and the influence of local spatial autocorrelation, hierarchical generalized linear models (HGLMs) were implemented within a Bayesian framework. Several recent, detailed descriptions of building hierarchical linear models can be found (Raudenbush and Bryk 2002, Gelman and Hill 2007, McMahon and Diez 2007), including for spatial data (Banerjee et al. 2004), and similar models in the context of analyzing species distributions can be found in Latimer et al. (2006). These models may be considered one implementation of a larger class of spatial models that, in a variety of ways, deal with the "change of support" problem associated with data collected from multiple scales (e.g., Gotway and Young 2002, Banerjee et al. 2004, Wikle and Berliner 2005).

The basic structure of the model is as follows. The occurrence of a species in each cell is modeled as a binary Bernoulli process, $Y_{ig} \sim \text{Bernoulli}(\phi_{ig})$, where Y_{ig} is the binary occurrence data for each cell i and grid g , and ϕ_{ig} are the estimated probabilities of occurrence in those cells. These probabilities of occurrence are related to linear predictors via the logit-link function as

$$\log\left(\frac{\phi_{ig}}{1 - \phi_{ig}}\right) = \alpha_g + \sum_{m=1}^M \beta_m X_{mi} + \rho_i$$

where α_g is a grid-level intercept term, β_m are M

regression coefficients, and X_{mi} are soil moisture and light availability covariate data at the cell (microsite) level, including their squared terms to allow for unimodal responses. The random effects for each cell, ρ_i , are allowed to be spatially dependent via a conditional autoregressive (CAR) specification of the prior, as defined by Besag (1974) and recently used by Gelfand et al. (2003) and Latimer et al. (2006) for similar applications. Thus, the distribution of random effects for cell i is made conditional on values from cells j , as follows:

$$\rho_i | \rho_j \sim \text{Normal} \left(\frac{\sum_{i \neq j} a_{ij} \rho_j}{n_i}, \frac{\sigma_g^2}{n_i} \right)$$

where $a_{ij} = 1$ if cells i and j are neighbors, and 0 otherwise. The n_i are the total number of neighbors of cell i , as defined by the neighborhood structure of choice. Neighbors here are defined as those cells directly adjacent, including those diagonal. Although a variety of spatial covariance structures are possible (for examples, see Banerjee et al. 2004), the CAR structure in this case was justified by the localized spread patterns of these plants and the areal form of data (2×2 m cells, vs. point locations).

The grid-level intercepts α_g provide the link between the cell and higher levels. They are modeled as random effects drawn from a global distribution with a grid-specific mean determined by a global intercept and grid-level covariates as follows:

$$\alpha_g = \mu + \sum_{k=1}^K \pi_k Z_{kg} + \varepsilon_g.$$

Here, Z_{kg} are K covariates at the grid level (pH, soil texture, organic content, and soil moisture), and there are K regression coefficients, π_k , describing those relationships. As for soil moisture in this case, the same abiotic covariates can be modeled at the level at which the data are collected and at higher levels, which allows one to test the importance of variables at multiple scales and avoids bias due to correlated covariates and group-level intercepts (J. Bafumi and A. Gelman, *unpublished manuscript*). The error terms at this level are normally distributed $\varepsilon_g \sim \text{Normal}(0, \sigma_g^2)$, with the variation attributable to the grid level given a noninformative prior, allowing the data to drive their estimation. The intercept (μ) and regression coefficients (π_k) of the grid-level model are given noninformative prior distributions (Appendix B).

To allow for covariance among the light and moisture coefficients (β_m) in the cell-level model (e.g., more moisture may be required in high light environments), these coefficients were modeled with a multivariate normal distribution $\beta \sim \text{MVN}(\mu, \mathbf{R})$ with a Wishart (Ω, ρ) prior for the precision matrix \mathbf{R} . The Wishart is a multivariate generalization of the gamma distribution

and forms a natural prior for the precision matrix of the multivariate normal (Gelman et al. 2004). To represent vague prior knowledge, we chose the degrees of freedom ρ to be 4, the smallest value allowed (because the rank of the precision matrix is also 4), and low parameter values (0.1 along the diagonal and 0.01 elsewhere) for the scale matrix Ω . Implementing the model, we found that the covariance terms were not significantly different from zero, and the final model was simplified to exclude covariance among coefficients.

The models of abundance use a similar specification of hierarchical structure and spatial random effects, with the replacement of the Bernoulli sampling distribution and logit-link function with a Poisson distribution and log-link function. Unconditional models were initially fit to determine the basic variance partitioning of presence and abundance between the microsite and population scales via calculation of interclass correlation coefficients (Raudenbush and Bryk 2002; Appendix B). As a Bayesian model, all parameters are considered random variables and are assigned prior distributions. All models were fit using WinBugs 1.4, which uses Markov chain Monte Carlo (MCMC) sampling methods to characterize the posterior distributions of model parameters (Gilks et al. 1996, Spiegelhalter et al. 2000). Covariate data were standardized by subtracting the overall mean and dividing by the standard deviation in order to improve model convergence and aid interpretation of coefficient estimates. Further model details, including choice of priors and approach to modeling covariance, can be found in Appendix B and the WinBugs code in the Supplement.

Prediction of species responses (occurrence or abundance) across a range of abiotic conditions is achieved within the same model framework as that used to estimate abiotic coefficients. Referred to as "posterior prediction" in statistical literature, prediction of new responses is conditional on the observed responses and estimated parameters. At each iteration of the MCMC, parameter estimates, including regression coefficients at each level of the model, β_m and π_k , and intercepts α_g and μ are used to predict occurrence or abundance at a range of abiotic conditions. Inclusion of this step within the MCMC allows propagation of uncertainty associated with all parameters to influence prediction.

RESULTS

Variance partitioning using unconditional models revealed that 63% of the variation of *Goodyera* presence was distributed among cells compared to 37% among grids, and similarly for *Tipularia*, with 69% among cells and 31% among grids. The patterns were nearly reversed for abundance, with 32% of the variation in *Goodyera* abundance among cells and 68% among grids, and 25% and 75% distributed among cells and grids, respectively, for *Tipularia*. The two species exhibited different relationships with abiotic variables in the full hierarchical models. Both the presence and abundance of *G.*

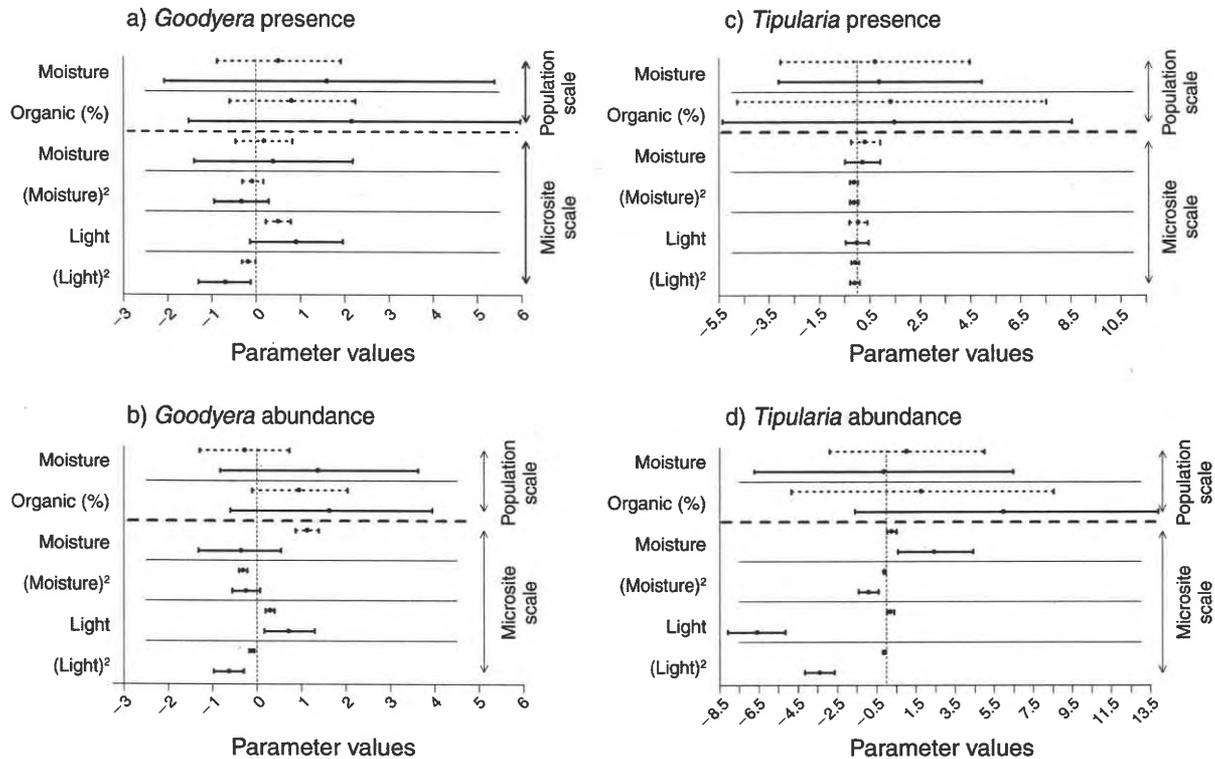


FIG. 2. Parameter estimates of the hierarchical model are represented by 95% credible intervals: literally the interval over which 95% of the density of a posterior probability distribution falls. The models consider presence and abundance of the orchids *Goodyera pubescens* and *Tipularia discolor*. Those variables with intervals that do not overlap the vertical zero line may be considered significantly different from zero at the 95% level. For each model, the top portion of the graph shows those explanatory variables considered at the population (grid) level, π_m , whereas the bottom moisture and light are at the microsite (cell) level, β_m . Second-order moisture and light are included in models to allow for unimodal or saturating responses to abiotic variables. Dotted lines represent estimates from nonspatial models; solid lines are from models including spatial effects.

pubescens showed a significant unimodal response to winter light at the microsite scale and a positive relationship with organic content of the soil at the population level (Fig. 2). *T. discolor* presence showed no significant responses to abiotic variables at either scale, but abundance was significantly correlated with both moisture and light (Fig. 2). Reflecting the hierarchical structure of the model, greater uncertainty was associated with grid-level effects than with the cell-level effects, as judged by the width of the 95% credible intervals (Fig. 2).

Each iteration of the MCMC yields an estimate of the probability of occurrence or abundance and of the spatial random effect for each cell, which yield posterior densities for each cell (displayed as means of those densities in Appendix B: Fig. B1). The significance of spatial effects was judged by summing the number of spatial error terms, ρ_i , whose 95% credible interval did not overlap zero. *G. pubescens* exhibited significant spatial error terms for both presence and abundance (87 and 163 cells, respectively, out of a total of 1020 cells; 8.5% and 16.0% of all cells). *T. discolor*, on the other hand, had no significant spatial effects in the presence model and 156 in the abundance model (out of total possible of 734; 21.2% of all cells).

To understand the impact of including spatial effects, it is informative to compare models with and without spatial error terms, ρ_i (Fig. 2, solid and dotted lines, respectively). The effect of including spatial autocorrelation varies across models and parameters. In most cases there is a movement of parameter estimates toward greater uncertainty when spatial effects are included. The notable exception is for *T. discolor* presence models, in which estimates are nearly the same, matching the result that no spatial effects were estimated to be different from zero for *T. discolor* presence models. In addition to the greater uncertainty with spatial effects, some factors (e.g., moisture) lost apparent significance in the spatial models, suggesting that lack of spatial effects could lead to the wrong conclusion about abiotic driving variables. This is especially important regarding forecasts of the impact of climate change.

Prediction performed within the same model framework as the model used to estimate abiotic coefficients produced volumes of the likelihood of occurrence and abundance as a function of the abiotic environment (Fig. 3). The uncertainty associated with these predictions is large, reflecting the variability across the study region and the complexity of factors that contribute to shaping species distribution patterns.

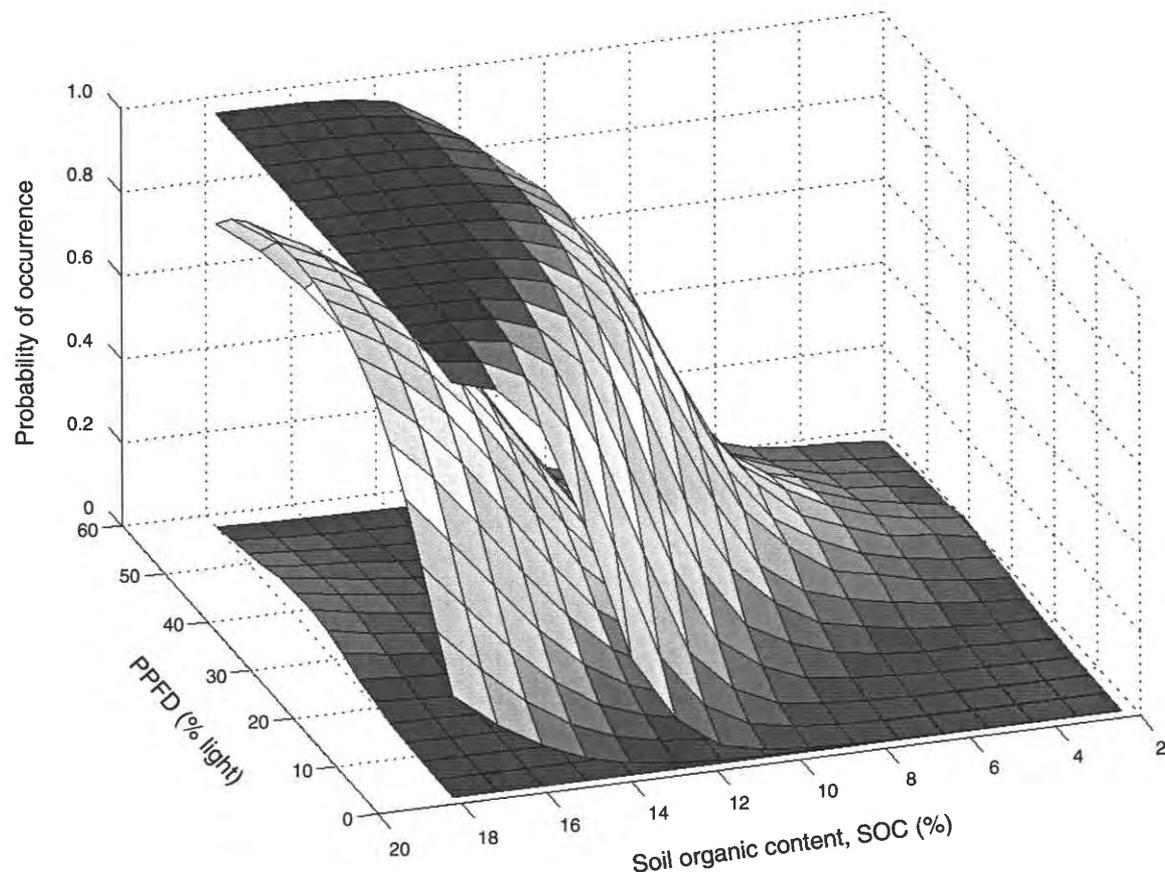


FIG. 3. Posterior parameter estimates can be used to reconstruct predictions of a response variable across a range of the explanatory variables with associated uncertainty. Shown here are three slices of a probability volume predicting the probability of occurrence for *G. pubescens*. The middle surface represents the mean posterior probability prediction (the best guess, given the current data), and the upper and lower surfaces represent the bounds within which we are 95% sure that the probability of occurrence will fall for a given combination of abiotic variables. PPFD is photosynthetic photon flux density (measured as the percentage of incident light transmitted through the canopy to the ground level).

DISCUSSION

An important, general effort in ecology is to understand how deterministic and stochastic forces combine to shape ecological processes across spatial and temporal scales. This may be addressed through experimental designs that can capture both kinds of processes, and also by using statistical methods that can accommodate both. This study used one such approach, within a hierarchical Bayesian framework, to evaluate the responses of two terrestrial orchids to environmental gradients while estimating additional, unexplained local processes that mediate these responses. We found support for the idea that species distributions can be linked to key abiotic variables at multiple scales, but the data also suggest that additional biotic or abiotic processes lead to local autocorrelation in patterns of presence and abundance. Importantly, it was shown that inclusion of spatial autocorrelation can change the magnitude and the certainty of estimated effects of the abiotic environment.

Other studies have used spatially correlated errors to investigate autocorrelation at landscape scales (e.g., Gelfand et al. 2003, Hooten et al. 2003, Latimer et al. 2006), and inclusion of spatial autocorrelation is known to allow more robust prediction of responses to abiotic gradients (Lichstein et al. 2002), but this study suggests their potential for understanding spatial processes at a local, within-population scale. The finding in this study of significant spatial effects for *G. pubescens* presence and not for *T. discolor* presence is consistent with the hypothesis that life history traits may contribute to detectable patterns of spatial autocorrelation in species distributions. The substrate specificity of the fungi needed by *T. discolor*, an apparently strict association with decaying logs (Rasmussen and Whigham 1998; R. Geffen and J. M. Diez, unpublished data), and the different identities of fungi associated with adults and protocorms, suggest a more limited potential for lateral spread of *T. discolor*. Likely due to the ability of *G. pubescens* to germinate with the same fungi associated with adults, probabilities of germination have been



PLATE 1. The two study species, terrestrial orchids found in forests of eastern North America: (left) adult and juvenile (perhaps a clonal offspring) *Goodyera pubescens*; (right) a juvenile *Tipularia discolor* emerging from the underside of a decaying log (this is often where their symbiotic fungi are found). Photo credits: Robert Warren (*Goodyera*) and Scott Eustis (*Tipularia*).

shown to increase dramatically close to adult plants (Diez 2007). Together with the greater proficiency for clonal spread of *G. pubescens* via creeping rhizomes, there is good reason to believe that these life history differences are at least partly responsible for the differences in estimated spatial structure in the presence models.

The fact that both species showed stronger spatial autocorrelation and responses to abiotic variables suggests somewhat different dynamics for abundance than for simple occurrence. *T. discolor*, in fact, showed no spatial autocorrelation for presence but a significant amount for abundance. It is plausible that the highly clumped recruitment of the species based on locations of decaying logs is enough to drive the pattern of spatial dependence in abundance, but the limited ability to clonally spread or recruit off the log using the fungi associated with adults prevents the type of lateral spread that leads to spatial autocorrelation in presence.

Studies investigating the relationship between patterns of species' occupancy (presence-absence) and abundance predominately find positive relationships between species' distributions and local abundance, but this relationship appears to vary across taxa and habitats (Blackburn et al. 2006). Although the mechanisms underlying this pattern are not well understood, it is likely that heterogeneity of habitat suitability and scale of measurement influence observed abundance-occupancy patterns and the interpreted process (He and Gaston 2000, Freckleton et al. 2006). Although our focus in this study was not the direct relationship between abundance and occupancy, a few observations derived from our study may be relevant to that area of research. In particular, the degree to which habitat suitability influences patterns of site occupancy and abundance may now be examined more explicitly, while accounting for spatial autocorrelation. Further, it may be useful to more explicitly incorporate differences in

species life histories for predicting the relative concordance of abundance and occupancy patterns.

Species distributions, scale, and prediction

Recent reviews of species distribution modeling have appropriately called for increased attention to scale (Pearson and Dawson 2003, Guisan and Thuiller 2005) because it is generally understood that different processes are likely to influence species at multiple spatial scales (Maurer and Taper 2002). This can make identification of driving variables difficult in the field, and challenging to estimate and interpret within models. Recent developments easing the implementation of hierarchical generalized linear models (HGLM) are establishing these approaches as very useful for simultaneously evaluating information from different scales (Raudenbush and Bryk 2002, Gelman and Hill 2007, McMahon and Diez 2007). Although various methods exist for identifying scales of variability in continuous response variables (e.g., Borcard et al. 2004, Keitt and Urban 2005), the ecological data are often structured hierarchically, due both to inherent structure (e.g., streams within watersheds) or imposed by sampling designs (e.g., plots within grids), and tools are increasingly available to identify and quantify key relationships at different scales using such data.

Relatively simple hierarchical models, as shown in this study, can be used to explicitly estimate the effects of different factors at multiple spatial scales. The finding that soil organic content helps to explain *G. pubescens* abundances at the grid level is far more informative than simply finding a "grid effect," as might be done with standard mixed models. The likely ecological explanation for such a pattern is related to the tight relationships with saprotrophic fungi. Moreover, models that ignore the hierarchical structure in processes and in data risk the "ecological fallacy" of applying group-level attributes to individuals (Gelman et al. 2001) and the

related problem of pseudoreplication (Hurlbert 1984), or must lose information by averaging over lower levels.

An additional opportunity afforded by Bayesian models is the relatively straightforward transition between parameter estimation and prediction with estimates of uncertainty (Gelman et al. 2004, Clark 2005). As shown here, estimated relationships with abiotic variables can be used to make predictions about likely responses to gradients with associated estimates of uncertainty. The full posterior distributions of estimated effects are well suited for such predictions, and could facilitate placement of such predictions within a dynamic model incorporating dispersal processes. The explicitly multilevel structure of the models importantly allows prediction at any scale to be fully conditional on uncertainties across other scales (Gelman et al. 2004). The uncertainty in the predictions in this study is large, probably reflecting the importance of other variables, potentially both biotic and abiotic, that may influence the species' distributions at the measured scales. It is critical that such uncertainty be estimated, however, in order to assess the degree to which reliable predictions can be made using key driving variables (Clark et al. 2001).

In summary, the ongoing debate about the relative importance of niche partitioning and stochastic processes for shaping species distributions and community composition suggests a great utility in approaches to inference that can incorporate and estimate both kinds of processes. Both niche theory and efforts to predict species and community responses to environmental change are increasingly incorporating both niche-based and stochastic, nonequilibrium processes. Extending this discussion to include scale-explicit responses to abiotic variables and the role of local spatial interactions will greatly aid the effort to understand species dynamics.

ACKNOWLEDGMENTS

The field research was supported by NSF grants DEB-0235371 to H. R. Pulliam and DEB-9632854 (the Coweeta LTER) to the University of Georgia Research Foundation. The authors thank Itamar Giladi, Robert Warren, and Scott Eustis for comments on an earlier draft, and the helpful comments of two anonymous reviewers. J. M. Diez is grateful to Alan Gelfand and Andrew Latimer for useful discussions of GLMMs during NSF-funded Summer Institute on Ecological Forecasting at Duke University, and also Mike Conroy, Chris Fonnesebeck, and lab group for helpful discussions of hierarchical modeling.

LITERATURE CITED

- Andrewartha, H. G., and L. C. Birch. 1954. The distribution and abundance of animals. University of Chicago Press, Chicago, Illinois, USA.
- Banerjee, S., B. P. Carlin, and A. E. Gelfand. 2004. Hierarchical modeling and analysis for spatial data. Chapman and Hall/CRC, Boca Raton, Florida, USA.
- Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems (with discussions). *Journal of the Royal Statistical Society, Series B* 36:192–236.
- Blackburn, T. M., P. Cassey, and K. J. Gaston. 2006. Variations on a theme: sources of heterogeneity in the form of the interspecific relationship between abundance and distribution. *Journal of Animal Ecology* 75:1426–1439.
- Borcard, D., P. Legendre, C. Avois-Jacquet, and H. Tuomisto. 2004. Dissecting the spatial structure of ecological data at multiple scales. *Ecology* 85:1826–1832.
- Chase, J. M., and M. A. Leibold. 2003. Ecological niches: linking classical and contemporary approaches. University of Chicago Press, Chicago, Illinois, USA.
- Clark, J. S. 2005. Why environmental scientists are becoming Bayesians. *Ecology Letters* 8:2–14.
- Clark, J. S., et al. 2001. Ecological forecasts: an emerging imperative. *Science* 293:657–660.
- Diez, J. M. 2005. Ecological dynamics of a terrestrial orchid symbiosis. Dissertation. University of Georgia, Athens, Georgia, USA.
- Diez, J. M. 2007. Hierarchical patterns of symbiotic orchid germination linked to adult proximity and environmental gradients. *Journal of Ecology* 95:159–170.
- Freckleton, R. P., D. Noble, and T. J. Webb. 2006. Distributions of habitat suitability and the abundance–occupancy relationship. *American Naturalist* 167:260–275.
- Gee, G. W., and J. W. Bauder. 1986. Particle size analysis. Pages 383–411 in A. Klute, editor. *Methods of soil analysis. Part 1: Physical and mineralogical methods*. Agronomy Monographs 9(1). American Society of Agronomy, Madison, Wisconsin, USA.
- Gelfand, A. E., J. A. Silander, W. Shanshan, A. M. Latimer, P. O. Lewis, A. G. Rebelo, and M. Holder. 2003. Explaining species distribution patterns through hierarchical modeling. *Bayesian Analysis* 1:1–47.
- Gelman, A., J. B. Carlin, and H. S. S. B. Rubin. 2004. Bayesian data analysis. Second edition. Chapman and Hall/CRC, New York, New York, USA.
- Gelman, A., and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge, UK.
- Gelman, A., D. K. Park, S. Ansolabehere, P. N. Price, and L. C. Minnite. 2001. Models, assumptions and model checking in ecological regressions. *Journal of the Royal Statistical Society Series A, Statistics in Society* 164:101–118.
- Gilks, W. R., S. Richardson, and D. J. Spiegelhalter, editors. 1996. *Markov chain Monte Carlo in Practice*. Chapman and Hall, London, UK.
- Goovaerts, P. 2001. Geostatistical modelling of uncertainty in soil science. *Geoderma* 103:3–26.
- Gotway, C. A., and L. J. Young. 2002. Combining incompatible spatial data. *Journal of the American Statistical Association* 97:632–648.
- Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8:993–1009.
- Hampe, A. 2004. Bioclimate envelope models: what they detect and what they hide. *Global Ecology and Biogeography* 13: 469–471.
- He, F. L., and K. J. Gaston. 2000. Occupancy–abundance relationships and sampling scales. *Ecography* 23:503–511.
- Holyoak, M., and M. Loreau. 2006. Reconciling empirical ecology with neutral community models. *Ecology* 87:1370–1377.
- Hooten, M. B., D. R. Larsen, and C. K. Wikle. 2003. Predicting the spatial distribution of ground flora on large domains using a hierarchical Bayesian model. *Landscape Ecology* 18: 487–502.
- Hubbell, S. P. 2006. Neutral theory and the evolution of ecological equivalence. *Ecology* 87:1387–1398.
- Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54:187–211.
- Keitt, T. H., and D. L. Urban. 2005. Scale-specific inference using wavelets. *Ecology* 86:2497–2504.

- Latimer, A. M., S. Wu, A. E. Gelfand, and J. A. Silander. 2006. Building statistical models to analyze species distributions. *Ecological Applications* 16:33–50.
- Leibold, M. A., M. Holyoak, N. Mouquet, P. Amarasekare, J. M. Chase, M. F. Hoopes, R. D. Holt, J. B. Shurin, R. Law, D. Tilman, M. Loreau, and A. Gonzalez. 2004. The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters* 7:601–613.
- Leibold, M. A., and M. A. McPeck. 2006. Coexistence of the niche and neutral perspectives in community ecology. *Ecology* 87:1399–1410.
- Levin, S. A. 1992. The problem of pattern and scale in ecology. *Ecology* 73:1943–1967.
- Lichstein, J. W., T. R. Simons, S. A. Shiner, and K. E. Franzreb. 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs* 72:445–463.
- Maurer, B. A., and M. L. Taper. 2002. Connecting geographical distributions with population processes. *Ecology Letters* 5:223–231.
- McCormick, M. K., D. F. Whigham, and J. O'Neill. 2004. Mycorrhizal diversity in photosynthetic terrestrial orchids. *New Phytologist* 163:425–438.
- McCormick, M. K., D. F. Whigham, D. Sloan, K. O'Malley, and B. Hodkinson. 2006. Orchid–fungus fidelity: a marriage meant to last? *Ecology* 87:903–911.
- McMahon, S. M., and J. M. Diez. 2007. Scales of association: using hierarchical linear models to measure ecological systems. *Ecology Letters* 10:437–452.
- Parmesan, C., S. Gaines, L. Gonzalez, D. M. Kaufman, J. Kingsolver, A. T. Peterson, and R. Sagarin. 2005. Empirical perspectives on species borders: from traditional biogeography to global change. *Oikos* 108:58–75.
- Pearson, R. G., and T. P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? *Global Ecology and Biogeography* 12:361–371.
- Pulliam, H. R. 2000. On the relationship between niche and distribution. *Ecology Letters* 3:349–361.
- Rasmussen, H. N., and D. F. Whigham. 1998. Importance of woody debris in seed germination of *Tipularia discolor* (Orchidaceae). *American Journal of Botany* 85:829–834.
- Raudenbush, S. W., and A. S. Bryk. 2002. Hierarchical linear models: applications and data analysis methods. Sage Publications, Thousand Oaks, California, USA.
- Rossi, R. E., D. J. Mulla, A. G. Journel, and E. H. Franz. 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs* 62:277–314.
- Skov, F., and J. C. Svenning. 2004. Potential impact of climatic change on the distribution of forest herbs in Europe. *Ecography* 27:366–380.
- Spiegelhalter, D., A. Thomas, and N. Best. 2000. WinBUGS User Manual. MRC Biostatistics Unit, Cambridge, UK.
- Tilman, D. 2004. Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assembly. *Proceedings of the National Academy of Sciences (USA)* 101:10854–10861.
- Whigham, D., and J. O'Neill. 1991. The dynamics of flowering and fruit production in two eastern North American terrestrial orchids, *Tipularia discolor* and *Liparis lilifolia*. Pages 89–101 in T. Wells and J. H. Willems, editors. *Population ecology of terrestrial orchids*. SPB Academic Publishing, The Hague, The Netherlands.
- Whittaker, R. J., M. B. Araujo, J. Paul, R. J. Ladle, J. E. M. Watson, and K. J. Willis. 2005. Conservation biogeography: assessment and prospect. *Diversity and Distributions* 11:3–23.
- Wikle, C. K. 2003a. Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology* 84:1382–1394.
- Wikle, C. K. 2003b. Hierarchical models in environmental science. *International Statistical Review* 71:181–199.
- Wikle, C. K., and L. M. Berliner. 2005. Combining information across spatial scales. *Technometrics* 47:80–91.

APPENDIX A

Details of the study design and methods (*Ecological Archives* E088-194-A1).

APPENDIX B

Model details (*Ecological Archives* E088-194-A2).

SUPPLEMENT

WinBugs code (*Ecological Archives* E088-194-S1).

