

COMMUNITY INFERENCE USING HIERARCHICAL MULTI-SPECIES LOGISTIC REGRESSION MODELS

GREG ANDERSON, BRANDON PEOPLES, AND CRAIG ROGHAIR

INTRODUCTION

Species distributional models (SDMs) are powerful tools for making inferences on relationships between species and their environments. SDMs provide a framework to predict the response of a species to environmental covariates, and to predict the occurrence or abundance of a species throughout a landscape. Most SDMs have been developed for individual species (Ferrier and Guisan 2006), and perform best when species are relatively common, well-sampled, and have a relatively large geographic range (Fitzpatrick et al. 2011). Unfortunately, the species of greatest interest to many biologists are often those that have received little research attention or sampling effort, or have restricted distributions (i.e. they are ‘data poor’). Recent community-level modeling approaches show promise for use with these data poor species (Dorazio and Royle 2005, Dorazio et al. 2006, Ferrier and Guisan 2006, and Dorazio et al. 2010).

For many taxa, multiple species are encountered when sampling for a species of interest (e.g., fishes, birds, macroinvertebrates, etc.). However, standard SDMs ignore the additional information that non-focal species provide, and cannot predict occurrences within the context of biotic communities. This information is in the form of shared exposure to environmental conditions. When species are syntopic, they are exposed to similar environments and interactions, both biotic and abiotic (i.e., the species are not strictly independent). Although multiple species are seldom included in occupancy models, recent advances within the occupancy estimation literature have provided a framework to extend SDMs to multiple species using hierarchical models (Dorazio and Royle 2005, Dorazio et al. 2006; Kery and Royle 2008, 2009; DeWan and Zipkin et al. 2010). In contrast to monitoring species individually, the multi-species approach allows for the “borrowing” of information from similar species to improve model estimates, especially for rare or ‘data poor’ species (DeWan and Zipkin et al. 2010). In addition, this approach provides a better understanding of the aggregated response of groups of species or the entire community to changes over a landscape. This approach is based on the use of hierarchical models parameterized with random components representing individual species level parameters that belong to hyperdistributions with a mean and variance that can be estimated from the entire community.

In this paper, we used a dataset of fluvial fish occurrences to investigate the strengths and limitations associated with both the single- and multiple-species SDM approaches. These data are drawn from a typical dataset collected within a single-season of sampling 40 stream sites. Our specific objectives were to: 1) develop a typical single species SDM, 2) develop a community level SDM, 3) investigate the effect of additional grouping terms (e.g. reproductive guild, family, parental care) on the community level model, and 4) compare the benefits and limitations of each approach.

METHODS

STUDY AREA AND SAMPLING METHODS

A more detailed description of study area, sites, field sampling, and remote sensing methods can be found in Peoples et al. (2011). We modeled fish community data collected in 7 tributaries (2nd to 4th Strahler-order) to the middle New River, Virginia. Four of these streams (Toms, Stroubles, Crab, and Peak creeks) are located in the Ridge and Valley physiographic provinces, and three (Chestnut, Crooked, and Little Reed Island creeks) lie in the Blue Ridge province. Forty sites were selected to represent the range of anthropogenic and longitudinal habitat diversity found in the basin (Figure 1).

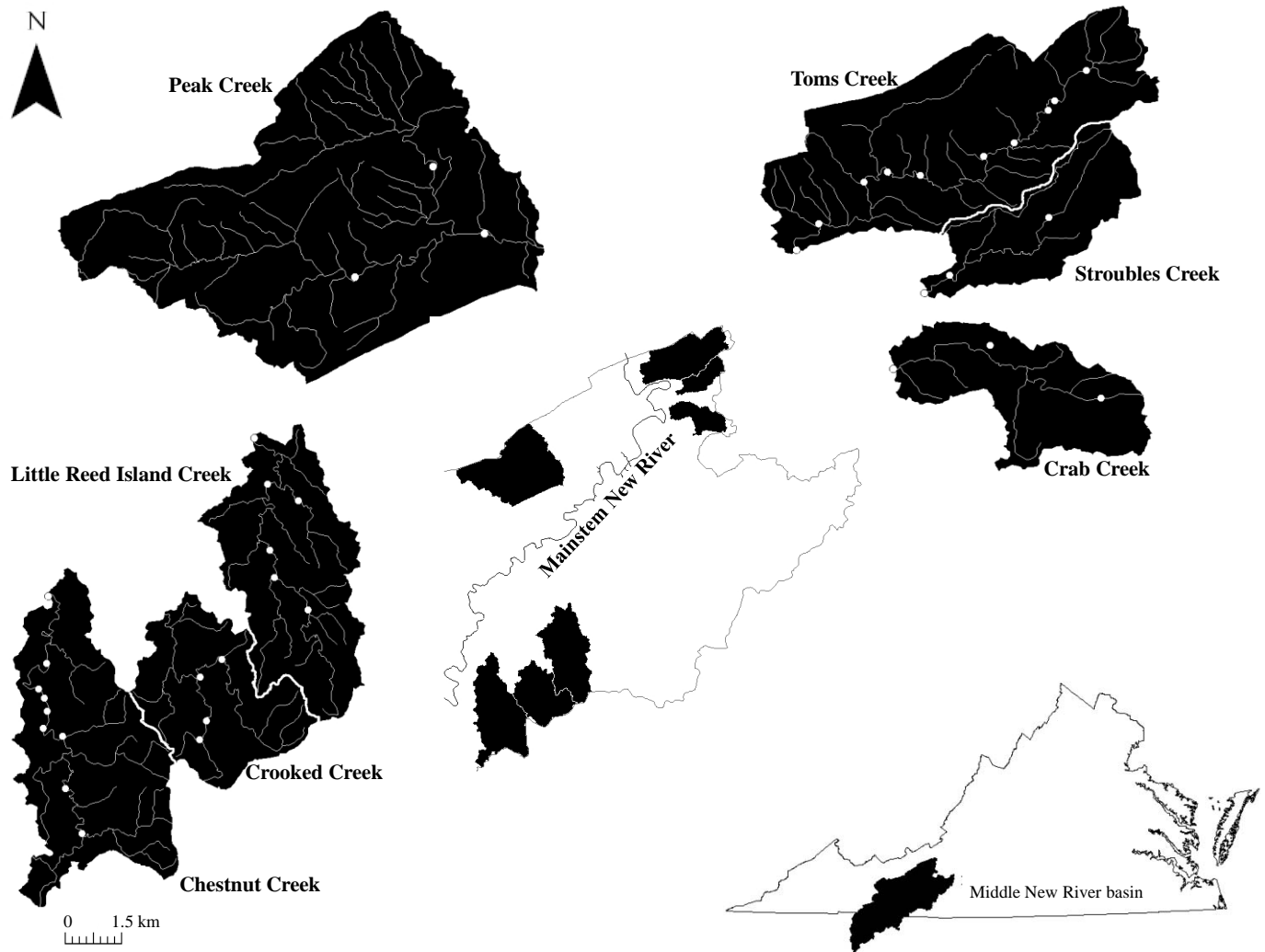


FIGURE 1. Map of sampling sites in the Middle New River basin.

Fishes were collected in July and August 2008 using single pass electrofishing with two Smith-Root LR-24 backpack electrofisher units. Reach lengths ranged from 180 to 240 m (20-30 times mean stream width). Instream habitat was quantified in early July of 2008 using methods based on the United States Environmental Protection Agency's Rapid Bioassessment Protocol (Barbour et al 1999). Instream habitat was quantified along 20 transects, spaced seven meters apart, at each site. Along each transect, mesohabitat type (pool, riffle, or run) was determined. Substrate embeddedness (hereafter, "embeddedness") was visually estimated at three equidistant points along each transect. Percentages of pool habitat (as a proxy for mesohabitat diversity) at each site were calculated by dividing the number of transects that occurred in pools by the total number of transects at each site. Embeddedness was averaged among all transects at each site.

We estimated each site's watershed area and landuse attributes in ArcGIS 9.2. Percentages of forest cover in each site's watershed were estimated using land cover maps in the National Land Cover Database. Prior to fitting models, watershed areas were log-transformed to achieve approximate distributional normality.

SINGLE-SPECIES SDM APPROACH

For binary data of presences and absences, generalized linear models (GLMs) provide a relatively strong framework to analyze species distribution models, especially for smaller datasets (Franklin 2009). In particular, logistic regression allows the modeling of presence/absence data by treating each observation (e.g., sites) as a Bernoulli trial with probability p_i , where i indexes the sites. This probability is then linked to the linear model with regressors using the logit link function.

To evaluate the single-species approach, we used the Appalachia darter *Percina gymnocephala* as a case species to create a SDM using logistic regression. This species is a benthic darter (Percidae) endemic to the New River drainage, and relatively little is known of its life history (Jenkins and Burkead 1994). Appalachia darter was moderately uncommon within our dataset, with a prevalence of 22.5% of the sites.

From our dataset of local habitat and remotely sensed features, we developed a set of six environmental covariates hypothesized to influence the presence of the Appalachia darter across the landscape: watershed area, watershed area², a binary indicator of whether the site was in the Blue Ridge (versus the Ridge and Valley; hereafter, we term this variable “physiographic province”), percent forest cover in the watershed, percent pools in the site, and average embeddedness. Given the modest sample size, we were unable to account for spatial autocorrelation (SAC) using spatial filtering methods. However, in an effort to deal with the nestedness of sites within tributaries, we modeled the tributary effect as a random effect within a generalized linear mixed model. All models were executed in R version 2.15.0 (R Development Core Team 2012) using the lme4 package (Bates et al. 2011). In an attempt to deal with possible multicollinearity, all possible combinations of the covariates were run with the exception that the quadratic term of watershed area² was not included unless watershed area was in the model. This approach resulted in a total of 48 candidate models. Model selection was performed using Akaike’s information criterion (AIC; Akaike 1973) as corrected for small sample size (AICc; Hurvich and Tsai 1989) using the R package AICmodavg (Marc 2012). This approach does not rely on the value of AICc per se. Instead, best models are selected based on differences between the value for the model of interest and that of the best-supported model (Δ_i ; Burnham and Anderson 2002). Based on the relative differences in AICc values for the fitted models, the weight of the evidence (w_i) for any particular model being the best-supported one in the set can be calculated (Akaike 1978). The evidence can then be judged by the ratio of Akaike weights (w_1/w_j), where 1 represents the best supported model and j the model under consideration. To determine the relative importance of a given parameter (θ) in the candidate set of models, the Akaike weights can be summed for all of the models in which the parameter is used ($\sum w_i$, Burnham and Anderson 2002).

To verify the model, we calculated the threshold independent metric area under the curve (AUC) of the receiver-operating characteristic (ROC) plot (Hanley & McNeil 1982). The ROC plot is a graph of the false-positive error rate versus the true positive rate based on all possible values of threshold probability. The AUC is then calculated by summing the area under the ROC curve (Franklin 2009). Although AUC should be judged for validation on out-of-sample sites or withheld sites (cross-validation), we were unable to do this type of validation due to small sample size.

MULTIPLE-SPECIES APPROACH

Using the full set of 27 species collected across the 40 sites in the New River drainage (Appendix A), we modeled the probability of presence using a multi-species logistic regression, where the probability of species j being located at site i was a Bernoulli random variable. Specifically, we modeled the probability that a species was present using the following set of equations:

$$\Pr(Y_{ij} = 1) = \text{logit}^{-1}(\alpha_j + \beta_{jk} * X_{ik}), \text{ for } i = 1, \dots, n. \text{ site}$$

$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2), \text{ for } j = 1, \dots, n. \text{ species}$$

$$\beta_{jk} \sim N(0, \Sigma), \text{ for } k = 1, \dots, n. \text{ parameters}$$

$$\Sigma = \begin{pmatrix} \sigma_{\beta_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{\beta_2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{\beta_K}^2 \end{pmatrix}$$

where α_j is the intercept for species j , β_{jk} is the regression coefficient for the covariate k and species j , X_{ik} is the value of covariate k at site i , and σ_θ^2 is the variance among the species for parameter θ .

Before the analysis, we visually tested for spatial autocorrelation using a semivariogram with hydrologic distances between pairs of sites. Upon inspection of the semivariogram, we detected a complex SAC pattern that resembled a sine curve (Figure 2). To better account for SAC, we used a spatial filtering approach using eigenfunction analysis from principle coordinates of neighborhood matrices based on Euclidian distance using the R packages `spdep` and `spacemaker`. To determine which sites should be considered neighbors, we increased the neighborhood distance by intervals of 2 kilometers until each site had at least one neighbor (6km). Then, beginning with 6km, we increased the distance at which sites should be considered neighbors by 2 kilometers until we reached 12 km. With each neighborhood definition, we extracted the positive eigenvectors and performed a permutation test to determine if Moran's I was significant for the eigenvector ($\alpha=0.05$). Using only significant eigenvectors, we fit a model with no other covariates except for the eigenvectors to estimate the weight of evidence for the neighborhood matrix using AICc. The model with the lowest AICc value was then used for spatial filtering for all candidate models.

Similar to the single-species approach, we fitted all possible combinations of the six covariates described above, where watershed area² was included in the model only if watershed area was. Each model was fit in R using package `lme4`. All regression coefficients were specified without correlation between the terms, and model support was judged using the approach described above. To verify models for each species, we used the estimated effect for each species and used the threshold independent metric AUC based on the ROC.

Using the global model (i.e., all covariates), we estimated the weight of evidence of including grouping variables as multi-level parameters to predict how species respond to different covariates. Three different types of grouping variables were assessed: (1) reproductive guilds, (2) families, and (3) whether or not adults provide parental care to broods (Appendix A). All possible combinations of these grouping variables were assessed by adding them to the global model. The weight of evidence was then judged by using AICc and comparing the model to one without any grouping terms.

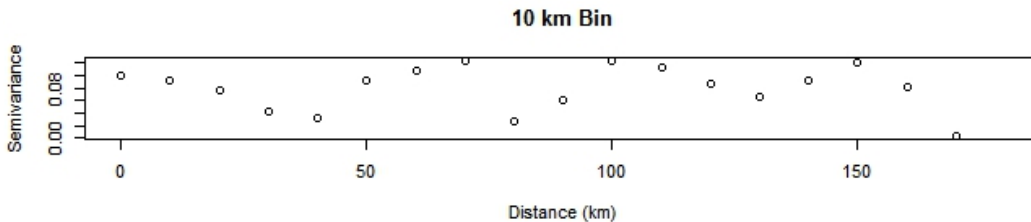


FIGURE 2. Example semivariogram for the Appalachia darter. All species in the dataset showed similar trends.

TABLE 1. Log likelihood (LN \mathcal{L}), number of parameters (K), AICc, delta AICc (Δ_i), and model weight (w_i) for best supported models ($\Delta_i < 2$) for the single-species analysis of the Appalachia darter.

Model	LN \mathcal{L}	K	AICc	Δ_i	w_i
Ln WS area	-14.75	3	36.17	0	0.19
Ln WS area, % Forest	-14.37	4	37.89	1.78	0.08
Ln WS area, % Pool	-14.44	4	38.02	1.85	0.07
Null Model	-20.78	2	45.88	9.70	0.001

RESULTS

SINGLE-SPECIES SDM APPROACH

A large amount of model uncertainty accompanied the single-species approach (Table 1). The best model included only watershed area. Secondary models also included this parameter, along with percent forest, percent pool, average embeddedness and physiographic province. However, the addition of these other parameters did not substantially improve the log likelihood (i.e., a change of less than two). The top model was 128 times better supported than the null model (i.e., model with no covariates). Using the entire candidate set of models fitted, the most important variable was watershed area, carrying a weight of 0.99; this parameter was found within all models with a ΔAICc of 9 or less. The other regression coefficients carried much less weight (between 0.25 and 0.37; Figure 3). Based on our top model, we expect the probability of Appalachia darter presence to increase as watershed area increases. Based on the ROC plot, we estimated AUC based on within-sample prediction to be 0.85 when the tributary system was not accounted for and 0.97 when the tributary system was included.

MULTIPLE-SPECIES APPROACH

No grouping term improved estimates of species presence. Therefore, all covariate models were fit without any grouping term. Similar to the single species approach, there was a large amount of model uncertainty within the multiple species modeling approach. No model carried a substantially large amount of weight, but the top model weight was much better supported than others. Our best supported model ($w_i=0.2$) included watershed area, watershed area², percent forest, percent pool, and average embeddedness. The top model was 1.83 times better supported than a model that not include all of these parameters (i.e., percent pool and percent forest; Table 2), and over 124 million times better supported than the null model. Unlike the single species approach, the log likelihood changed substantially between closely weighted models. Because the addition of other covariates required the estimation of two more parameters (i.e., mean and variance among species), models with more covariates were heavily penalized (receiving higher AICc values). Variance parameters were not estimated for percent forest or percent pool, probably because of multicollinearity among spatial eigenvectors (i.e., the gradient observed is better explained by spatial factors than changes in forest cover; VIF=3.6). Among the entire set of candidate models, the most important variable (i.e., highest

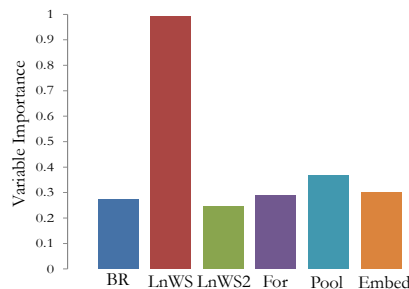


FIGURE 3. Variable importance for the single species logistic regression used for the Appalachia darter. Variable importance was calculated by summing the model weights for each model in which the parameter was found.

TABLE 2. Log likelihood (LN \mathcal{L}), number of parameters (K), AICc, delta AICc (Δ_i), and model weight (w_i) for best supported models ($\Delta_i < 2$) for the multiple species analysis.

Model	LN \mathcal{L}	K	AICc	Δ_i	w_i
LnWS, LnWS ² , % Forest, % Pool, Avg. Embed	-477.62	28	1012.8	0	0.20
LnWS, LnWS ² , Avg Embed	-482.43	24	1014	1.22	0.11
LnWS, LnWS ² , % Forest, Avg. Embed	-480.49	26	1014.3	1.53	0.09
LnWS, LnWS ² , % Pool, Avg. Embed	-480.51	26	1014.4	1.58	0.09
LnWS, LnWS ²	-484.83	22	1014.6	1.83	0.08
Null	-506.71	18	1050.1	37.28	0

variable importance) was watershed area (variable importance = 0.99). The second most important variable was watershed area² (0.83), then average embeddedness (0.71), followed by percent forest (0.60), percent pool (0.46), and physiographic province (0.19). Based on the estimated effects from the top model, we would expect these relationships to differ by species. However, the estimates of percent forest and percent pool did not have a variance component (but see above). The estimates for the other parameters varied substantially between species (Figure 4). The best supported model predicts species richness to increase with watershed area (Figures 4), but that community composition should change with increasing watershed size.

In addition to watershed area, average embeddedness of the substrate also had a high variable importance. Similar to watershed area, the effect differed by species, and we would expect that the number of species present to accumulate as average embeddedness increased (black line in Figure 5b). Generally, the more cosmopolitan species aligned on the positive side of the estimated effects (e.g., Centrarchids; Figure 4c); however, species groups that should be adversely affected by average embeddedness (e.g., simple lithophils, Table 2) showed no ascertainable responses to this variable.

Based on our top model, we estimated within sample AUC to be relatively high for all species (i.e., >0.8; Figure 6). However, the spatial effects were not included in the prediction of species presences at sites because the AUC was highly variable between species. This was especially true for the species that were found in only a few sites (e.g., cutlips minnow *Exoglossum maxillingua* only occurred in 12% of sites), or within a single tributary system (e.g., greenside darter *Ethostoma blennioides* was only found in Crooked Creek).

DISCUSSION

This study demonstrates the pervasive influence of longitudinal habitat heterogeneity on species' presence within communities. In general, stream fish species richness increases asymptotically with watershed area (Beecher et al. 1988, Matthews and Robison 1998), and community composition changes as headwater specialists are replaced by species that prefer larger habitats (Paller 1994). This pattern is largely attributable to the ubiquitous positive relationship between watershed size and habitat volume, diversity, and complexity (Gorman and Karr 1978, Downes et al. 1998). The high relative importance of watershed area as a predictor variable in our models, in addition to the relative balance of species-wise positive and negative effects of watershed area, corroborate the prevailing knowledge.

The relative importance of embeddedness and percent forest on approximating species occurrences most likely reflects response to anthropogenically induced land cover change. Most fishes upland North American streams evolved under forested conditions. Conversion of forests to agricultural or residential land use types can alter the physical, chemical, and biological processes that such fishes are adapted to (Hynes 1975; Allan 2004). As a result, changes (whether positive or negative) in species distributions often accompany such perturbations. Specifically, watershed-scale reductions in forested land cover increases fine sediment input to streams (Allan 1995). Increased sediment input

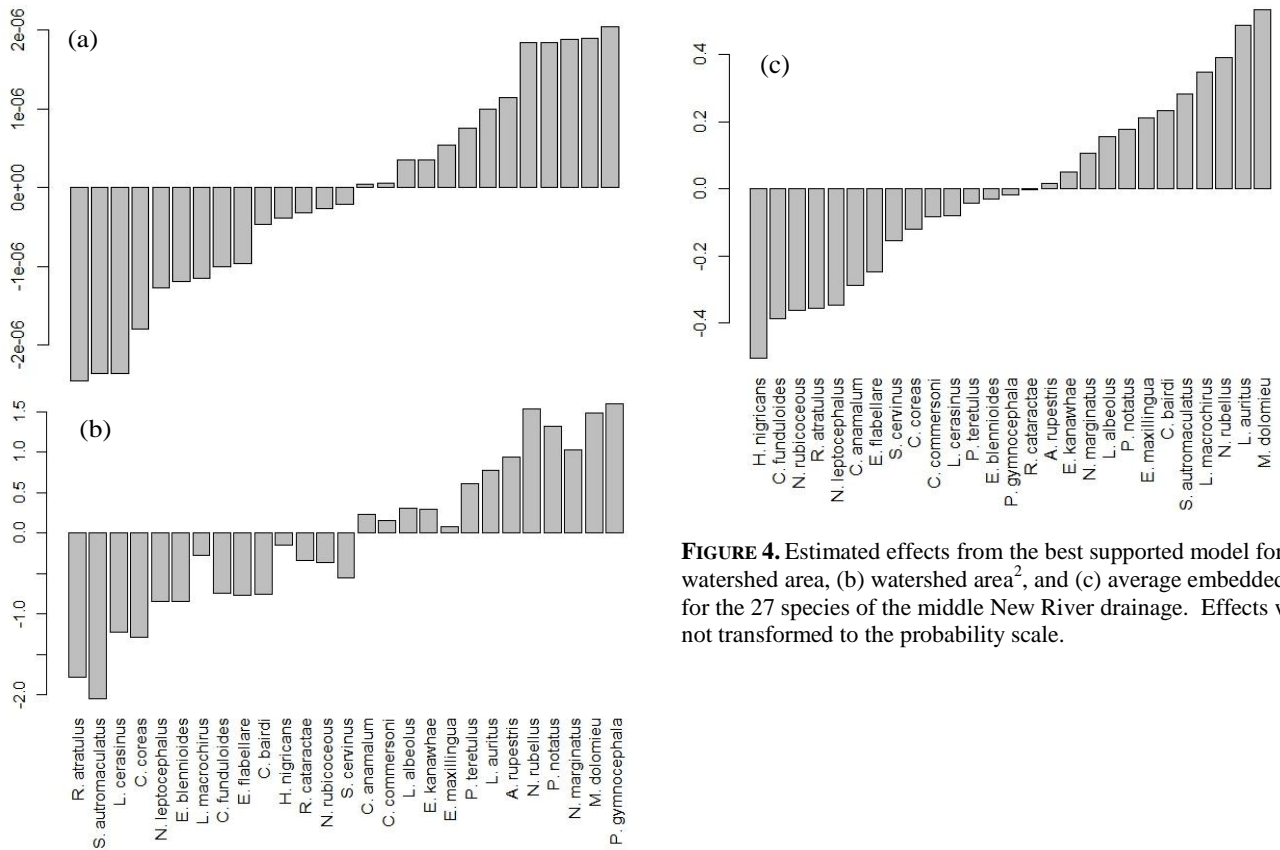


FIGURE 4. Estimated effects from the best supported model for (a) watershed area, (b) watershed area², and (c) average embeddedness for the 27 species of the middle New River drainage. Effects were not transformed to the probability scale.

to streams can disrupt feeding and reproductive habits of fishes, resulting in negative impacts on population processes and community structure (Muncy et al. 1979). This study demonstrates that within the context of stream size, the distributions of stream fishes are regulated by human land use changes at multiple spatial scales.

The lack of common responses among species characterized by our various grouping variables probably reflects the coarseness of the groups themselves. With respect to response to predictor variables, the diversity of trait syndromes within each group was large. Each group contained species adapted to headwaters and large watersheds (influencing responses to watershed area), to persistence in disturbed habitats (influencing responses to percent forest and embeddedness), to various mesohabitat types (influencing responses to percent pool), and that occurred in both physiographic provinces. Partitioning our coarse groupings into more specific ones may have yielded significant

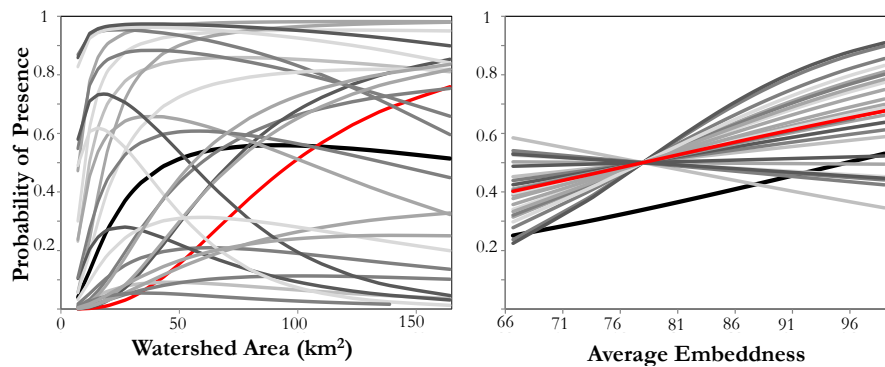


FIGURE 5. Predicted relationships with the environmental covariates for the best supported multi-species hierarchical model. The average effect is shown in black and the estimated effect for the Appalachian darter is shown in red.

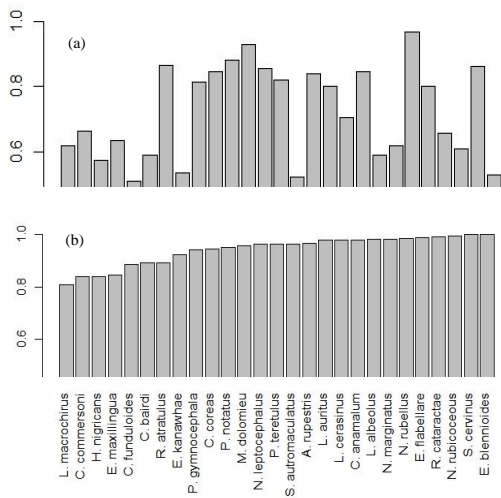


FIGURE 6. Estimates of the area under the curve based on the receiver operator characteristic using (a) only the environmental covariates and (b) both the environmental and spatial covariates.

continuum of fish life-history strategies proposed by Winemiller and Rose (1992) provides a useful framework for doing so. Performing ordination analyses on matrices of phylogenetically independent life-history traits (e.g. maximum longevity, relative fecundity, an ordinal variable describing parental care, etc...) would allow for the identification of much more meaningful groups.

Another setback to our approach is the failure to incorporate interspecific interactions. Within the broader context of habitat heterogeneity, interspecific interactions regulate species occurrences by expanding or shrinking realized niches (Bruno et al. 2003, and references therein). For instance, the presence of piscivorous fishes (which, as we demonstrate, typically prefer larger habitats) may inhibit the occurrence of prey species up to a point in the watershed where the predator can no longer occur. Likewise, competition between native and introduced species (e.g., Appalachia darter and Roanoke darter *P. roanoka*) may result in habitat-mediated competitive exclusion (i.e., competition restricts both to different portions of the watershed where the native species once occurred throughout), and thus may underlie distributional patterns. In both of these cases, interspecific interactions shrink species' realized niches. Alternatively, positive interactions such as nest association between *Nocomis* and other minnows may expand their distributions through habitat amelioration (beneficial to associates) and increased larval survival (beneficial to all participants). With such a limited dataset, the full incorporation of interspecific interactions would not be feasible for this study. Doing so would cause a situation where the number of factors approaches the number of observations, resulting in highly unstable models. The incorporation of interspecific interactions into multiple-species occupancy models appears to be a fruitful area of research, and should be further explored.

The benefits and setbacks of single- versus multiple-species occupancy models depend on the question of interest. Although researchers who seek to model distributions of single species may focus on that species' occurrence records and associated habitat variables, information on single species could be gained by placing that species within the context of the entire community; this would facilitate inferences on interspecific interactions. However, those who do so should be wary of assigning causation (e.g. predation, competition, or mutualism) to correlation (i.e. species

group-wise effects, but would have resulted in nearly as many groups as number of species. In such a case, grouping terms would provide little more information than species-wise responses. For example, one binary grouping term characterizes whether or not adults of a species provide parental care to broods (parental care is believed to increase larval fish survival by decreasing predation, *sensu* Johnston 1999). However, parental care differs greatly among species in this group: bluehead chubs *Nocomis leptoccephalus* construct nests and maintain them by turning stones, but do not directly guard eggs or fry; fantail darters *Etheostoma flabellare* guard eggs, but not fry; and sunfishes (Centrarchidae) guard broods through late larval states. Even among sunfishes, the length of parental guarding time differs (Jenkins and Burkhead 1994). Among constituent species in this group, heterogeneity in response to environmental factors probably outweighs the commonality of some level of parental care provided to broods. A more quantitative approach to grouping species would be to delineate groups based on life-history strategies. The triangular

associations arising from similar environmental requirements). Although multiple-species models were more likely to reach convergence than single-species models, overall interpretation of multiple-species models can be difficult. On the one hand, species-wise parameter estimates from multiple-species models are obtainable, but interpreting each one individually could quickly become complicated (although this issue could be circumvented by using relevant grouping terms). On the other hand, multiple-species models provide a means for researchers to take a more holistic approach to predicting species distributions. Multiple-species models also serve as a useful tool to succinctly characterize changes in species richness along environmental gradients. Such tools may provide fresh viewpoints in ecologists' inherent search to understand relationships between organisms and their habitats.

REFERENCES

- Akaike, H. 1978. A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics* 30:9–14.
- Allan, J.D. 1995. *Stream ecology: the structure and function of running waters*. Chapman and Hall, New York.
- Allan, J.D. 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems. *Annual Review of Ecology, Evolution, and Systematics* 35: 257-284.
- Balon, E. K. 1975. Reproductive guilds of fishes - proposal and definition. *Journal of the Fish Research Board of Canada* 32:821-864.
- Barbour, M.T., J. Gerristen, B.D. Snyder, and J.B. Stribling. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish, 2nd edition. U. S. Environmental Protection Agency, Office of Water EPA 841-B-99-002. Washington, DC.
- Bates, D., M. Maechler, M. and B. Bolker. 2011. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-42. <http://CRAN.R-project.org/package=lme4>
- Beecher, H.A., E. R. Dott, and R.F. Fernau. 1988. Fish species richness and stream order in Washington State streams. *Environmental Biology of Fishes* 22: 193-208.
- Bruno, J.F., J.J. Stachowicz, and M.D. Bertness. 2003. Inclusion of facilitation into ecological theory. *Trends in Ecology and Evolution* 18: 119-125.
- Burnham, K. P., and D. R. Anderson. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer-Verlag, New York.
- DeWan A.A. and E.F. Zipkin. 2010. An integrated sampling and analysis approach for improved biodiversity monitoring. *Environmental Management* 45:1223–1230.
- Dorazio, R.M. and J.A. Royle 2005. Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the American Statistical Association* 100:389-398.
- Dorazio, R.M., M. Kery, J.A. Royle, and M. Plattner. 2010. Models for inference in dynamic metacommunity systems. *Ecology* 91: 2466–2475.
- Dorazio, R.M., J.A. Royle, B. Soderstrom, and A. Glimskar. 2006. Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology* 87:842–854.
- Downes, B.J., P.S. Lake, S.G. Schreiber, and A. Glaister. 1998. Habitat structure and regulation of species diversity in a stony, upland stream. *Ecological Monographs* 68: 237-257.

- Ferrier, S. and Guisan, A. 2006. Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, 43: 393–404.
- Fitzpatrick, M. C., Sanders, N. J., Ferrier, S., Longino, J. T., Weiser, M. D. and Dunn, R. 2011. Forecasting the future of biodiversity: a test of single- and multi-species models for ants in North America. *Ecography*, 34: 836–847.
- Franklin J. 2009. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press, Cambridge.
- Gorman, O.T. and J.R. Karr. 1978. Habitat structure and stream fish communities. *Ecology* 59: 507-515.
- Hanley, J.A. and B.J. McNeil. 1982. The meaning and use of the area under a receiving operating characteristic curve. *Radiology* 143: 28-36.
- Hurvich, C. M., and C.-L. Tsai. 1989. Regression and time series model selection in small samples. *Biometrika* 76:297–307.
- Hynes, H.B.N. 1975. The stream and its valley. *Verhandlungen, Internationale Vereinigung für Theoretische und Angewandte Limnologie* 19: 1-15.
- Jenkins, R.E., and N.M. Burkhead. 1994. *Freshwater Fishes of Virginia*. American Fisheries Society, Bethesda.
- Johnston, C.E. 1999. The relationship of spawning mode to conservation of North American minnows (Cyprinidae). *Environmental Biology of Fishes* 55: 21-30.
- Kery, M. and J.A. Royle. 2008. Hierarchical Bayes estimation of species richness and occupancy in spatially replicated surveys. *Journal of Applied Ecology* 45: 589-598.
- Kery, M. and J.A. Royle. 2009. Inference about species richness and community structure using species-specific occupancy models in the National Swiss Breeding Bird Survey MHB. *In*: Thomson D.L., E.G. Cooch, and M.J. Conroy (eds). *Modeling demographic processes in marked populations*. Springer, New York, pp 639– 656.
- Marc, M. J. 2012. AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). R package version 1.24. <http://CRAN.R-project.org/package=AICcmodavg>.
- Matthews, W.J. and H.W. Robison. 1998. Influence of drainage connectivity, drainage area, and regional species richness on fishes in the Interior Highlands of Arkansas. *American Midland Naturalist* 139: 1-19.
- Muncy, R.J., G.J. Atchison, R.V. Bulkley, B.W. Menzel, L.G. Perry, and R.C. Summerfelt. Effects of suspended solids and sediment on reproduction and early life of warmwater fishes: a review. EPA Report 600/3-79-042. U.S. Environmental Protection Agency. Washington.
- Paller, M.H. 1994. Relationships between fish assemblage structure and stream order in South Carolina coastal plain streams. *Transactions of the American Fisheries Society* 123: 150-161.
- Peoples, B. K., M. B. Tainer, and E. A. Frimpong. 2011. Bluehead chub nesting activity: a potential mechanism of population persistence in degraded stream habitats. *Environmental Biology of Fishes* 90:379-391.
- R Development Core Team (2012). *R: A language and environment for statistical computing*. R foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Winemiller, K.O. and K.A. Rose. 1992. Patterns of life-history diversification in North American fishes: implications for population regulation. *Canadian Journal of fisheries and Aquatic Sciences* 49: 2196-2218.

APPENDIX A. Grouping terms of 27 New River fishes whose distributions were modeled. Groupings were assigned based on information found in Jenkins and Burkhead 1994, and references therein.

Species	Family	Reproductive guild*	Parental care
White sucker <i>Catostomus commersonii</i>	Catostomidae	Simple lithophil	No
Nothern hogsucker <i>Hypentelium nigricans</i>	Catostomidae	Simple lithophil	No
Blacktip jumprock <i>Moxostoma cervinus</i>	Catostomidae	Simple lithophil	No
Rock bass <i>Ambloplites rupestris</i>	Centrarchidae	Saucer-pit nester	Yes
Redbreast sunfish <i>Lepomis auritus</i>	Centrarchidae	Saucer-pit nester	Yes
Bluegill <i>Lepomis macrochirus</i>	Centrarchidae	Saucer-pit nester	Yes
Smallmouth bass <i>Micropterus dolimieu</i>	Centrarchidae	Saucer-pit nester	Yes
Mottled sculpin <i>Cottus bairdi</i>	Cottidae	Speleophil	Yes
Central stoneroller <i>Campostoma anomalum</i>	Cyprinidae	Simple lithophil	Yes
Mountain redbelly dace <i>Chrosomus oreas</i>	Cyprinidae	Simple lithophil/nest associate	No
Rosyside dace <i>Clinostomus funduloides</i>	Cyprinidae	Simple lithophil/nest associate	No
Cutlips minnow <i>Exoglossum maxillingua</i>	Cyprinidae	Gravel mound nester	Yes
White shiner <i>Luxilus albeolus</i>	Cyprinidae	Simple lithophil/nest associate	No
Crescent shiner <i>Luxilus cerasinus</i>	Cyprinidae	Simple lithophil/nest associate	No
Bluehead chub <i>Nocomis leptocephalus</i>	Cyprinidae	Gravel mound nester	Yes
Rosyface shiner <i>Notropis rubellus</i>	Cyprinidae	Simple lithophil/nest associate	No
Saffron shiner <i>Notropis rubricaceous</i>	Cyprinidae	Simple lithophil/nest associate	No
Kanawha minnow <i>Phenacobious teretulus</i>	Cyprinidae	Simple lithophil	No
Bluntnose minnow <i>Pimephales notatus</i>	Cyprinidae	Rheo-phytophil	Yes
Blacknose dace <i>Rhinichthys atratulus</i>	Cyprinidae	Simple lithophil	No
Longnose dace <i>Rhinichthys cataractae</i>	Cyprinidae	Simple lithophil	No
Creek chub <i>Semotilus atromaculatus</i>	Cyprinidae	Gravel mound nester	Yes
Margined madtom <i>Noturus marginatus</i>	Ictaluridae	Speleophil	Yes
Greenside darter <i>Etheostoma blennoides</i>	Percidae	rheo-phytophil	Yes
Fantail darter <i>Etheostoma flabellare</i>	Percidae	Speleophil	Yes
Kanawha darter <i>Etheostoma kanawhae</i>	Percidae	Simple lithophil	No
Appalachia darter <i>Percina gymnocephala</i>	Percidae	Psammophil	No

* Based on Balon's (1975) proposal for fish reproductive guilds

APPENDIX B. R code to load data, do manipulations, and call libraries.

```
#####
#Load Data#####
#####
dat<-read.csv("NewRiver_Fish.csv",header=TRUE) #Load fish and habitat data
dist1<-read.csv("NewRiver_HydroDist.csv",header=TRUE) #Load hydrologic data
dist<-as.matrix(dist1[,-1]) #Remove site names
rownames(dist)<-dist1[,1] #Name rows as site names
new.dat<-dat[,c(2:31,34,37,41)] #Create new data frame for scaling
rownames(new.dat)<-dat[,1] #Name rows as site names
new.dat$LogWS_area<-scale(log(dat$WS_area))[1:nrow(new.dat)] #Scale Log WS
new.dat$LogWS_area2<-
scale(log(dat$WS_area)*log(dat$WS_area))[1:nrow(new.dat)] #Scale Quadratic
n.site<-nrow(new.dat) #Define number of sites
new.dat$P_Forest<-scale(dat$P_Forest)[1:nrow(new.dat)] #Scale Forest
new.dat$P_pool<-scale(dat$P_pool)[1:nrow(new.dat)] #Scale Percent Forest
new.dat$AVE_EMB<-scale(dat$AVE_EMB)[1:nrow(new.dat)] #Average Embedddness
covariates<-new.dat[,c(29,31:35)] #Covariate data frame for testing corrs
Y<-new.dat[,1:27] #Specify response matrix
n.species<-ncol(Y) #Define number of species
group.dat<-read.csv("NewRiver_Groups.csv",header=TRUE) #Load group data
guild<-group.dat$Group #Define guild information
nest<-group.dat$Nest_Assoc #Define nesting information
family<-group.dat$Family #Define family information
parent<-c(1,1,1,rep(2,5),1,3,3,1,3,3,2,1,2,1,2,1,1,2,2,2,2,1,1) #parental info
#####
#Create data for hierarchical model #
#####
yvec<-as.vector(as.matrix(Y)) #vector of responses
species<-as.factor(rep(1:n.species,each=n.site)) #factor for species id
sites<-rep(1:n.site,n.species) #vector of sites as factors
obs<-1:length(yvec) #Vector representing each site
lmer.dat<-data.frame(yvec,species,sites,obs) #Create new dataframe
lmer.dat$system<-c(dat$System,rep(0,n.site*(n.species-1))) #vector of tribes
lmer.dat$samparea<-rep(dat$Area,n.species) #vector of area sampled
lmer.dat$LogWS_area<-rep(scale(log(dat$WS_area)),n.species) #Ln WS area
lmer.dat$LogWS_area2<-scale(rep(log(dat$WS_area),n.species)*
rep(log(dat$WS_area),n.species)) #LN WS area sq
lmer.dat$BlueRidge<-rep(new.dat$BlueRidge,n.species) #Blue Ridge
lmer.dat$P_Forest<-rep(new.dat$P_Forest,n.species) #Percent Forest
lmer.dat$P_pool<-rep(new.dat$P_pool,n.species) #Percent Pool
lmer.dat$AVE_EMB<-rep(new.dat$AVE_EMB,n.species) #Average Embedddness
lmer.dat$guilds<-rep(guild,each=n.site) #Guilds
lmer.dat$families<-rep(family,each=n.site) #Families
lmer.dat$parents<-rep(parent,each=n.site) #Parental Care
nest.full<-nest[species] #Nests
parent.1<-ifelse(parent==2,1,0) #Parental Care
parent.2<-ifelse(parent==3,1,0) #Parental Care
parent.1.full<-parent.1[species] #Parental Care
parent.2.full<-parent.2[species] #Parental Care
#####
#Load libraries#
#####
library(lme4);library(arm);library(AICcmodavg);library(spdep)
library(spacemaker)
```

APPENDIX C. R code to create semivariogram based on hydrologic distance.

```
ro <- function(x, r) {return(round(x/r)*r)} #Function to round to any integer
#####
#Note: to do semivariogram for other species change the 27 to the nubmer of#
#the species of interest
#####
##Semivariogram for appalachia darter
#####
global<-
glmer(Y[,27]~BlueRidge+LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+
      (1|System),data=new.dat,family=binomial(link="logit")) #Global model
res<-Y[,27]-fitted(global) #create residuals
Z<-array(dim=dim(dist)) #Create empty array
for(i in 1:length(res)){
  for(j in 1:length(res)){
    Z[i,j]<-0.5*(res[i]-res[j])^2}}
semi.dat<-data.frame(as.vector(dist),as.vector(Z))
semi.dat<-semi.dat[order(semi.dat[,1]),]
semi.dat<-semi.dat[-(1:38),] #Remove missing values
semi.dat<-semi.dat[-(1407:1562),] #Remove missing values
semi.dat[,1]<-semi.dat[,1]/1000 #scale to km scale
semi.dat$bin1<-ro(semi.dat[,1],1) #Create 1 km bins
semi.dat$bin2<-ro(semi.dat[,1],2) #Create 2 km bins
semi.dat$bin3<-ro(semi.dat[,1],3) #Create 3 km bins
semi.dat$bin10<-ro(semi.dat[,1],10) #Create 10 km bins
bin1<-seq(1,171.5,1);bin2<-seq(0,172,2);bin3<-seq(0,174,3);bin10<-
seq(0,180,10)
result.bin1<-rep(NA,length(bin1))
for(i in 1:length(bin1)){
  result.bin1[i]<-1/(2*sum(ifelse(semi.dat$bin1==bin1[i],1,0)))
  *sum(ifelse(semi.dat$bin1==bin1[i],semi.dat[,2],0))}
result.bin2<-rep(NA,length(bin2))
for(i in 1:length(bin2)){
  result.bin2[i]<-1/(2*sum(ifelse(semi.dat$bin2==bin2[i],1,0))) *
  sum(ifelse(semi.dat$bin2==bin2[i],semi.dat[,2],0))}
result.bin3<-rep(NA,length(bin3))
for(i in 1:length(bin3)){
  result.bin3[i]<-1/(2*sum(ifelse(semi.dat$bin3==bin3[i],1,0))) *
  sum(ifelse(semi.dat$bin3==bin3[i],semi.dat[,2],0))}
result.bin10<-rep(NA,length(bin10))
for(i in 1:length(bin10)){
  result.bin10[i]<-1/(2*sum(ifelse(semi.dat$bin10==bin10[i],1,0))) *
  sum(ifelse(semi.dat$bin10==bin10[i],semi.dat[,2],0))}
#Create Plots
par(mfrow=c(4,1))
plot(bin1,result.bin1,main="1 km Bin",ylab="Semivariance")
plot(bin2,result.bin2,main="2 km Bin",ylab="Semivariance")
plot(bin3,result.bin3,main="3 km Bin",ylab="Semivariance")
plot(bin10,result.bin10,main="10 km Bin",ylab="Semivariance")
```

APPENDIX D. R code to perform logistic regression on a Appalachia darter.

```
all.models<-vector("list",48)
all.models[[1]]<-glmer(Y[,27]~1+(1|System),data=new.dat,
  family=binomial(link="logit"))
all.models[[2]]<-glmer(Y[,27]~BlueRidge+(1|System),data=new.dat,
  family=binomial(link="logit"))
all.models[[3]]<-glmer(Y[,27]~LogWS_area+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[4]]<-glmer(Y[,27]~p_Forest+(1|System),data=new.dat,
  family=binomial(link="logit"))
all.models[[5]]<-glmer(Y[,27]~p_pool+(1|System),data=new.dat,
  family=binomial(link="logit"))
all.models[[6]]<-glmer(Y[,27]~AVE_EMB+(1|System),data=new.dat,
  family=binomial(link="logit"))
all.models[[7]]<-glmer(Y[,27]~BlueRidge+LogWS_area+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[8]]<-glmer(Y[,27]~BlueRidge+p_Forest+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[9]]<-glmer(Y[,27]~BlueRidge+p_pool+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[10]]<-glmer(Y[,27]~BlueRidge+AVE_EMB+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[11]]<-glmer(Y[,27]~LogWS_area+LogWS_area2+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[12]]<-glmer(Y[,27]~LogWS_area+p_Forest+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[13]]<-glmer(Y[,27]~LogWS_area+p_pool+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[14]]<-glmer(Y[,27]~LogWS_area+AVE_EMB+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[15]]<-glmer(Y[,27]~p_Forest+p_pool+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[16]]<-glmer(Y[,27]~p_Forest+AVE_EMB+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[17]]<-glmer(Y[,27]~p_pool+AVE_EMB+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[18]]<-glmer(Y[,27]~BlueRidge+LogWS_area+LogWS_area2+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[19]]<-glmer(Y[,27]~BlueRidge+LogWS_area+p_Forest+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[20]]<-glmer(Y[,27]~BlueRidge+LogWS_area+p_pool+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[21]]<-glmer(Y[,27]~BlueRidge+LogWS_area+AVE_EMB+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[22]]<-glmer(Y[,27]~BlueRidge+p_Forest+p_pool+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[23]]<-glmer(Y[,27]~BlueRidge+p_Forest+AVE_EMB+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[24]]<-glmer(Y[,27]~BlueRidge+p_pool+AVE_EMB+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[25]]<-glmer(Y[,27]~LogWS_area+LogWS_area2+p_Forest+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[26]]<-glmer(Y[,27]~LogWS_area+LogWS_area2+p_pool+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[27]]<-glmer(Y[,27]~LogWS_area+LogWS_area2+AVE_EMB+(1|System),
  data=new.dat,family=binomial(link="logit"))
```

#Continued from previous page

```
all.models[[28]]<-glmer(Y[,27]~LogWS_area+p_Forest+p_pool+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[29]]<-glmer(Y[,27]~LogWS_area+p_Forest+AVE_EMB+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[30]]<-glmer(Y[,27]~LogWS_area+p_pool+AVE_EMB+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[31]]<-glmer(Y[,27]~p_Forest+p_pool+AVE_EMB+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[32]]<-
glmer(Y[,27]~BlueRidge+LogWS_area+LogWS_area2+p_Forest+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[33]]<-
glmer(Y[,27]~BlueRidge+LogWS_area+LogWS_area2+p_pool+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[34]]<-
glmer(Y[,27]~BlueRidge+LogWS_area+LogWS_area2+AVE_EMB+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[35]]<-
glmer(Y[,27]~BlueRidge+LogWS_area+p_Forest+p_pool+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[36]]<-
glmer(Y[,27]~BlueRidge+LogWS_area+p_Forest+AVE_EMB+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[37]]<-
glmer(Y[,27]~BlueRidge+LogWS_area+p_pool+AVE_EMB+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[38]]<-glmer(Y[,27]~BlueRidge+p_Forest+p_pool+AVE_EMB+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[39]]<-
glmer(Y[,27]~LogWS_area+LogWS_area2+p_Forest+p_pool+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[40]]<-
glmer(Y[,27]~LogWS_area+LogWS_area2+p_Forest+AVE_EMB+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[41]]<-
glmer(Y[,27]~LogWS_area+LogWS_area2+p_pool+AVE_EMB+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[42]]<-glmer(Y[,27]~LogWS_area+p_Forest+p_pool+AVE_EMB+(1|System),
  data=new.dat,family=binomial(link="logit"))
all.models[[43]]<-glmer(Y[,27]~BlueRidge+LogWS_area+LogWS_area2+p_Forest+
  p_pool+(1|System),data=new.dat,family=binomial(link="logit"))
all.models[[44]]<-glmer(Y[,27]~BlueRidge+LogWS_area+LogWS_area2+p_Forest+
  AVE_EMB+(1|System),data=new.dat,family=binomial(link="logit"))
all.models[[45]]<-glmer(Y[,27]~BlueRidge+LogWS_area+LogWS_area2+p_pool+
  AVE_EMB+(1|System),data=new.dat,family=binomial(link="logit"))
all.models[[46]]<-glmer(Y[,27]~BlueRidge+LogWS_area+p_Forest+p_pool+
  AVE_EMB+(1|System),data=new.dat,family=binomial(link="logit"))
all.models[[47]]<-glmer(Y[,27]~LogWS_area+LogWS_area2+p_Forest+p_pool+
  AVE_EMB+(1|System),data=new.dat,family=binomial(link="logit"))
all.models[[48]]<-
glmer(Y[,27]~BlueRidge+LogWS_area+LogWS_area2+p_Forest+p_pool+
  AVE_EMB+(1|System),data=new.dat,family=binomial(link="logit"))
#Get Model Fit Statistics
model.name<-rep(NA,48)
tmp<-1:48
```

```

#Continued from previous page
model.name[1]<-"null model"
for(i in 2:48){
model.name[i]<-paste("model.",as.list(tmp[i]),sep="")
model.table<-array(dim=c(length(all.models),12))
rownames(model.table)<-model.name
colnames(model.table)<-c("Ln
L","K","AICc","delta","wi","Intercept","BlueRidge","LogWS_area","LogWS_area2"
,"p_Forest","p_pool","AVE_EMB")
for(i in 1:length(all.models)){
  model.table[i,1]<-deviance(all.models[[i]])/-2
  model.table[i,2]<-AICc(all.models[[i]],return.K=TRUE)
  model.table[i,3]<-AICc(all.models[[i]],return.K=FALSE)
  model.table[i,6]<-fixef(all.models[[i]])["(Intercept)"]
  model.table[i,7]<-fixef(all.models[[i]])["BlueRidge"]
  model.table[i,8]<-fixef(all.models[[i]])["LogWS_area"]
  model.table[i,9]<-fixef(all.models[[i]])["LogWS_area2"]
  model.table[i,10]<-fixef(all.models[[i]])["p_Forest"]
  model.table[i,11]<-fixef(all.models[[i]])["p_pool"]
  model.table[i,12]<-fixef(all.models[[i]])["AVE_EMB"]
}
for(i in 1:length(all.models)){
  model.table[i,4]<-model.table[i,3]-min(model.table[,3])
}
for(i in 1:length(all.models)){
  model.table[i,5]<-exp(-0.5*model.table[i,4])/sum(exp(-
0.5*model.table[,4]))
}
model.table<-model.table[order(model.table[,3]),]

```


APPENDIX E. R code to perform PCNM on the New River fish community data.

```
sites<-read.csv("sites_gps.csv",header=TRUE) #Read data in
sites<-as.matrix(sites) #Create data as matrix
rownames(sites)<-dat$Site
nbnear2000<-dnearneigh(sites,0,2000,row.names=dat$Site,longlat=FALSE)
nbnear4000<-dnearneigh(sites,0,4000,row.names=dat$Site,longlat=FALSE)
nbnear6000<-dnearneigh(sites,0,6000,row.names=dat$Site,longlat=FALSE)
nbnear8000<-dnearneigh(sites,0,8000,row.names=dat$Site,longlat=FALSE)
nbnear10000<-dnearneigh(sites,0,10000,row.names=dat$Site,longlat=FALSE)
nbnear12000<-dnearneigh(sites,0,12000,row.names=dat$Site,longlat=FALSE)
par(mfrow=c(3,2))
plot(nbnear2000,sites,col="red",pch=20,cex=2)
title(main="neighbors if 0<d<2km")
plot(nbnear4000,sites,col="red",pch=20,cex=2)
title(main="neighbors if 0<d<4km")
plot(nbnear6000,sites,col="red",pch=20,cex=2)
title(main="neighbors if 0<d<6km")
plot(nbnear8000,sites,col="red",pch=20,cex=2)
title(main="neighbors if 0<d<8km")
plot(nbnear10000,sites,col="red",pch=20,cex=2)
title(main="neighbors if 0<d<10km")
plot(nbnear12000,sites,col="red",pch=20,cex=2)
title(main="neighbors if 0<d<12km")

#####
#Start at 6 where each site has at least one neighbor#
#####
dist_sites<-nbdists(nbnear6000,sites)
fdist<-lapply(dist_sites,function(x) 1-x/max(dist(sites)))
listw_sites<-nb2listw(nbnear6000,glist=fdist,style="W")
eigen_sites=scores.listw(listw_sites,echo=TRUE)
moranI<-test.scores(eigen_sites,listw_sites,999)
vectors6<-eigen_sites$vectors[,1:10]
lmer.dat$vec1<-rep(vectors6[,1],n.species)
lmer.dat$vec2<-rep(vectors6[,2],n.species)
lmer.dat$vec3<-rep(vectors6[,3],n.species)
lmer.dat$vec4<-rep(vectors6[,4],n.species)
lmer.dat$vec5<-rep(vectors6[,5],n.species)
lmer.dat$vec6<-rep(vectors6[,6],n.species)
lmer.dat$vec7<-rep(vectors6[,7],n.species)
lmer.dat$vec8<-rep(vectors6[,8],n.species)
lmer.dat$vec9<-rep(vectors6[,9],n.species)
lmer.dat$vec10<-rep(vectors6[,10],n.species)
models1<-
glmer(yvec~1+(1|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+vec9+vec10+(
0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|sp
ecies)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species)+(0+vec9|species)+(0
+vec10|species),data=lmer.dat,family=binomial(link="logit"))

#8 km
dist_sites<-nbdists(nbnear8000,sites)
fdist<-lapply(dist_sites,function(x) 1-x/max(dist(sites)))
listw_sites<-nb2listw(nbnear8000,glist=fdist,style="W")
eigen_sites=scores.listw(listw_sites,echo=TRUE)
moranI<-test.scores(eigen_sites,listw_sites,999)
vectors8<-eigen_sites$vectors[,1:8]
```

#continued from previous page

```
lmer.dat$vec1<-rep(vectors8[,1],n.species)
lmer.dat$vec2<-rep(vectors8[,2],n.species)
lmer.dat$vec3<-rep(vectors8[,3],n.species)
lmer.dat$vec4<-rep(vectors8[,4],n.species)
lmer.dat$vec5<-rep(vectors8[,5],n.species)
lmer.dat$vec6<-rep(vectors8[,6],n.species)
lmer.dat$vec7<-rep(vectors8[,7],n.species)
lmer.dat$vec8<-rep(vectors8[,8],n.species)
models2<-glmer(yvec~1+(1|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+
(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+
(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),
data=lmer.dat,family=binomial(link="logit"))
```

#10 km

```
dist_sites<-nbdists(nbnear10000,sites)
fdist<-lapply(dist_sites, function(x) 1-x/max(dist(sites)))
listw_sites<-nb2listw(nbnear10000,glist=fdist,style="W")
eigen_sites=scores.listw(listw_sites, echo=TRUE)
morani<-test.scores(eigen_sites,listw_sites,999)
vectors10<-eigen_sites$vectors[,1:7]
lmer.dat$vec1<-rep(vectors10[,1],n.species)
lmer.dat$vec2<-rep(vectors10[,2],n.species)
lmer.dat$vec3<-rep(vectors10[,3],n.species)
lmer.dat$vec4<-rep(vectors10[,4],n.species)
lmer.dat$vec5<-rep(vectors10[,5],n.species)
lmer.dat$vec6<-rep(vectors10[,6],n.species)
lmer.dat$vec7<-rep(vectors10[,7],n.species)
models3<-
glmer(yvec~1+(1|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+(0+vec1|species)+
(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|s
pecies)+(0+vec7|species),data=lmer.dat,family=binomial(link="logit"))
```

#12 km

```
dist_sites<-nbdists(nbnear12000,sites)
fdist<-lapply(dist_sites, function(x) 1-x/max(dist(sites)))
listw_sites<-nb2listw(nbnear12000,glist=fdist,style="W")
eigen_sites=scores.listw(listw_sites, echo=TRUE)
morani<-test.scores(eigen_sites,listw_sites,999)
vectors12<-eigen_sites$vectors[,1:7]
lmer.dat$vec1<-rep(vectors12[,1],n.species)
lmer.dat$vec2<-rep(vectors12[,2],n.species)
lmer.dat$vec3<-rep(vectors12[,3],n.species)
lmer.dat$vec4<-rep(vectors12[,4],n.species)
lmer.dat$vec5<-rep(vectors12[,5],n.species)
lmer.dat$vec6<-rep(vectors12[,6],n.species)
lmer.dat$vec7<-rep(vectors12[,7],n.species)
models4<-
glmer(yvec~1+(1|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+(0+vec1|species)+
(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|s
pecies)+(0+vec7|species),data=lmer.dat,family=binomial(link="logit"))
```

#Continued from previous page

```
#####  
#Use 8km neighborhood#  
#####  
lmer.dat$vec1<-rep(vectors8[,1],n.species)  
lmer.dat$vec2<-rep(vectors8[,2],n.species)  
lmer.dat$vec3<-rep(vectors8[,3],n.species)  
lmer.dat$vec4<-rep(vectors8[,4],n.species)  
lmer.dat$vec5<-rep(vectors8[,5],n.species)  
lmer.dat$vec6<-rep(vectors8[,6],n.species)  
lmer.dat$vec7<-rep(vectors8[,7],n.species)  
lmer.dat$vec8<-rep(vectors8[,8],n.species)
```

APPENDIX F. R code to perform hierarchical multi-species logistic regression to determine weight of evidence for grouping terms.

#No grouping terms

```
model.1<-  
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+vec1+vec2+vec3+vec4  
+vec5+vec6+vec7+vec8+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|specie  
s)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE_EMB|species)+(0+vec1|specie  
s)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec  
6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(li  
nk="logit"))
```

#Guilds

```
model.2<-  
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+vec1+vec2+vec3+vec4  
+vec5+vec6+vec7+vec8+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|specie  
s)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE_EMB|species)+(0+vec1|specie  
s)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec  
6|species)+(0+vec7|species)+(0+vec8|species)+(1|guilds),data=lmer.dat,family=  
binomial(link="logit"))
```

#Families

```
model.3<-  
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+vec1+vec2+vec3+vec4  
+vec5+vec6+vec7+vec8+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|specie  
s)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE_EMB|species)+(0+vec1|specie  
s)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec  
6|species)+(0+vec7|species)+(0+vec8|species)+(1|families),data=lmer.dat,famil  
y=binomial(link="logit"))
```

#Parental Care

```
model.4<-  
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+parent.1.full+paren  
t.2.full+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(1|species)+(0+LogWS_area|sp  
ecies)+(0+LogWS_area2|species)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE  
_EMB|species)+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|spec  
ies)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data  
=lmer.dat,family=binomial(link="logit"))
```

#Nest Associates

```
model.5<-  
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+nest.full+vec1+vec2  
+vec3+vec4+vec5+vec6+vec7+vec8+(1|species)+(0+LogWS_area|species)+(0+LogWS_ar  
ea2|species)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE_EMB|species)+(0+v  
ec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|speci  
es)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=b  
inomial(link="logit"))
```

#Families and Guilds

```
model.6<-  
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+vec1+vec2+vec3+vec4  
+vec5+vec6+vec7+vec8+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|specie  
s)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE_EMB|species)+(0+vec1|specie  
s)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec  
6|species)+(0+vec7|species)+(0+vec8|species)+(1|families)+(1|guilds),data=lme  
r.dat,family=binomial(link="logit"))
```

#Continued from previous page

#Parental Care and Guilds

```
model.7<-  
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+parent.1.full+paren  
t.2.full+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(1|species)+(0+LogWS_area|sp  
ecies)+(0+LogWS_area2|species)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE  
_EMB|species)+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|spec  
ies)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species)+(1|g  
uilds),data=lmer.dat,family=binomial(link="logit"))
```

#Nest associates and Guilds

```
model.8<-  
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+nest.full+vec1+vec2  
+vec3+vec4+vec5+vec6+vec7+vec8+(1|species)+(0+LogWS_area|species)+(0+LogWS_ar  
ea2|species)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE_EMB|species)+(0+v  
ec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|speci  
es)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species)+(1|guilds),data=lmer.d  
at,family=binomial(link="logit"))
```

#Parental Care and Families

```
model.9<-  
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+parent.1.full+paren  
t.2.full+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(1|species)+(0+LogWS_area|sp  
ecies)+(0+LogWS_area2|species)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE  
_EMB|species)+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|spec  
ies)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species)+(1|f  
amilies),data=lmer.dat,family=binomial(link="logit"))
```

#Nest associates and Families

```
model.10<-  
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+nest.full+vec1+vec2  
+vec3+vec4+vec5+vec6+vec7+vec8+(1|species)+(0+LogWS_area|species)+(0+LogWS_ar  
ea2|species)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE_EMB|species)+(0+v  
ec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|speci  
es)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species)+(1|families),data=lmer  
.dat,family=binomial(link="logit"))
```

#Parental Care and nest associates

```
model.11<-  
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+parent.1.full+paren  
t.2.full+nest.full+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(1|species)+(0+Log  
WS_area|species)+(0+LogWS_area2|species)+(0+p_Forest|species)+(0+p_pool|speci  
es)+(0+AVE_EMB|species)+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0  
+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|spe  
cies),data=lmer.dat,family=binomial(link="logit"))
```

#Parental Care, Families , and Guilds

```
model.12<-  
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+parent.1.full+paren  
t.2.full+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(1|species)+(0+LogWS_area|sp  
ecies)+(0+LogWS_area2|species)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE  
_EMB|species)+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|spec  
ies)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species)+(1|g  
uilds)+(1|families),data=lmer.dat,family=binomial(link="logit"))
```

#Continued from previous page

#Nest associates, Families, and Guilds

model.13<-

```
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+nest.full+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE_EMB|species)+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species)+(1|guilds)+(1|families),data=lmer.dat,family=binomial(link="logit"))
```

#Parental Care, Nest associates , and Guilds

model.14<-

```
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+parent.1.full+parent.2.full+nest.full+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE_EMB|species)+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species)+(1|guilds),data=lmer.dat,family=binomial(link="logit"))
```

#Parental Care, Nest associates , and Families

model.15<-

```
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+parent.1.full+parent.2.full+nest.full+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE_EMB|species)+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species)+(1|families),data=lmer.dat,family=binomial(link="logit"))
```

#Parental Care, Nest associates ,Guilds and Families

model.16<-

```
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+parent.1.full+parent.2.full+nest.full+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE_EMB|species)+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species)+(1|families)+(1|guilds),data=lmer.dat,family=binomial(link="logit"))
```

APPENDIX G. R code to perform hierarchical multi-species logistic regression using environmental covariates.

```
models<-vector("list",48)
models[[1]]<-
glmer(yvec~1+(1|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[2]]<-
glmer(yvec~LogWS_area+(1|species)+(0+LogWS_area|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[3]]<-
glmer(yvec~LogWS_area+(1|species)+(0+LogWS_area|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[4]]<-
glmer(yvec~p_Forest+(1|species)+(0+p_Forest|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[5]]<-
glmer(yvec~p_pool+(1|species)+(0+p_pool|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[6]]<-
glmer(yvec~AVE_EMB+(1|species)+(0+AVE_EMB|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[7]]<-
glmer(yvec~BlueRidge+(1|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[8]]<-
glmer(yvec~LogWS_area+LogWS_area2+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[9]]<-
glmer(yvec~LogWS_area+p_Forest+(1|species)+(0+LogWS_area|species)+(0+p_Forest|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[10]]<-
glmer(yvec~LogWS_area+p_pool+(1|species)+(0+LogWS_area|species)+(0+p_pool|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
```

#Continued from previous page

```
models[[11]]<-
glmer(yvec~LogWS_area+AVE_EMB+(1|species)+(0+LogWS_area|species)+(0+AVE_EMB|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[12]]<-
glmer(yvec~LogWS_area+BlueRidge+(1|species)+(0+LogWS_area|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[13]]<-
glmer(yvec~p_Forest+p_pool+(1|species)+(0+p_Forest|species)+(0+p_pool|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[14]]<-
glmer(yvec~p_Forest+AVE_EMB+(1|species)+(0+p_Forest|species)+(0+AVE_EMB|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[15]]<-
glmer(yvec~p_Forest+BlueRidge+(1|species)+(0+p_Forest|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[16]]<-
glmer(yvec~p_pool+AVE_EMB+(1|species)+(0+p_pool|species)+(0+AVE_EMB|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[17]]<-
glmer(yvec~p_pool+BlueRidge+(1|species)+(0+p_pool|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[18]]<-
glmer(yvec~AVE_EMB+BlueRidge+(1|species)+(0+AVE_EMB|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[19]]<-
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+(0+p_Forest|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[20]]<-
glmer(yvec~LogWS_area+LogWS_area2+p_pool+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+(0+p_pool|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
```


#Continued from previous page

```
models[[21]]<-
glmer(yvec~LogWS_area+LogWS_area2+AVE_EMB+(1|species)+(0+LogWS_area|species)+
(0+LogWS_area2|species)+(0+AVE_EMB|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec
7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0
+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.d
at,family=binomial(link="logit"))
models[[22]]<-
glmer(yvec~LogWS_area+LogWS_area2+BlueRidge+(1|species)+(0+LogWS_area|species
)+(0+LogWS_area2|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6
+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species
)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lm
er.dat,family=binomial(link="logit"))
models[[23]]<-
glmer(yvec~LogWS_area+p_Forest+p_pool+(1|species)+(0+LogWS_area|species)+(0+p
_Forest|species)+(0+p_pool|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(
0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|sp
ecies)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,famil
y=binomial(link="logit"))
models[[24]]<-
glmer(yvec~LogWS_area+p_Forest+AVE_EMB+(1|species)+(0+LogWS_area|species)+(0+
p_Forest|species)+(0+AVE_EMB|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8
+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|
species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,fam
ily=binomial(link="logit"))
models[[25]]<-
glmer(yvec~LogWS_area+p_Forest+BlueRidge+(1|species)+(0+LogWS_area|species)+(
0+p_Forest|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+
vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+v
ec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat
,family=binomial(link="logit"))
models[[26]]<-
glmer(yvec~LogWS_area+p_pool+AVE_EMB+(1|species)+(0+LogWS_area|species)+(0+p_
pool|species)+(0+AVE_EMB|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+
vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|spec
ies)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=
binomial(link="logit"))
models[[27]]<-
glmer(yvec~LogWS_area+p_pool+BlueRidge+(1|species)+(0+LogWS_area|species)+(0+
p_pool|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8
+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|
species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,fam
ily=binomial(link="logit"))
models[[28]]<-
glmer(yvec~LogWS_area+AVE_EMB+BlueRidge+(1|species)+(0+LogWS_area|species)+(0
+AVE_EMB|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+ve
c8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec
5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,f
amily=binomial(link="logit"))
models[[29]]<-
glmer(yvec~p_Forest+p_pool+AVE_EMB+(1|species)+(0+p_Forest|species)+(0+p_pool
|species)+(0+AVE_EMB|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1
|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)
+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=bino
mial(link="logit"))
```

#Continued from previous page

```
models[[30]]<-
glmer(yvec~p_Forest+p_pool+BlueRidge+(1|species)+(0+p_Forest|species)+(0+p_pool|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[31]]<-
glmer(yvec~p_Forest+AVE_EMB+BlueRidge+(1|species)+(0+p_Forest|species)+(0+AVE_EMB|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[32]]<-
glmer(yvec~p_pool+AVE_EMB+BlueRidge+(1|species)+(0+p_pool|species)+(0+AVE_EMB|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[33]]<-
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+(0+p_Forest|species)+(0+p_pool|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[33]]<-
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+(0+p_Forest|species)+(0+p_pool|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[34]]<-
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+AVE_EMB+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+(0+p_Forest|species)+(0+AVE_EMB|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[35]]<-
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+BlueRidge+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+(0+p_Forest|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[36]]<-
glmer(yvec~LogWS_area+LogWS_area2+p_pool+AVE_EMB+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+(0+p_pool|species)+(0+AVE_EMB|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[37]]<-
glmer(yvec~LogWS_area+LogWS_area2+p_pool+BlueRidge+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+(0+p_pool|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
```

#Continued from previous page

```
models[[38]]<-
glmer(yvec~LogWS_area+LogWS_area2+AVE_EMB+BlueRidge+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+(0+AVE_EMB|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[39]]<-
glmer(yvec~LogWS_area+p_Forest+p_pool+AVE_EMB+(1|species)+(0+LogWS_area|species)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE_EMB|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[40]]<-
glmer(yvec~LogWS_area+p_Forest+p_pool+BlueRidge+(1|species)+(0+LogWS_area|species)+(0+p_Forest|species)+(0+p_pool|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[41]]<-
glmer(yvec~LogWS_area+p_Forest+AVE_EMB+BlueRidge+(1|species)+(0+LogWS_area|species)+(0+p_Forest|species)+(0+AVE_EMB|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[42]]<-
glmer(yvec~LogWS_area+p_pool+AVE_EMB+BlueRidge+(1|species)+(0+LogWS_area|species)+(0+p_pool|species)+(0+AVE_EMB|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[43]]<-
glmer(yvec~p_Forest+p_pool+AVE_EMB+BlueRidge+(1|species)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE_EMB|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[44]]<-
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE_EMB|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[45]]<-
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+BlueRidge+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+(0+p_Forest|species)+(0+p_pool|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
```

#Continued from previous page

```
models[[46]]<-
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+AVE_EMB+BlueRidge+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+(0+p_Forest|species)+(0+AVE_EMB|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[47]]<-
glmer(yvec~LogWS_area+LogWS_area2+p_pool+AVE_EMB+BlueRidge+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+(0+p_pool|species)+(0+AVE_EMB|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))
models[[48]]<-
glmer(yvec~LogWS_area+LogWS_area2+p_Forest+p_pool+AVE_EMB+BlueRidge+(1|species)+(0+LogWS_area|species)+(0+LogWS_area2|species)+(0+p_Forest|species)+(0+p_pool|species)+(0+AVE_EMB|species)+(0+BlueRidge|species)+vec1+vec2+vec3+vec4+vec5+vec6+vec7+vec8+(0+vec1|species)+(0+vec2|species)+(0+vec3|species)+(0+vec4|species)+(0+vec5|species)+(0+vec6|species)+(0+vec7|species)+(0+vec8|species),data=lmer.dat,family=binomial(link="logit"))

all.models<-models
model.name<-rep(NA,48)
tmp<-1:48
model.name[1]<-"null model"
for(i in 2:48){
  model.name[i]<-paste("model.",as.list(tmp[i]),sep="")
  model.table<-array(dim=c(length(all.models),12))
  rownames(model.table)<-model.name
  colnames(model.table)<-c("LnL", "K", "AICc", "delta", "wi", "Intercept", "BlueRidge", "LogWS_area", "LogWS_area2", "p_Forest", "p_pool", "AVE_EMB")
  for(i in 1:length(all.models)){
    model.table[i,1]<-deviance(all.models[[i]])/-2
    model.table[i,2]<-AICc(all.models[[i]],return.K=TRUE)
    model.table[i,3]<-AICc(all.models[[i]],return.K=FALSE)
    model.table[i,6]<-lme4::fixef(all.models[[i]])["(Intercept)"]
    model.table[i,7]<-lme4::fixef(all.models[[i]])["BlueRidge"]
    model.table[i,8]<-lme4::fixef(all.models[[i]])["LogWS_area"]
    model.table[i,9]<-lme4::fixef(all.models[[i]])["LogWS_area2"]
    model.table[i,10]<-lme4::fixef(all.models[[i]])["p_Forest"]
    model.table[i,11]<-lme4::fixef(all.models[[i]])["p_pool"]
    model.table[i,12]<-lme4::fixef(all.models[[i]])["AVE_EMB"]
  }
  for(i in 1:length(all.models)){
    model.table[i,4]<-model.table[i,3]-min(model.table[,3])
  }
  for(i in 1:length(all.models)){
    model.table[i,5]<-exp(-0.5*model.table[i,4])/sum(exp(-0.5*model.table[,4]))
  }
  model.table<-model.table[order(model.table[,3]),]
```