

Seeing the forest and the trees: multilevel models reveal both species and community patterns

MICHELLE M. JACKSON,^{1,†} MONICA G. TURNER,¹ SCOTT M. PEARSON,² AND ANTHONY R. IVES¹

¹*Department of Zoology, University of Wisconsin, Madison, Wisconsin 53706 USA*

²*Department of Natural Sciences, Mars Hill College, Mars Hill, North Carolina 28754 USA*

Citation: Jackson, M. M., M. G. Turner, S. M. Pearson, and A. R. Ives. 2012. Seeing the forest and the trees: multilevel models reveal both species and community patterns. *Ecosphere* 3(9):79. <http://dx.doi.org/10.1890/ES12-00116.1>

Abstract. Studies designed to understand species distributions and community assemblages typically use separate analytical approaches (e.g., logistic regression and ordination) to model the distribution of individual species and to relate community composition to environmental variation. Multilevel models (MLMs) offer a promising strategy for integrating species and community-level analyses. Here, we demonstrate how MLMs can be used to analyze differences in species composition of communities across environmental gradients. We first use simulated data to show that MLMs can outperform three standard methods that researchers use to identify environmental drivers of the species composition of communities, redundancy analysis (RDA), canonical correspondence analysis (CCA), and nonmetric multidimensional scaling (NMDS). In particular, MLMs can separate the effects of collinearity among environmental drivers and factor out the effect of changes in overall species abundances or occurrences that do not involve changes in composition. We then apply MLMs to presence/absence data for 14 species of understory herbs and topographic, biotic, and edaphic variables measured in 54 forested plots in the Southern Appalachian Mountains. In addition to providing information about community composition, MLMs simultaneously identify the responses of individual species to the environmental variables. Thus, MLMs not only have potentially superior statistical properties in analyses of community composition compared to standard methods, but they simultaneously provide detailed information about species-specific responses underlying the changes in community composition.

Key words: environmental gradients; generalized linear mixed models; multilevel models; ordination; understory herbs; Southern Appalachian Mountains, USA; species distributions.

Received 7 May 2012; revised 15 August 2012; accepted 24 August 2012; **published** 24 September 2012. Corresponding Editor: S. Collins.

Copyright © 2012 Jackson et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits restricted use, distribution, and reproduction in any medium, provided the original author and sources are credited.

† **E-mail:** mgooch@wisc.edu

INTRODUCTION

Ecologists have a long-standing interest in understanding the distributions of species and differences among species assemblages (Clements 1916, Gleason 1926, Whittaker 1956). There are numerous quantitative methods for modeling species distributions, all of which have different advantages and disadvantages (see Legendre and Legendre 1998, Guisan and Zimmermann

2000, Elith et al. 2006). Methods for investigating the effects of environmental variables on species and ecological communities generally fall into two categories: those that focus on the distributions of individual species and those that focus on differences in the composition of communities. While these approaches might be combined within the same study, they are typically performed as separate analyses (e.g., Turner et al. 2004, Driscoll and Weir 2005, Hottola and

Siitonen 2008). The composition of communities is often analyzed using ordination methods that describe how species assemblages vary (Bray and Curtis 1957) by placing communities along one or multiple axes that explain a high proportion of the variance among communities (McCune and Grace 2002). Although ordination nicely describes differences in community composition, disentangling the drivers of these patterns is often indirect, as many environmental variables may be correlated with a single axis. To analyze distributions of individual species, logistic or linear regression is often used (e.g., Pearce et al. 1994, Turner et al. 2003). Regression analyses can be difficult to interpret, however, because they require separate models for each species within an ecological community (see Guisan and Thuiller 2005, Dormann 2007). Some studies have addressed these difficulties by describing the shapes of multiple species' ecological response curves along environmental gradients (Oksanen and Minchin 2002, Rydgren et al. 2003) and using descriptive hierarchical models (Huisman et al. 1993). Still, it is difficult to tie together results from ordination on community composition with results from regression on separate species, even though they come from the same data.

Multilevel models (MLMs; Gelman and Hill 2007) offer a promising strategy for integrating species- and community-level analyses from which inference can be drawn at both levels simultaneously. To understand this approach, consider a data set giving the presence/absence of N species among M sites, with several environmental variables measured for each site. A MLM could be constructed as a set of simultaneous logistic regressions for the presence/absence of each species using the environmental measurements as independent variables. In addition to information about the distributions of individual species, the MLM allows estimation and statistical tests for variation among species in their response to the different environmental variables. Because differences among communities result from variation among species' responses to environmental variables, by testing for variation among individual species the MLM will identify differences in community composition and associate these with environmental variables. MLMs have been used in ecology to incorporate variables into predictive models at different

spatiotemporal scales (Cushman and McGarigal 2004, Duffy et al. 2010, Qian et al. 2010). Furthermore, Qian and Shen (2007) show that MLMs are able to discern a treatment effect that is smaller than detected by conventional approaches. However, the application of MLMs for easily combining species- and community-level analyses of community composition is new.

To demonstrate how MLMs can be used to analyze changes in species composition of communities across environmental gradients, we first used simulated data to assess the ability of MLMs to separate the effects of multiple environmental drivers on community composition. Using the simulated data, we compared the MLM with three frequently used multivariate methods, including two canonical analysis methods (redundancy analysis, RDA, and canonical correspondence analysis, CCA) and one indirect ordination method (nonparametric multidimensional scaling, NMDS), to identify environmental drivers of community composition statistically. We then applied MLMs to field data on the presence/absence of 14 herbaceous species common to southern Appalachian deciduous forests to address two questions: (1) What environmental variables predict the presence/absence and community composition of forest herbs? (2) How do the results from a MLM approach compare to those from RDA, CCA, and NMDS? Our main goal is to introduce MLMs and demonstrate their performance in identifying environmental gradients that may determine both species distributions and community composition. Therefore, we use the real data set to demonstrate the usefulness of the MLM approach, rather than present a comprehensive ecological analysis of herbaceous species distributions in the Southern Appalachians. Furthermore, we compare the MLM approach with three standard methods, rather than attempt to conduct a comprehensive comparison of more-recent methods for analyzing community composition data.

METHODS

Here, we first present a simulation model for communities assembled along environmental gradients to compare MLM with RDA, CCA, and NMDS; we describe the simulations before describing the MLM approach to give a clear

initial picture of the types of data we consider. After describing the methods, we present the data from the southern Appalachian deciduous forests and the ecological background for this study.

Simulations

We simulated community composition using a pool of 20 hypothetical species for each of 30 sites that varied along 5 environmental gradients. We assumed that the number of individuals of each species i at each site j follows a Poisson distribution with mean λ_i that depends on $x_{k,j}$, the value of environmental variable k in site j ,

$$\lambda_i = \exp(a_i + b_{i,1}x_{1,j} + b_{i,2}x_{2,j} + b_{i,3}x_{3,j} + b_{i,4}x_{4,j} + b_{i,5}x_{5,j}). \quad (1)$$

For simplicity, we assume that species respond linearly to the environmental variables, although quadratic or other nonlinear functions could equally be used (and were used in the analysis of the forest herb data set). To generate variation among communities, we assumed that species differ in the value of their coefficients $b_{i,k}$ that dictate their responses to the environmental gradients. We set up challenges for statistical methods to identify species and community responses to environmental gradients by simulating data under the following scenario:

1. Environmental variable x_1 varies among sites independently from other environmental variables, and the values of $b_{i,1}$ are given by a Gaussian random variable with mean zero and standard deviation 0.4.
2. The distribution of x_2 among sites is highly correlated with the distribution of x_3 ($\rho = 0.7$), and like x_1 the values of $b_{i,2}$ are given by a Gaussian random variable with mean zero and standard deviation 0.4.
3. The distribution of x_3 is highly correlated with that of x_2 ($\rho = 0.7$), yet environmental variable 3 has no effect on the distribution of species ($b_{i,3} = 0$).
4. Environmental variable x_4 varies among sites independently of other environmental variables, and the values of $b_{i,4} = 0.5$ for all species, implying all species respond in the same way to x_4 .
5. Environmental variable x_5 varies among sites independently from other environmen-

tal variables, yet species do not respond to x_5 ($b_{i,5} = 0$).

Simulated in this way, a statistical test should identify environmental variables x_1 and x_2 as explaining variation in community composition among sites, while the remaining variables x_3 , x_4 , and x_5 should not be significant.

We have made an explicit assumption about the definition of community composition. We assumed that two communities that have the same relative abundance of species have the same composition, regardless of the total abundance of species. Thus, even though environmental variable x_4 increases the abundance of species, it does not change composition because it changes the abundances of all species in the same way. Specifically, if the mean abundances of species 1 and 2 are $\exp(a_1 + b_{1,4}x_{4,j})$ and $\exp(a_2 + b_{2,4}x_{4,j})$ in site j , and if $b_{1,4} = b_{2,4}$, then the relative mean abundance is $\lambda_1/\lambda_2 = \exp(a_1 - a_2)$ which is independent of x_4 .

These assumptions created community data in which the least and most common species occurred in an average of 4.9 and 29.3 sites, respectively, and on average species occurred in 18.4 sites. These characteristics of the simulated data sets correspond roughly to the characteristics of the real data set from the southern Appalachian deciduous forests, but with a slightly greater range of occurrences between least and most common species (see *Results*).

Multilevel models

We used a multilevel modeling (MLM) framework in which the abundance or presence/absence of species is the response (dependent) variable and environmental variables are the predictor (independent) variables (Gelman and Hill 2007). The MLM makes it possible to estimate simultaneously the responses of multiple species to environmental variables, thereby also giving a summary of the environmental determinants of community composition.

For the case of abundance data, a formal description of the model with only one environmental variable (x) is

$$\text{PR}(Y_q = n) = \text{Poisson}(\lambda_q)$$

$$\lambda_q = \exp(a_{\text{spp}[q]} + b_{\text{spp}[q]}x_{\text{site}[q]})$$

$$\begin{aligned}
 a_{\text{spp}[q]} &= \alpha + d_{\text{spp}[q]} \\
 b_{\text{spp}[q]} &= \beta + e_{\text{spp}[q]} \\
 d &\sim \text{Gaussian}(0, \sigma_{\text{intercept}}^2) \\
 e &\sim \text{Gaussian}(0, \sigma_{\text{slope}}^2)
 \end{aligned}
 \tag{2}$$

where Y_q is the number of individuals of each species at each site ($q = 1, \dots, NM$, where N is the number of species and M is the number of sites). The expectations of the Poisson distributions λ_q are themselves treated as random variables, with the distribution of $\log(\lambda_q)$ containing both fixed and random effects. The intercepts and slopes for each species, $a_{\text{spp}[q]}$ and $b_{\text{spp}[q]}$, have mean values of α and β (treated as fixed effects) and variances $\sigma_{\text{intercept}}^2$ and σ_{slope}^2 given in the Gaussian random variables (random effects) d and e . The functions $\text{spp}[q]$ and $\text{site}[q]$ are used to give the identity of the species and site that corresponds to observation q in the data set (Gelman and Hill 2007).

The MLM given by Eqs. 2 can be interpreted as a set of Poisson regressions in which differences in slopes and intercepts among species are random variables. Thus, for example, if there were a data set with only three species having slopes from Poisson regressions of b_1 , b_2 and b_3 , then the model would assume that the values of b_1 , b_2 and b_3 were drawn independently from a Gaussian distribution with mean β and variance σ_{slope}^2 . Therefore, the fixed effect β gives the average response of the individual species to the environmental variable, and the variance of the random effect σ_{slope}^2 gives the variability among species in their response to the environmental variable (i.e., changes in community composition). Differences in community composition are embodied in σ_{slope}^2 , because the greater the difference among species in response to the environmental variable, the greater will be the changes in community composition along the corresponding environmental gradient. Although we have presented Eqs. 2 with only a single environmental variable, it can be extended to include numerous environmental variables. Similarly, although we have assumed that $\log(\lambda_q)$ is a linear function of the environmental gradi-

ents, quadratic and higher-order polynomial terms could be added (as both fixed and random effects) to account for nonlinear responses of species to environmental gradients.

The MLM for presence/absence of species can be constructed similarly;

$$\begin{aligned}
 \Pr(Y_q = 1) &= \mu_q \\
 \mu_q &= \text{logit}^{-1}(a_{\text{spp}[q]} + b_{\text{spp}[q]}x_{\text{site}[q]}) \\
 a_{\text{spp}[q]} &= a + d_{\text{spp}[q]} \\
 b_{\text{spp}[q]} &= \beta + e_{\text{spp}[q]} \\
 d &\sim \text{Gaussian}(0, \sigma_{\text{intercept}}^2) \\
 e &\sim \text{Gaussian}(0, \sigma_{\text{slope}}^2)
 \end{aligned}
 \tag{3}$$

where Y_q is the presence (1) or absence (0) of species among sites. The logit function, $\text{logit}(p) = \log(p/(1 - p))$, takes values from $-\infty$ to $+\infty$ as p varies from 0 to 1. Otherwise, the MLM is the same as that for Poisson-distributed data (Eqs. 2). A similar model, but including phylogenetic relationships among species, is derived in Ives and Helmus (2011).

Identifying environmental effects on community composition involves testing whether variances σ_{slope}^2 differ from zero. For this we performed likelihood ratio tests, comparing the models with and without the variance term using maximum likelihood (ML). Because the possible values of σ^2 are constrained to be positive, the asymptotic values of the log-likelihood ratio, $-2\log L$, are distributed by a 50:50 mixture of χ^2_0 and χ^2_1 distributions, so that the P-values given by the constrained likelihood ratio test are half the values of those calculated from the χ^2_1 that would be computed in an unconstrained likelihood ratio test (Self and Liang 1987, Stram and Lee 1994). The estimates of the mean responses of species to environmental gradient k , β_k , asymptotically follows a Gaussian distribution, which is used for statistical tests. The MLMs were constructed using lmer (Bates et al. 2008) in R version 2.14.1 (R Development Core Team 2011), and example code is provided in the Supplement.

The results from MLM analyses can be displayed as biplots like those produced by canonical analysis methods. To focus on the

variation among sites caused by variation in species-specific responses to environmental gradients, we computed the predicted values of λ_q (Eqs. 2) or μ_q (Eqs. 3), and subtracted from these the predicted values calculated without the environmental random effects ($e_{\text{spp}[i]} = 0$); the resulting values represent variation among sites that is explained by differences among species in their responses to the environmental gradients. We then performed a principal components analysis (PCA) on these values. This use of PCA on predicted values is similar to redundancy analysis (RDA) (Palmer 1993), although for RDA the PCA is performed on the predicted values from standard regression when plotting the results in terms of the linear combination (LC) scores (Oksanen et al. 2011). Arrows depicting the influence of each environmental variable are calculated by weighting each site by the value of its environmental variable and plotting the centroid at the tip of the arrow.

RDA, CCA, and NMDS

Redundancy analysis (RDA) and canonical correspondence analysis (CCA) are canonical analyses that are designed to identify variation among communities that can be explained by data on environmental gradients (Legendre and Legendre 1998). They are closely related, differing in that RDA is based on the Euclidean distance between communities whereas CCA is based on the χ^2 distance. Nonmetric multidimensional scaling (NMDS) is an indirect ordination method (Clark 1993), although in conjunction with permutation tests involving environmental variables, it is often used to assess the relationships between community composition and environmental variables. NMDS has the potential advantage of preserving the ordering of relationships among communities but not the distance, in this sense not requiring a strict adherence to an a priori distance metric. Nonetheless, a distance metric must still be specified to determine distance ranks. We used the Bray-Curtis (Bray and Curtis 1957) distance and a maximum of 200 iterations for the NMDS. We compared the MLM with all three of these established methods, because all are frequently used in analyses of community composition along environmental gradients.

For application to the simulated data, we used

partial RDA and CCA in which the effect of each environmental gradient x_i was determined after “partialing out” the effects of the other environmental variables (Legendre and Legendre 1998); this should improve the separation of effects in the face of collinearity such as between x_2 and x_3 . For abundance data, we used a log(+1) transformation. Statistical significance of the associations of environmental variables with variation in community composition was obtained using 10,000-permutation tests. Finally, to compare the results among the four methods, we calculated the procrustes score between each pairwise set of results. The procrustes score gives a measure of minimum dissimilarity between two matrices, in our specific case the community values on the first two axis representing the greatest variation among communities. All analyses were performed using the vegan library (Oksanen et al. 2011) in R version 2.14.1 (R Development Core Team 2011), and example code is provided in the Supplement.

Application of MLM to understory herbs

To illustrate the use of MLM and other methods, we used new field data from a study of the distribution of native understory herbs in the Southern Appalachians, for which surprisingly little is known despite their ecological significance. Understory forest herbs represent the majority of plant diversity in temperate forests and are especially sensitive to disturbance across broad temporal and spatial scales (Gilliam 2007). Furthermore, many have extremely limited dispersal (Ehrlén and Eriksson 2000) and must respond to changing conditions rapidly or face local extinction. Many forest herbs occupy their southerly range limits in the Southern Appalachians due to cool, moist, high-elevation climates. If this region warms significantly as predicted (as much as 1–7°C according to Mearns et al. 2003), certain species may be faced with severely reduced ranges as high elevation habitats shrink in size and are replaced by warmer, drier climates currently found at lower elevations.

We applied MLM, RDA, CCA, and NMDS to presence/absence data for 14 native herbaceous species among 54 plots in the French Broad River Basin of western North Carolina, USA (see Appendix for detailed field sampling methods).

Environmental variables recorded for each plot included total herb and shrub cover, litter depth, basal area of all trees using the point-quarter count method (Stearns 1949), elevation, aspect (converted to a relative moisture value), slope, terrain shape, and distance to stream. Five soil cores were obtained from each plot and analyzed for soil texture, total nitrogen, organic matter, phosphorus, potassium, calcium, magnesium, and pH. In the statistical analyses, we included the basal area of sugar maple (*Acer saccharum*), tulip poplar (*Liriodendron tulipifera*), and chestnut oak (*Quercus prinus*) because they produce litter of varying quality and grow on sites with different moisture levels. *A. saccharum* and *L. tulipifera* dominate moist sites, and *Q. prinus* dominates drier sites (Whittaker 1956). *L. tulipifera* is a fast-growing tree common to historically disturbed sites (Kuhman et al. 2011).

After checking for collinearity among variables, we excluded soil nitrogen and organic matter from the models because they were highly correlated with elevation ($\rho = 0.72$, $\rho = 0.73$), which was retained because it is a strong proxy for climate (Bolstad et al. 1998). We used phosphorus and calcium as representative soil variables in the models. Calcium was strongly correlated with pH ($\rho = 0.70$), magnesium ($\rho = 0.70$), and potassium ($\rho = 0.59$). We transformed all environmental variables by subtracting the mean and dividing by the standard deviation, so that each variable had mean zero and variance 1; this makes it possible to compare regression coefficients directly because they measure effect sizes. To reduce the number of variables in the MLM, we used Akaike's information criterion (AIC) for model selection, always including the fixed effect β for an environmental variable if the corresponding random effect $e_{\text{spp}[i]}$ was included. Quadratic forms of all environmental variables were included as both fixed and random effects in the model selection to account for possible nonlinear effects on species occurrences.

We performed RDA, CCA, and NMDS ordinations on the herb data set using the 7 environmental variables that were significant as either fixed or random effects in the best MLM; these were elevation, soil calcium, soil phosphorus, basal area of *L. tulipifera* and *Acer saccharum*, and herb cover. We limited the analyses to the 7 variables identified by MLM, because RDA and

CCA are recommended for use with only environmental variables likely to be important, rather than for data exploration (Legendre and Legendre 1998), and to simplify comparisons among methods. We excluded quadratic effects from the ordination analyses because they can cause warping of ordination space (ter Braak and Prentice 1988), and because no significant quadratic terms were identified as random effects using the MLM. The analyses of the data were performed in the same way as those in the simulation study.

RESULTS

Simulations

We simulated 2000 data sets and analyzed each using MLM, RDA, CCA, and NMDS to detect the effect of 5 environmental variables on community composition. We were primarily interested in assessing type I and type II errors, which are determined by the proportion of data sets in which "statistically significant" effects of environmental variables were identified at the 0.05 level (Fig. 1). From the design of the simulations, environmental variables x_1 and x_2 affect community composition, whereas x_3 , x_4 , and x_5 do not.

MLM had good power to detect the effects of x_1 and x_2 , while simultaneously failing to (falsely) identify effects of x_3 , x_4 , and x_5 . The power of the tests was slightly higher for the count data than the presence/absence data, reflecting the greater information available when abundance is included. The nominal rejection rates of the null hypotheses for environmental variables x_3 , x_4 , and x_5 were near 0.05, as they should be.

RDA performed well in identifying x_1 and x_2 as drivers of community composition, with similar power to MLM for x_1 but lower power for x_2 (Fig. 1). The lower power of RDA to detect x_2 relative to x_1 suggests that RDA is prone to confusion by collinearity (between x_2 and x_3). The nominal rejection rates for x_3 and x_5 were 5%, as they should be. However, x_4 was identified as having a strong effect on composition in many simulated data sets. Thus, RDA is very sensitive to the overall abundance of species, identifying communities as having different composition even if the relative abundanc-

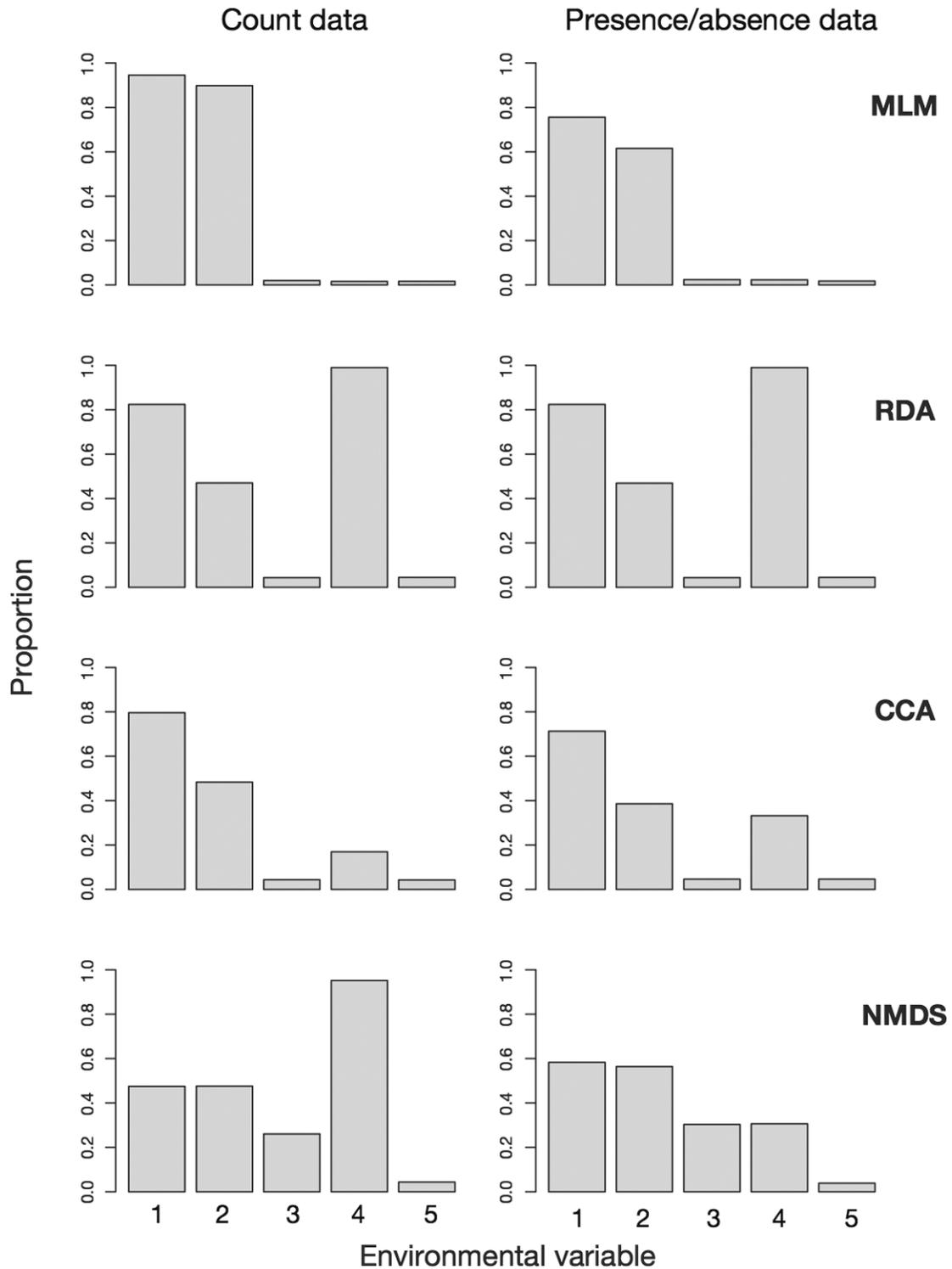


Fig. 1. Results from 2,000 simulations using multilevel modeling (MLM), redundancy analysis (RDA), canonical correspondence analysis (CCA), and non-metric multidimensional scaling (NMDS) for count (left panels) and presence/absence data (right panels). Bars show the proportion of data sets for which $P < 0.05$.

Table 1. Coefficients of fixed and random effects in the best (lowest AIC) multilevel model for data on Southern Appalachian understory herbs. Basal area of trees (m^2) are indicated by ACSA = *Acer saccharum* and LITU = *Liriodendron tulipifera*.

Variable	Fixed effects	Random effects
Intercept	0.65*	1.20
Elevation	0.83**	0.99***
Elevation ²	-0.26**	...
ACSA	0.35**	...
TreeBA (total tree basal area)	0.18†	...
% herb cover	0.16	0.15†
% herb cover ²	-0.32**	...
LITU	0.13	0.28**
Soil Ca	0.50†	0.61***
Soil Ca ²	-0.29**	...
Soil P	0.09	0.27**

† $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

es of species are the same. The performance of CCA was similar to RDA, although CCA had lower power in all cases. This likely occurs because in the simulations species abundances depended linearly on the environmental variables, and when this is the case RDA should be superior to CCA (Doledec et al. 2000).

NMDS generally had lower power than MLM and often than RDA and CCA. In addition to falsely identifying x_4 as a determinant of community composition, NMDS also frequently falsely identified x_3 . Even though x_3 did not affect the abundance of species, it was strongly correlated with x_2 , which did. We performed partial RDA and CCA analyses that are explicitly designed to factor out collinearity, while NMDS did not have this capability. Therefore, the susceptibility of NMDS to collinearity was expected.

Overall, statistical advantages of MLM include good statistical power and the ability to separate differences in community composition *sensu stricto* from differences due to overall abundances, such as those caused by x_4 . This does not mean, however, that MLM did not identify the effect of x_4 to increase the abundance of all species. In fact, the fixed coefficient β_4 for the shared effect of x_4 on all species was statistically significant in all 2000 simulations.

Analyses of forest herbs

Of the 14 focal herb species, the six most common species each occurred in >50% of plots,

with the least common occurring in 26% of the plots (Appendix: Table A1). In the MLM, the best-fitting model had no competing models with $\Delta AIC < 2$. The probability of observing any focal species in a plot was greatest at high values of elevation (but with a negative quadratic effect) and basal area of sugar maple (ACSA), with negative quadratic effects for herb cover and soil calcium (fixed effects; Table 1). The among-species variation in presence was strongly influenced by elevation, soil calcium, basal area of tulip poplar, and soil phosphorous (random effects; Table 1).

Application of RDA, CCA, and NMDS yielded similar, though not identical results. MLM, RDA, and CCA all identified elevation, soil calcium, basal area of tulip poplar, and soil phosphorous as statistically significant environmental variables determining community composition. RDA also identified basal area of sugar maple as marginally significant ($P = 0.06$). While NMDS identified elevation, soil calcium, and basal area of sugar maple, it did not identify basal area of tulip poplar and soil phosphorous as statistically significant environmental variables.

Biplots of the results reveal similar but not identical results in the distribution of communities in the first two ordination axes (Fig. 2). The most similar results were for RDA and CCA, with a procrustes dissimilarity score of 0.14, while these both differed from MLM with scores of 0.19. NMDS was the most dissimilar, with scores of 0.73, 0.71, and 0.63 from RDA, CCA, and MLM, respectively. That NMDS is the most dissimilar method parallels the result that it did not identify the same environmental variables as the other methods (Table 2). Finally, in the MLM, RDA, and CCA the effect of elevation was nearly orthogonal to the effects of basal area of tulip poplars (LITU) and soil calcium (Ca), indicating that these two sets of variables partition communities separately.

In addition to comparing communities, MLM simultaneously gave information at the species level in the form of regression coefficients of the effects of separate environmental variables on each species; environmental variables that have significant variability among species explain differences in composition among communities (Table 3). Elevation was the most important environmental variable for eight species, all of

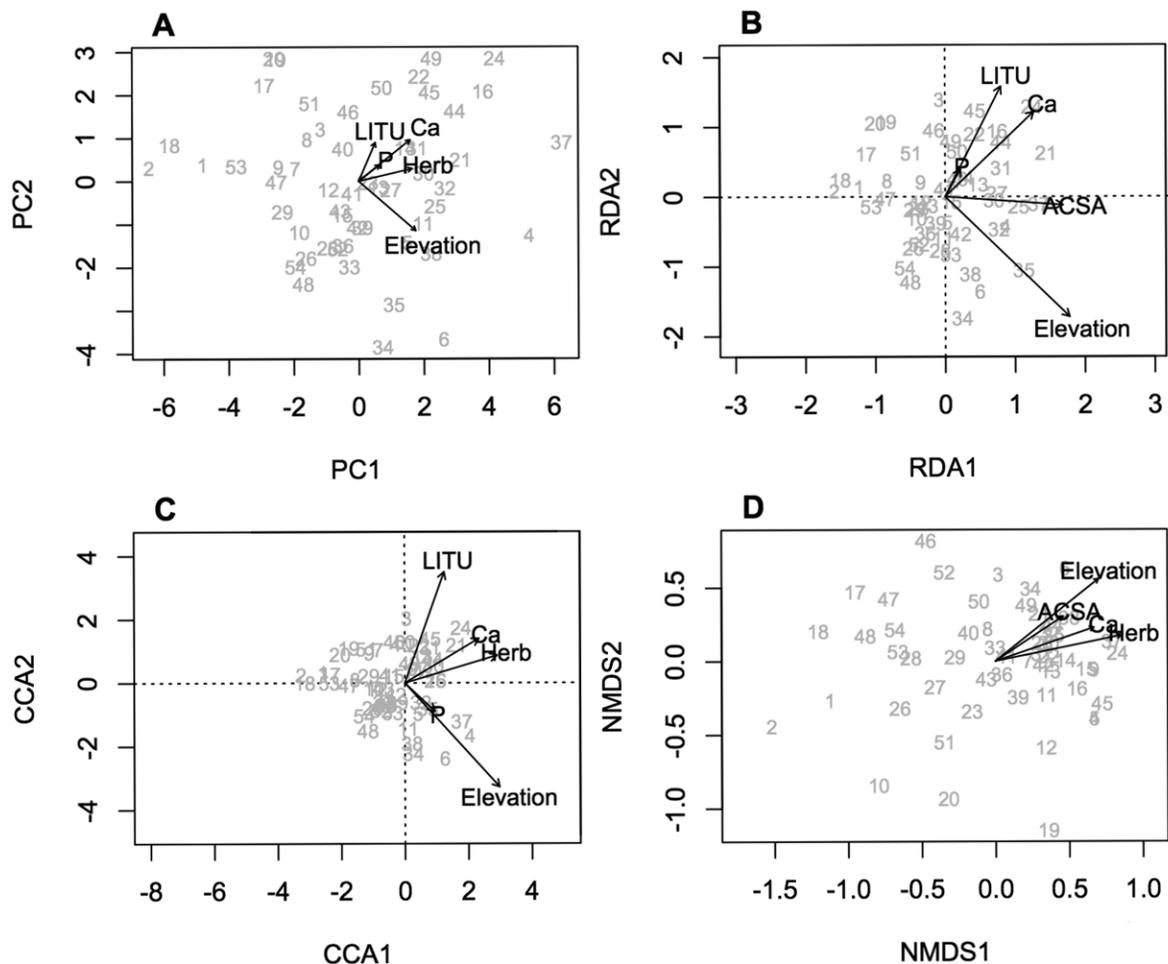


Fig. 2. Ordination plots of understory herb data using multilevel modeling (A), redundancy analysis (B), canonical correspondence analysis (C), and non-metric multidimensional scaling (D), with significant environmental vectors overlain. Elevation is the elevation of the plot, TreeBA is the basal area of all trees, ACSA is the total basal area of *Acer saccharum*, LITU is the total basal area of *Liriodendron tulipifera*, Ca is soil calcium, P is soil phosphorous, and herb is the average percent cover of all herbs.

which were associated with high elevation sites except for one, while soil calcium was the most important driver for five species; four species were positively associated with soil calcium and one was negatively associated. Basal area of tulip poplars and soil phosphorus were each the most important variables for one and two species, respectively.

DISCUSSION

We used simulations to test the statistical performance of MLM and analyses of a new

data set to compare results with three other common methods used to analyze community composition, RDA, CCA, and NMDS. Using simulated data, MLM had better power to detect environmental drivers (x_1 and x_2) of community composition than RDA, CCA, and NMDS. MLM was also less susceptible to loss of power caused by collinearity between environmental variances (x_2 and x_3); even using partial RDA and CCA, which removed the false identification of environmental variables (x_3), these methods had lower power to detect the collinear environmental driver (x_2). NMDS falsely identified collinear

Table 2. Significance of environmental variables for determining community composition using likelihood ratio tests for random effects in the best multilevel model (MLM) and 10,000 permutations for three ordination techniques: redundancy analysis (RDA), canonical correspondence analysis (CCA) and non-metric multidimensional scaling (NMDS).

Variable	MLM	RDA	CCA	NMDS
Elevation	***	***	***	***
Herb	†	†		
Ca	***	**	**	**
LITU	**	**	**	
Soil P	**	**	**	
ACSA		†		*
TreeBA (total tree basal area)				†

† $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

environmental variables as affecting community composition, even when they did not (x_3). Finally, MLM was the only method consistently distinguishing between environmental variables that affected all species in the same way (x_4) from those that affected all species differently (x_1 and x_2).

We believe it makes most sense to consider communities with the same relative abundances of species, or communities with the same relative probabilities of containing each species, as having the same composition. This makes a strict distinction between the number of species in a community and the identities of those species. MLM is the only method that made this distinction. For the analyses of abundance data

using RDA, CCA and NMDS, differences in abundances among sites can be first removed by dividing the abundances of each species in a site by the total abundance of species in that site; this should make RDA, CCA, and NMDS less sensitive to differences in mean abundances among sites when identifying differences among sites in community composition. However, this standardization of abundance data cannot be applied to presence/absence data. Furthermore, ordination methods that estimate species optima as weighted averages of plot scores (e.g., CA and DCA) ignore species relative abundance differences (Økland and Eilertsen 1996). Thus, the extent to which multivariate methods differentiate communities with the same relative abundances differs among methods.

The MLM had advantages over other methods when applied to the forest herb data set in its ability to disentangle drivers of individual species distributions, along with overall community trends. The MLM method is hierarchical; thus, variables can be nested within groups (species), requiring only one model to explain the distributions of many species within a framework of community composition. This contrasts the commonly used approach of performing separate logistic regressions on the presence/absence of the 14 species. MLMs can identify species with narrow environmental gradient ranges that may be especially sensitive to disturbance or changes in habitat suitability. For example, five species (*C. thalictroides*, *L.*

Table 3. MLM random effect coefficients by species of southern Appalachian understory herbs. Coefficients are the random effect plus the estimate for fixed effects in order to account for the mean slope. The most important predictor variable (largest effect) for each species is bold.

Species	Elevation	Herb	LITU	Ca	P
<i>Actaea racemosa</i>	0.59	0.51	0.11	1.24	0.13
<i>Arisaema triphyllum</i>	0.72	0.45	0.45	1.19	0.23
<i>Caulophyllum thalictroides</i>	1.07	0.44	0.16	0.68	0.44
<i>Chimaphila maculatum</i>	-1.11	-0.27	-0.95	-0.61	-0.16
<i>Galium latifolia</i>	0.38	-0.09	0.06	0.35	-0.69
<i>Geranium maculatum</i>	0.42	0.43	0.44	0.24	0.38
<i>Lilium superbum</i>	1.58	0.54	0.30	-0.03	0.39
<i>Polygonatum biflorum</i>	0.18	-0.06	-0.20	0.37	-0.59
<i>Prosartes lanuginosum</i>	2.25	0.08	0.08	0.60	0.01
<i>Sanguinaria canadensis</i>	0.04	0.27	0.68	0.94	-0.22
<i>Thalictrum dioicum</i>	0.53	-0.13	-0.13	0.74	0.38
<i>Trillium grandiflorum</i>	1.74	0.09	0.47	0.39	0.03
<i>Uvularia grandiflora</i>	1.92	0.07	-0.26	1.41	0.22
<i>Viola rotundifolia</i>	0.48	-0.03	0.34	-0.83	0.64

superbum, *P. lanuginosum*, *T. grandiflorum*, and *U. grandiflora*) were associated with high elevation sites using the MLM. These species may be particularly sensitive to climate warming and at risk of losing their southern ranges if high elevation habitats become unsuitable. Four of these species are ant-dispersed (all except *C. thalictroides*), furthering limiting their ability to adapt to shifting ranges because they cannot migrate very far or fast. *Trillium* species have also been shown to be sensitive to land-use history (Vellend 2004, Jenkins and Webster 2009), suggesting that even if areas of suitable climate exist, this group may be vulnerable if it cannot establish viable populations in secondary forests. Furthermore, basal area of tulip poplar, *L. tulipifera*, was important as a random effect in the MLM. In the southern Appalachians, tulip poplar dominates previously disturbed sites with mesic, fertile soils and generates a thin leaf litter layer (Kuhman et al. 2011). Most species were positively associated with tulip poplar, particularly *G. maculatum* and *S. canadensis*, possibly indicating a correlation with suitable edaphic conditions. When the spatial pattern in species composition of the herb layer is significantly correlated with that of the overstory layer, the two strata demonstrate the “linkage phenomenon” described by Gilliam (2007). While we do not explicitly test for such linkages, the importance of *L. tulipifera* in explaining both species and community patterns in our data set points towards the existence of this phenomenon.

The results from the MLM at the individual species level were consistent with the results from community analysis methods. Elevation was the most important correlate of community composition using all four methods (Table 2) and was also the environmental variable most frequently identified as most important for individual species (Table 3). Soil calcium was the second- or third-most important environmental variable across all community analysis methods and the second-most important at the species level; furthermore, the effects of basal area of tulip poplar were closely associated with the effects of soil calcium. In the community analyses the effects of elevation were largely orthogonal to those of soil calcium/basal area of tulip poplars (Fig. 2). Therefore, these environmental variables are influential for explaining the distribution of individual species and

hence the composition of communities across the environmental gradients we considered.

As essentially regression-based, our MLM approach requires decisions typical of multiple regression. Multivariate analyses often deal with large numbers of environmental predictor variables, and therefore it is often useful to exclude highly collinear variables or those that are known to have similar biological effects, as we did in our analyses of our forest herb data set. Similarly, inclusion of quadratic or interaction terms should be done in a biologically informed way so as not to build a statistical model that contains more coefficients than could sensibly be estimated from a data set. Thus, we recommend caution in applying our MLM approach to any data set, and use of appropriate diagnostics (Gelman and Hill 2007), rather than any blanket recommendations regarding possible problems with collinearity and model over specification.

We have only compared four possible approaches to understanding variation in community composition, and we recognize that other approaches may have different strengths and weaknesses. More advanced multivariate analyses likely outperform RDA, CCA, and NMDS. For example, co-inertia analyses such as outlying mean index (OMI; Doledec et al. 2000, Thuiller et al. 2004) have been shown to explain more variance than CCA or RDA. Other recent approaches such as random forest (Breiman 2001) or boosted regression trees (Elith et al. 2008) have been shown to provide better prediction than logistic regression for assessing individual species distributions. In particular, REEMtree, a regression-tree based linear mixed-effects model, is capable of building interactions between variables and has no a priori expectation for the response of species to environmental gradients (Sela and Simonoff 2012). Here, however, we have restricted our comparison with MLM to RDA, CCA and NMDS because they are standard, commonly used approaches.

We believe that MLM provides a useful alternative to commonly used methods for analyses of community composition, while at the same time giving species-level information to disentangle the drivers of species distributions and assemblages. Furthermore, MLM provides a straightforward analysis that produces information easily accessible to scientists, policy makers,

and resource managers. For instance, the effects of land-use history (e.g., logging and farming) on forest understory communities are of long-standing interest to ecologists (see Hermy and Verheyen 2007 for review). MLM could be used to disentangle the effects of such legacies on individual species versus community composition. Indeed, Wyatt and Silman (2010) found that individual abundance was greater in old growth forests than mature secondary forests, and that species composition differed significantly between the two. MLM would have the advantage of combining such analyses into a single model. The MLM approach also provides a flexible framework that can be further modified to perform different tasks. For example, it can be used to model phylogenetic correlations of relatedness among species (Ives and Helmus 2011), and it could be similarly modified to incorporate other types of correlations, such as spatial autocorrelation among sites. Finally, MLM can easily be integrated into a multi-model inference strategy, and model-averaged parameter estimates and estimates of unconditional sampling variances can be computed. Future research might develop MLMs to explore the community composition of other taxa across a range of scales and landscapes, and to assess changes in species assemblages as environmental conditions change over time. Thus, MLMs offer a powerful and efficient approach for understanding community assemblages.

ACKNOWLEDGMENTS

This study was funded by the Long-term Ecological Research Program of the National Science Foundation (DEB-0823293 Coweeta LTER), the Bunde Fund Research Grant from the Department of Zoology, University of Wisconsin-Madison, and NSF-DEB-0816613 to ARI. We thank Jaimie Little, Aki Masunaga, Adam Milch, and Bryan Moore for their field assistance, and the U.S. Forest Service for a permit to conduct research in the Pisgah National Forest. Tom Albright, Heather Lumpkin, Tammy Wilson, Wilfried Thuiller, and one anonymous reviewer provided valuable comments on the manuscript.

LITERATURE CITED

- Bates, D., M. Maechler and B. Dai. 2008. lme4: Linear mixed-effects models using Eigen and Eigen++ software. R package version 0.999375-28. <http://lme4.r-forge.r-project.org/>
- Bolstad, P. V., L. Swift, F. Collins, and J. Regnier. 1998. Measured and predicted air temperatures at basin to regional scales in the southern Appalachian mountains. *Agricultural and Forest Meteorology* 91:161–176.
- Braun, E. L. 1950. *Deciduous forests of eastern North America*. McGraw-Hill, New York, New York, USA.
- Braun-Blanquet, E. L. 1932. *Plant sociology: the study of plant communities*. McGraw-Hill, New York, New York, USA.
- Bray, J. R. and J. T. Curtis. 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs* 27:325–349.
- Breiman, L. 2001. Random forests. *Machine Learning* 45:5–32.
- Clark, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology* 18(1):117–143.
- Clements, F. E. 1916. *Plant succession: an analysis of the development of vegetation*. Publication 242. Carnegie Institute.
- Cushman, S. A. and K. McGarigal. 2004. Hierarchical analysis of forest bird species-environment relationships in the Oregon coast range. *Ecological Applications* 14:1090–1105.
- Doledec, S., D. Chessel, and C. Gimaret-Carpentier. 2000. Niche separation in community analysis: a new method. *Ecology* 81:2914–2927.
- Dormann, C. F. 2007. Promising the future? Global change projections of species distributions. *Basic and Applied Ecology* 8:387–397.
- Driscoll, D. A. and T. Weir. 2005. Beetle responses to habitat fragmentation depend on ecological traits, habitat condition, and remnant size. *Conservation Biology* 19:182–194.
- Duffy, M. A., C. E. Caceres, S. R. Hall, A. J. Tessier, and A. R. Ives. 2010. Temporal, spatial, and between-host comparisons of patterns of parasitism in lake zooplankton. *Ecology* 91:3322–3331.
- Ehrlén, J. and O. Eriksson. 2000. Dispersal limitation and patch occupancy in forest herbs. *Ecology* 81:1667–1674.
- Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151.
- Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77:802–813.
- Ford, W. M., R. H. Odom, P. E. Hale, and B. R. Chapman. 2000. Stand-age, stand characteristics, and landform effects on understory herbaceous communities in southern Appalachian cove-hardwoods. *Biological Conservation* 93:237–246.
- 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*. American Society of Agronomy and Soil Science Society of America, Madison, Wisconsin, USA.
- Gelman, A. and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York, New York, USA.
- Gilliam, F. S. 2007. The ecological significance of the herbaceous layer in temperate forest ecosystems. *BioScience* 57:845–858.
- Gleason, H. A. 1926. The individualistic concept of plant association. *Bulletin of the Torrey Botany Club* 53:7–26.
- Glenn-Lewin, D. C. 1977. Species diversity in North American temperate forests. *Vegetatio* 33:153–162.
- Graham, R. C. 1990. Soil-geomorphic relations on the Blue Ridge Front: soil characteristics and pedogenesis. *Soil Science Society of America* 54:1367–1377.
- Guisan, A. and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8:993–1009.
- Guisan, A. and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147–186.
- Hermly, M. and K. Verheyen. 2007. Legacies of the past in the present-day forest biodiversity: a review of past land-use effects on forest plant species composition and diversity. *Ecological Research* 22:361–371.
- Hottola, J. and J. Siitonen. 2008. Significance of woodland key habitats for polypore diversity and red-listed species in boreal forests. *Biodiversity Conservation* 17:2559–2577.
- Huisman, J., H. Olff, and L. S. M. Fresco. 1993. A hierarchical set of models for species response analysis. *Journal of Vegetation Science* 4(1):37–46.
- Ibanez, I., J. S. Clark, and M. C. Dietz. 2008. Evaluating the sources of potential migrant species: implications under climate change. *Ecological Applications* 18:1664–1678.
- Ives, A. R. and M. Helmus. 2011. Generalized linear mixed models for phylogenetic analyses of community structure. *Ecological Monographs* 81(3):511–525.
- Jenkins, M. A. and C. R. Webster. 2009. Spatial patterning and population structure of a common woodland herb, *Trillium erectum*, in primary and post-logging secondary forests. *Forest Ecology and Management* 258:2569–2577.
- Kuhman, T. R., S. M. Pearson, and M. G. Turner. 2011. Agricultural land-use history increases non-native plant invasion in a Southern Appalachian forest a century after abandonment. *Canadian Journal of Forest Research* 41(5):920–929.
- Legendre, P. and L. Legendre. 1998. *Numerical ecology*. Elsevier Science, Amsterdam, The Netherlands.
- McCune, B. and J. B. Grace. 2002. *Analysis of ecological communities*. MjM Software Design, Gleneden Beach, Oregon, USA.
- McNab, W. H. 1993. A topographic index to quantify the effects of mesoscale landform on site productivity. *Canadian Journal of Forest Research* 23:1100–1107.
- McNab, W. H. 1996. Classification of local- and landscape-scale ecological types in the southern Appalachian Mountains. *Environmental Monitoring and Assessment* 39:215–229.
- Mearns, L. O., F. Giorgi, L. McDaniel, and C. Shields. 2003. Climate scenarios for the southeastern US based on GCM and regional model simulations. *Climatic Change* 60:7–35.
- Økland, R., and O. Eilertsen. 1996. Dynamics of understory vegetation in an old-growth boreal coniferous forest, 1988–1993. *Journal of Vegetation Science* 7(5):747–762.
- Oksanen, J., R. Kindt, and R. B. O’Hara. 2011. *vegan: community ecology package*. Version 2.0-2. <http://CRAN.R-project.org/package=vegan>
- Oksanen, J. and P. Minchin. 2002. Continuum theory revisited: What shape are species responses along ecological gradients? *Ecological Modelling* 157:119–129.
- Palmer, M. W. 1993. Putting things in even better order: The advantages of canonical correspondence analysis. *Ecology* 74:2215–1130.
- Pearce, J. L., M. A. Burgman, and D. C. Franklin. 1994. Habitat selection by helmeted honeyeaters. *Wildlife Resources* 21:53–63.
- Qian, S. S., T. F. Cuffney, I. Alameddine, G. McMahon, and K. H. Reckhow. 2010. On the application of multilevel modeling in environmental and ecological studies. *Ecology* 91:355.
- Qian, S. S. and Z. Shen. 2007. Ecological applications of multilevel analysis of variance. *Ecology* 88:2489–2495.
- R Development Core Team. 2011. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rydgren, K., R. Økland, and T. Økland. 2003. Species response curves along environmental gradients. A case study from SE Norwegian swamp forests. *Journal of Vegetation Science* 14:869–880.
- Sela, R. J. and J. S. Simonoff. 2012. RE-EM trees: a data mining approach for longitudinal and clustered data. *Machine Learning* 86:169–207.
- Self, S. G. and K. Y. Liang. 1987. Asymptotic properties of maximum-likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* 82:605–610.
- Southeast Regional Climate Center. 2007. Climate data, Station 310301. <http://www.sercc.com/>
- Stearns, F. W. 1949. Ninety years of change in a northern hardwood forest in Wisconsin. *Ecology*

- 30:350–358.
- Stram, D. O. and J. W. Lee. 1994. Variance-components testing in the longitudinal mixed effects model. *Biometrics* 50:1171–1177.
- ter Braak, C. J. F. and I. C. Prentice. 1988. A theory of gradient analysis. *Advances in Ecological Research* 18:271:317.
- Thuiller, W., S. Lavorel, G. Midgley, S. Lavergne, and T. Rebelo. 2004. Relating plant traits and species distributions along bioclimatic gradients for 88 *Leucadendron* taxa. *Ecology* 85:1688–1699.
- Trivedi, M. R., M. D. Morecroft, P. M. Berry, and T. P. Dawson. 2008. Potential effects of climate change on plant communities in three montane nature reserves in Scotland, UK. *Biological Conservation* 141(6):1665–1675.
- Turner, M. G., W. H. Romme, R. A. Reed, and G. A. Tuskan. 2003. Postfire aspen seedling recruitment across the Yellowstone (USA) landscape. *Landscape Ecology* 18:127–140.
- Turner, M. G., S. E. Gergel, and M. D. Dixon. 2004. Distribution and abundance of trees in floodplain forests of the Wisconsin River: environmental influences at different scales. *Journal of Vegetation Science* 15:729–738.
- Vellend, M. 2004. Parallel effects of land-use history on species diversity and genetic diversity of forest herbs. *Ecology* 85:3043–3055.
- Whittaker, R. H. 1956. *Vegetation of the Great Smoky Mountains*. *Ecological Monographs* 26:1–80.
- Wyatt, J. L. and M. R. Silman. 2010. Centuries-old logging legacy on spatial and temporal patterns in understory herb communities. *Forest Ecology and Management* 260:116–124.

SUPPLEMENTAL MATERIAL

APPENDIX

Study Area

This field data was collected in the French Broad River Basin in Buncombe and Madison Counties of western North Carolina (Fig. A1).

The study region lies within the Southern Blue Ridge physiographic province, which is characterized by steep terrain and mixed-mesophytic deciduous forest (Braun 1950). This area receives

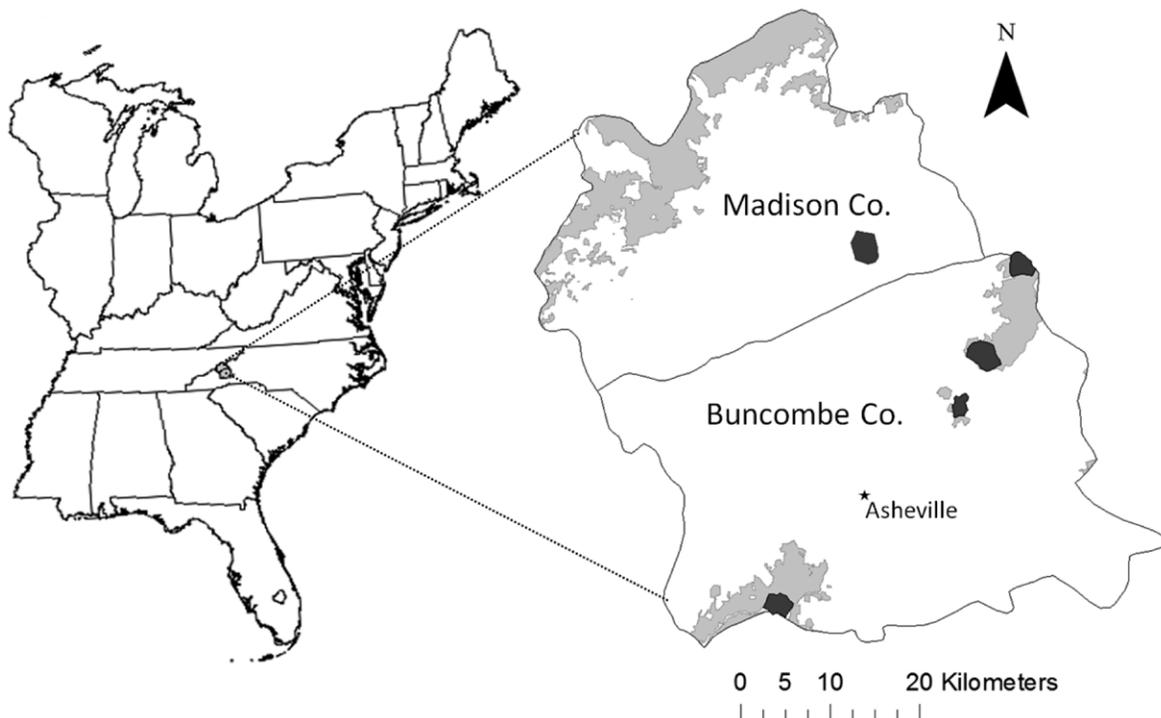


Fig. A1. Map of the study region. The enlarged area shows the two counties where the fieldwork was conducted. Light gray shading indicates Pisgah National Forest. Dark gray shading indicates the 5 watersheds in which the study plots ($n = 54$) are located.

Table A1. Focal native herbaceous species.

Species	Common name	Family	Presence in plots (%)
<i>Actaea racemosa</i>	Black cohosh	Ranunculaceae	51.9
<i>Arisaema triphyllum</i>	Jack in the pulpit	Araceae	79.6
<i>Caulophyllum thalictroides</i>	Blue cohosh	Berberidaceae	46.3
<i>Chimaphila maculatum</i>	Spotted wintergreen	Ericaceae	38.9
<i>Galium latifolium</i>	Purple bedstraw	Rubiaceae	75.9
<i>Geranium maculatum</i>	Wild geranium	Geraniaceae	64.8
<i>Lilium superbum</i>	Turk's cap lily	Lilaceae	42.6
<i>Polygonatum biflorum</i>	Solomon's seal	Lilaceae	29.6
<i>Prosartes lanuginosum</i>	Yellow mandarin	Lilaceae	88.9
<i>Sanguinaria canadensis</i>	Bloodroot	Papaveraceae	61.1
<i>Thalictrum dioicum</i>	Early meadow rue	Ranunculaceae	38.9
<i>Trillium grandiflorum</i>	White trillium	Lilaceae	50.0
<i>Uvularia grandiflora</i>	Large-flowered bellwort	Lilaceae	33.3
<i>Viola rotundifolia</i>	Roundleaf violet	Violaceae	25.9

on average 125 cm of precipitation per year. Mean winter temperature is 4°C and mean summer temperature is 23°C (Southeast Regional Climate Center 2007). Elevation ranges from 350 to 1900 m. Overstory canopies are dominated by *Quercus velutina*, *Q. prinus*, *Q. alba*, *Oxydendron arboreum*, and *Pinus strobus* on upper slopes and ridges, and by *Liriodendron tulipifera*, *Acer saccharum*, *A. rubrum*, *Fagus grandifolia*, *Q. rubra*, *Tilia americana*, and *Betula lenta* on more mesic sites (McNab 1996). Soils consist of Ultisols on gradual slopes and ridges, and Inseptisols on steeper slopes and coves (Graham 1990). The biological diversity of herbaceous communities is particularly rich, especially in mesic cove forests (Whittaker 1956, Glenn-Lewin 1977, Ford et al. 2000). The topographic variability that characterizes the region creates strong gradients of temperature and moisture (Bolstad et al. 1998) that can be used as proxies for climate (Ibanez et al. 2008, Trivedi et al. 2008).

Field Sampling

Prior to sampling, we identified 14 native herbaceous species that represented a range of habitat generalists and specialists, and were common to the area (Table A1). Sampling took place from May–August 2009. We selected sites by generating random points across 5 watersheds in ArcGIS and choosing a subset of 54 sites that were accessible, not recently logged, and not dominated by Eastern hemlock (*Tsuga canadensis*), rhododendron (*Rhododendron* spp.), or mountain laurel (*Kalmia latifolia*).

A 20 × 20 m plot was established at each site containing 3 transects at 0 m, 10 m, and 20 m, perpendicular to the slope. Using a modified

version of the Braun Blanquet (1932) scale, we recorded percent cover for each focal species within 1-m² quadrats at 6 evenly spaced intervals along each transect (18 quadrats per plot). The entire plot was thoroughly searched and the presence of any focal species not recorded in a quadrat was noted and assigned to the lowest cover class (0–0.1%). Within each quadrat, we also measured percent total herb and shrub cover, and litter depth (an average of 4 measurements). Slope and terrain shape were determined using a clinometer. A terrain shape index (TSI) estimates the concavity or convexity of the local landform by averaging slope gradients from the plot center to its edge in the 8 sub-cardinal directions (McNab 1993). A positive TSI indicates a concave landform, whereas a negative TSI indicates a convex landform. We extracted elevation and distance to stream for each plot using 30-m digital elevation models (DEMs) in ArcGIS. Distance to stream represents relative slope position and was obtained by calculating flow accumulation and the distance from plots to the nearest point of highest accumulation, using a threshold of 100 pixels. High values indicate positions upslope near ridges and low values indicate positions downslope close to streams or coves.

Five soil cores (5 cm diameter, 15 cm depth) were obtained from each plot, composited, sieved (4 mm) and oven-dried. Soils were analyzed for total N, organic matter, P, K⁺, Ca⁺⁺, Mg⁺⁺, and pH by the University of Wisconsin-Madison Soil and Plant Analysis Laboratory using standard methods. Soil texture was determined in the laboratory using the hydrometer method (Gee and Bauder 1986).

Table A2. Description of the environmental variables used in this study, including observed range, mean, and 1 standard deviation ($n = 54$ plots).

Category	Variable	Description	Observed range	Mean	SD
Topographic	Elevation	Elevation a.s.l. (m)	683–1,558	1,086	210.4
	Aspect	Degrees converted to moisture index	1–16	7.8	3.9
	Slope	Slope gradient (%)	8–30	22.8	9.1
	Terrain	Local terrain shape index (convex < 0 < concave)	–66–55	7.0	22.2
	Distance to stream	Measure of slope position relative to nearest stream (m)	0–340	108.3	87.2
Biotic	TreeBA	Relative basal area of trees >10 cm DBH (m ²)	4,645–30,342	15,762	6,741
	ACSA	Relative basal area of <i>Acer saccharum</i> (m ²)	0–0.49	0.08	0.14
	LITU	Relative basal area of <i>Liriodendron tulipifera</i> (m ²)	0–0.79	0.22	0.25
	QUPR	Relative basal area of <i>Quercus prinus</i> (m ²)	0–0.58	0.09	0.16
	Shrub	Average total shrub cover (0.5–3 m above ground)	1–40	16.4	8.5
	Herb	Average total herb cover (<0.5 m above ground)	7–79	35.0	17.5
Soils and litter	Litter	Average litter depth (cm)	0.7–3.5	1.9	0.5
	Sand	Amount of sand in soil (%)	35–64	0.5	0.1
	pH	Soil pH	4.2–6.1	5.0	0.4
	OM	Soil organic matter (%)	4.9–32.8	12.5	5.6
	Ca	Soil calcium (ppm)	28–1,332	399.9	323.8
	Mg	Soil magnesium (ppm)	18–287	92.5	50.5
	P	Soil phosphorous (ppm)	2–16	4.6	2.3
	K	Soil potassium (ppm)	35–155	91.0	26.9
	N	Soil total nitrogen (%)	0.1–1.13	0.4	0.2

SUPPLEMENT

R script and input data file for creating multilevel models using the forest herb data set described in the main text (*Ecological Archives* C003-010-S1).