A Hierarchical Community Occurrence Model for North Carolina Stream Fish

Stephen R. Midway[a], Tyler Wagner[b] & Bryn H. Tracy[c]

[a] Pennsylvania Cooperative Fish and Wildlife Research Unit, Pennsylvania State University, 406 Forest Resources Building, University Park, Pennsylvania 16802, USA

[b] U.S. Geological Survey, Pennsylvania Cooperative Fish and Wildlife Research Unit, Pennsylvania State University, 402 Forest Resources Building, University Park, Pennsylvania 16802, USA

[c] North Carolina Department of Environment and Natural Resources, Division of Water Resources, 1621 Mail Service Center, Raleigh, North Carolina 27699, USA
Published online: 30 Aug 2014.

PLEASE SCROLL DOWN FOR ARTICLE

ARTICLE

A Hierarchical Community Occurrence Model for North Carolina Stream Fish

Stephen R. Midway*

Pennsylvania Cooperative Fish and Wildlife Research Unit, Pennsylvania State University,
406 Forest Resources Building, University Park, Pennsylvania 16802, USA

Tyler Wagner

U.S. Geological Survey, Pennsylvania Cooperative Fish and Wildlife Research Unit, Pennsylvania State
University, 402 Forest Resources Building, University Park, Pennsylvania 16802, USA

Bryn H. Tracy

North Carolina Department of Environment and Natural Resources, Division of Water Resources,
1621 Mail Service Center, Raleigh, North Carolina 27699, USA

Abstract

The southeastern USA is home to one of the richest—and most imperiled and threatened—freshwater fish assemblages in North America. For many of these rare and threatened species, conservation efforts are often limited by a lack of data. Drawing on a unique and extensive data set spanning over 20 years, we modeled occurrence probabilities of 126 stream fish species sampled throughout North Carolina, many of which occur more broadly in the southeastern USA. Specifically, we developed species-specific occurrence probabilities from hierarchical Bayesian multispecies models that were based on common land use and land cover covariates. We also used index of biotic integrity tolerance classifications as a second level in the model hierarchy; we identify this level as informative for our work, but it is flexible for future model applications. Based on the partial-pooling property of the models, we were able to generate occurrence probabilities for many imperiled and data-poor species in addition to highlighting a considerable amount of occurrence heterogeneity that supports species-specific investigations whenever possible. Our results provide critical species-level information on many threatened and imperiled species as well as information that may assist with re-evaluation of existing management strategies, such as the use of surrogate species. Finally, we highlight the use of a relatively simple hierarchical model that can easily be generalized for similar situations in which conventional models fail to provide reliable estimates for data-poor groups.

Flowing waters are increasingly viewed as part of the larger landscape they occupy, and together rivers and their landscapes provide a more holistic context in which to study ecosystems. Despite this recognition, the number of threats to and stressors on rivers is growing (Malmqvist and Rundle 2002; Carpenter et al. 2011). In addition to local threats, such as altered flow, pollution, and habitat degradation, we now know that seemingly distant threats like catchment land use and climate change can have devastating impacts on the health of streams.

The numerous pressures that are now placed on rivers frequently manifest in habitat degradation, pollution, and other impacts that collectively threaten freshwater biodiversity (Dudgeon et al. 2006). Although much can be done to ameliorate proximate threats to rivers (e.g., riparian buffers and increased discharge regulation), changes in catchment land use are increasingly considered a major threat (Allan 2004), as nonforested catchment lands alter the quality and fate of surface water (Foley et al. 2005). For example, agricultural land

depletes freshwater, erodes soil, and increases nutrient runoff (as reviewed by Carpenter et al. 2011). Urbanization of catchment land creates problems similar to those generated by agriculture, with the addition of increased stream flashiness (Walsh et al. 2005) and greater volumes of inorganic contaminants (e.g., arsenic and lead; Paul and Meyer 2001).

Within the setting of changing and uncertain threats to streams is the need to develop useful species distribution models that can assist in conservation efforts without relying on comprehensive data. Species distribution models have a long history in ecology, and much work continues on their development and validation (reviewed by Elith and Leathwick 2009). One particular area of species distribution models that has drawn attention for improvement is the reduction in and evaluation of model uncertainty (Elith and Leathwick 2009). Leung and Steele (2013) cautioned that there is no magic bullet for dealing with uncertainty; however, even in data-poor situations, certain species–environment relationships can be improved with very little data. This practice of species distribution models is particularly important in cases of rare and imperiled species—often those species with the most to gain from the application of models to conservation efforts.

Although aquatic species distribution models have evaluated a wide variety of possible environmental effects, the continued investigation of catchment land cover has proven to be informative (e.g., Allan et al. 1997; Gevrey et al. 2009). The negative impacts of catchment deforestation on streams are widely known; however, a growing literature is reporting negative impacts of deforestation at much finer scales. For example, Sutherland et al. (2002) reported elevated sediment levels at base flow conditions in catchments with as little as 22% and 13% deforestation. Lu et al. (2013) found that degradation rates of dissolved organic matter differed based on source; dissolved organic matter originating from agricultural and urban catchments degraded much slower and remained in streams longer than dissolved organic matter originating from forested catchments. At the organismal level, Blevins et al. (2013) reported that riparian land use influenced the stress responses of Creek Chub *Semotilus atromaculatus*, thereby highlighting the sublethal physiological stresses that altered land use can place on fish. Collectively, these studies demonstrate how altered land use impacts the fine-scale physical, chemical, and biological conditions that may influence species distributions.

The southeastern USA is home to one of the richest—and most imperiled and threatened—freshwater fish assemblages in North America (Warren et al. 2000). Although the high degree of endemism may contribute to species vulnerability (Brooks et al. 1992), it also means that the southeastern USA is an area where a variety of land use impacts on streams will first be detected and therefore is a harbinger for other regional fish assemblages. Much work with stream fishes, particularly imperiled species, in the southeastern USA has generated valuable knowledge regarding their conservation status (see Jelks et al. 2008). However, aside from conservation status and

range, conservation efforts rarely have additional information upon which to base conservation action. Often, imperiled stream fish assemblages benefit from specific actions that are designed for priority or surrogate species (Caro and O'Doherty 1999; Chittick et al. 2001). For example, protecting forested land in an effort to improve stream habitat for one species may improve stream habitat for a suite of species. (Although we avoid a discussion on the differences among terms such as "umbrella species," "indicator species," and "flagship species," throughout this study we refer to the underlying concept with the term "surrogate species.") The surrogate species approach is not a poor tactic; however, it operates on the assumption that multiple threatened species benefit from the same conditions. This highlights the potential disconnect between species and habitats: we tend to focus biodiversity thinking at the species level, yet we act at the ecosystem level. Additionally, because it is rare for one species to have a distribution identical to that of another species, there is an inherent violation of the assumption that what is optimal for one species is also optimal for the community.

We recognize that it is unrealistic to collect detailed information on all species in a community and that the surrogate species approach does have utility. However, improved estimates of landscape-related occurrence uncertainty for an entire assemblage of species would greatly improve the evaluation of surrogate species conservation strategies by better characterizing the species-specific heterogeneity of a group or assemblage. In the present study, we used hierarchical Bayesian multispecies models (HBMMs), which provide a number of advantages over conventional models. First, HBMMs constitute one method of estimating information on individual species in situations where comprehensive data are lacking. Specifically, the random effects in hierarchical models expand the scope of inference so that generalizations can be made in a case where a limited sample of populations can yield information representing unobserved populations (Kéry and Schaub 2012). The grouped nature of random effects also eliminates the assumption of independence; thus, estimates for data-poor species are improved through the sharing of group-level information. Of additional importance is the ability of HBMMs to explicitly model and draw inference from multiple ecological levels (Royle and Dorazio 2008). This increases the overall realism of the model in addition to providing a robust and flexible framework in which many covariates may be evaluated.

We used HBMMs to quantify uncertainty in stream fish species presence based on common land use attributes. Drawing from an extensive 20-year sampling effort covering the state of North Carolina, we modeled different a priori species groupings to not only estimate a group response but also to provide species-specific estimates that improve upon existing information and may be used to manage and conserve threatened stream fish assemblages within North Carolina and throughout the southeastern USA.

METHODS

Study area and fish sampling.—The data used in this study came from an ongoing stream sampling program conducted by the North Carolina Department of Environment and Natural Resources (Division of Water Resources, Biological Assessment Branch). Since 1991, a standardized protocol has been used to sample 835 stream reaches throughout North Carolina (Figure 1). Sample sites are wadeable, 183-m (600-ft) reaches. Although sites are sampled approximately once every 5 years, mainly between April and June, we selected only the most recent annual survey for each site. Stream conditions were also relatively similar among samples, reducing concern about fish detection; a regular time of year for all samples meant that water temperatures and conductivity were consistent (B. H. Tracy, unpublished data), and sampling did not occur unless turbidity was very low. Sampling included backpack electrofishing units (most frequently two units) along with an appropriate number of dipnetters based on the stream size. Reaches were sampled using two-pass depletion, covering all available habitats; the first pass moved upstream, and the second pass returned downstream. All fish were collected and identified to species. Unidentifiable individuals of all sizes were preserved in 10% neutral buffered formalin and were identified upon return to the laboratory. Additional programmatic details can be found in the standard operating procedures (NCDENR 2006).

Land use and land cover data.—In addition to mean network catchment elevation and slope, we examined percentages of developed, agricultural, forested, and impervious surfaces in the upstream network catchment. These landscape characteristics were chosen because of their documented effects on fish occurrence. For example, forested catchment cover is positively correlated with the presence of native fish assemblages and endemic fish species. Deforested lands, including developed and agricultural lands, are known to have negative impacts on many native stream fish assemblages and endemic species.
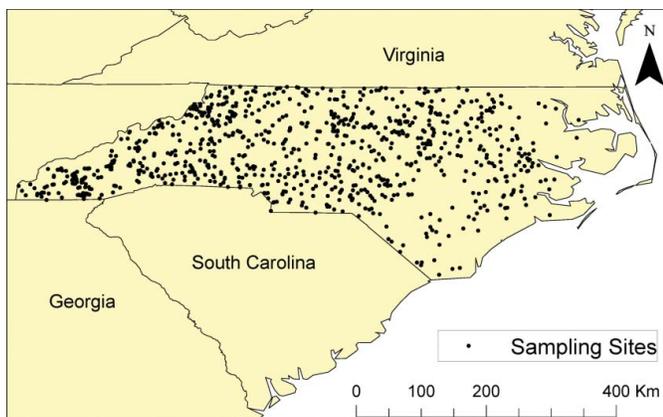
Landscape covariates were quantified for each site at the network catchment level (i.e., the entire upstream catchment of the reach as opposed to the local catchment) and included percentages agricultural land, developed land, and forested land. Data on percentage agricultural land and forested land were from the National Land Cover Database (Homer et al. 2007), and human population density was expressed as the number of people per square kilometer (NOAA 2010).

Data analysis.—We used community models based on detection–nondetection data (i.e., apparent species distribution models; Kéry et al. 2010) to examine the effects of natural and anthropogenic landscape characteristics on the fish community's occurrence probability. We conducted a sensitivity analysis on the lower limit of observations that defined a species' inclusion in our model. Models that were run with cutoffs of 5, 10, and 15 observations yielded nearly identical parameter estimates, suggesting that inclusion or exclusion of infrequently occurring species did not detectably impact the overall model. Therefore, because we were interested in estimating occurrence for rare (i.e., infrequently detected) species, we chose a cutoff of five observations (occurrences > 5) for use in all subsequent modeling. We adopted the HBMM approach to quantify among-species variability in the effects of landscape characteristics on occurrence probability. The modeling framework allowed for the inclusion of site- and species-specific covariates: site-specific covariates were landscape characteristics of the upstream network catchment for each sample site, and the species-specific covariate was a species' tolerance level (intolerant, intermediate, or tolerant) with respect to stream physiochemical habitat and water quality degradation. Tolerance levels were taken directly from the existing North Carolina index of biotic integrity program, which was modified from Karr (1981). The response variable for the analysis was binary, with $y_{(i,j)} = 1$ if species i was detected at site j, and $y_{(i,j)} = 0$ otherwise ($y_{[i,j]} \sim \text{Bernoulli}[\pi_{i,j}]$). The general form of the model was

$$\text{logit}\left(\pi_{i,j}\right) = \beta_{(0,i)} + \beta_{(1,i)} \cdot X_j,$$

where

$$\beta_{(0,i)} \sim \text{N}(\gamma_0^0 + \gamma_1^0 \cdot Z_{1,i} + \gamma_2^0 \cdot Z_{2,i}, \sigma_{\beta 0}^2),$$
$$\beta_{(1,i)} \sim \text{N}(\gamma_0^1 + \gamma_1^1 \cdot Z_{1,i} + \gamma_2^1 \cdot Z_{2,i}, \sigma_{\beta 1}^2),$$

$\beta_{(0,i)}$ is the species-specific intercept, and $\beta_{(1,i)}$ is the species-specific effect of the site-specific landscape covariate X_j on the logit probability of occurrence for species i (all covariates were standardized prior to analysis: $[X_j - X]/\text{SD}[X]$). The species-specific intercepts and slopes were modeled as normally distributed random effects and as a function of habitat tolerance level, for which there were three categories (intolerant, intermediate, and tolerant) and two covariates

FIGURE 1. Map of stream fish sampling sites ($n = 835$) in North Carolina. [Color figure available online.]

($Z_{1,i} = 1$ if species i is intermediate, and $Z_{1,i} = 0$ otherwise; $Z_{2,i} = 1$ if species i is intolerant, and $Z_{2,i} = 0$ otherwise; tolerant species was the reference cell). Thus, γ_0^x is the grand mean intercept or slope for tolerant species; γ_1^x and γ_2^x are the effects of intermediate species and intolerant species, respectively.

Vague normal priors (N $\sim$ [0, 1,000]) were used for all slope and intercept parameters, and vague uniform priors (Uniform $\sim$ [0, 10]) were used for $\sigma_{\beta0}$ and $\sigma_{\beta1}$. The fish survey sampling design did not allow for the estimation of detection probability (i.e., repeat visits to sites within a season were not performed; data from the two electrofishing passes were pooled). As such, we recognize that the effects of landscape covariates on species occurrence may be biased if detection probability is less than 1.0 (Gu and Swihart 2004), resulting in underestimates of the effects of covariates for some species (Tyre et al. 2003). The fish community data, however, were from surveys that were performed with the specific goal of assessing the entire fish community, and sampling followed standardized methods by trained field crews. Furthermore, studies have suggested that in many cases, stream reach lengths of 235–555 m (reaches in our study totaled 366 m) are sufficient for presence–absence sampling (Paller 1995). Therefore, efforts were made to minimize the possibility of making false-negative errors (i.e., recording a species as absent when it was in fact present). All models were fitted using WinBUGS version 1.4 (Spiegelhalter et al. 2004).

RESULTS

In total, 126 fish species from 835 stream sites were included in the multispecies hierarchical modeling. Of the 126 species, there were 27 intolerant species, 84 intermediate species, and 15 tolerant species (see Supplementary Table S.1 in the online version of this article for the full list of species). Of the landscape covariates considered, our final covariates for analysis included the percentages of developed, agricultural, and forested lands in the upstream network catchment. These land use and land cover types were chosen because (1) they were correlated with other landscape metrics (e.g., percentage developed land and percentage impervious surface: Spearman's rank correlation coefficient $r = 0.96$; percentage forested land and mean slope: $r = 0.82$) and (2) they represent anthropogenic (developed and agricultural lands) and natural (forested land) landscape characteristics that were hypothesized to structure aquatic communities, mediate the invasion and spread of tolerant species, and mediate the loss and range contraction of intolerant species across the landscape. The percentage forested land in the upstream network catchment (mean $\pm$ SD) was $57 \pm 25\%$ (range $= 0.6$–100%). The percentages developed land and agricultural land (mean $\pm$ SD) were $13 \pm 18\%$ (range $= 0.1$–98%) and $19 \pm 15\%$ (range $= 0.0$–66%), respectively.

Multispecies Hierarchical Modeling

Two separate multispecies models were fitted: one that included the percentages agricultural land and developed land as covariates; and one that contained only forested land as a covariate. The three covariates could not be included in a single model because of the relatively high correlation between the percentage forested land and the percentage developed land ($r = 0.56$) or agricultural land ($r = 0.57$). The r-value for the correlation between percentage developed land and percentage agricultural land was -0.20.

Across all intolerant species, the effect of percentage forested land in the network catchment was positive (posterior mean $= 0.54$; 95% credible interval [CI] $= 0.08$–0.99). In contrast, the effect of percentage forested land on intermediate species did not differ from zero (posterior mean $= 0.03$; 95% CI $= -0.44$ to 0.47) and the effect on tolerant species was negative (posterior mean $= -0.74$, 95% CI $= -1.21$ to -0.30; Figure 2). However, it was not until forested land reached approximately 70% and 90% that the probability of occurrence for intermediate and intolerant species, respectively, became higher than the probability of occurrence for tolerant species. The effects of percentages developed land and agricultural land were examined while holding the other land use type at the minimum value (i.e., the effect of developed land was assessed while holding agricultural land at 0%; the effect of agricultural land was assessed while holding developed land at 0.1%). The effects of developed land and
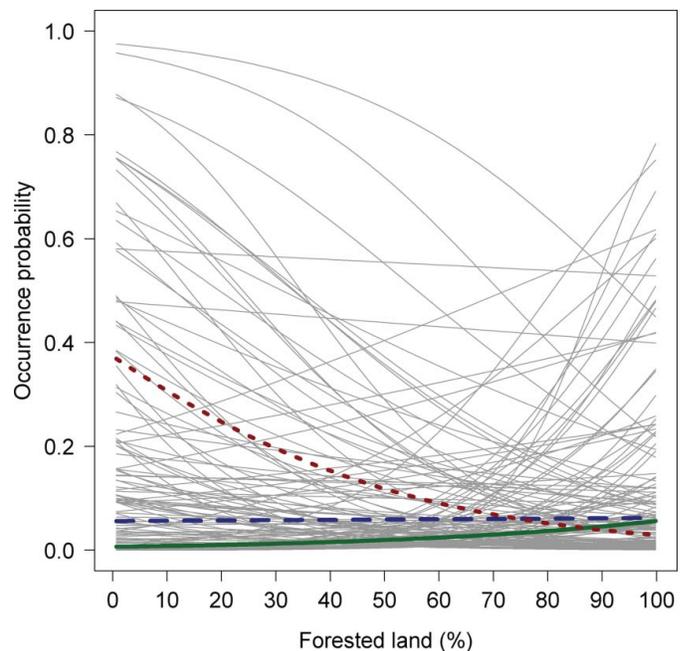

FIGURE 2. Species-specific occupancy probabilities (thin gray lines) in response to percentage forested land in the upstream network catchment; mean responses for intolerant (solid green line), intermediate (dashed blue line), and tolerant (dotted red line) fish species are also presented. [Color figure available online.]

agricultural land on intolerant species were negative (developed land: posterior mean = −0.75, 95% CI = −0.99 to −0.50; agricultural land: posterior mean = −0.37, 95% CI = −0.64 to 0.01); although the 95% CI overlapped zero for the effect of agricultural land, the 90% CI (−0.58 to −0.05) did not overlap zero. The effect of percentage developed land on intermediate species was relatively weak and negative (posterior mean = −0.21; 95% CI = −0.45 to 0.04; 90% CI = −0.41 to −0.002), whereas the 95% CI for the effect of percentage agricultural land overlapped zero (posterior mean = 0.04; 95% CI = −0.29 to 0.34).

The effects of developed land and agricultural land on tolerant species were positive (developed land: posterior mean = 0.46, 95% CI = 0.22–0.71; agricultural land: posterior mean = 0.55, 95% CI = 0.22–0.87; Figures 3, 4). The probability of occurrence for tolerant species was lower than those for intolerant species and intermediate species at very low percentages of developed land (less than ∼5%). However, once developed land reached approximately 10% and 20%, the probability of occurrence for tolerant species exceeded the probabilities of occurrence for intolerant species and intermediate species, respectively. A similar pattern was observed for the effect of percentage agriculture in the network catchment; however, agricultural land had to exceed approximately 10% and 20% for the occurrence probability of tolerant species to exceed those of intolerant species and intermediate species, respectively (Figures 3, 4).

FIGURE 4. Species-specific occupancy probabilities (thin gray lines) in response to percentage agricultural land in the upstream network catchment; mean responses for intolerant (solid green line), intermediate (dashed blue line), and tolerant (dotted red line) fish species are also presented. [Color figure available online.]

Individual species occurrence curves and estimates are presented in Supplementary Figure S.1 and Table S.1; however, to highlight heterogeneity in occurrence probability, we present results from nine species (three species from each tolerance level; Figures 5–7). The species in each group represent those with some of the largest sample sizes, thus allowing for comparisons to highlight the heterogeneity in occurrence probabilities and to reduce uncertainty associated with some infrequently sampled species. Occurrence for select intolerant species (Highback Chub *Hybopsis hypsinotus*, Piedmont Darter *Percina crassa*, and Roanoke Darter *Percina roanoka*) was low for all three land use covariates, although this was expected based on their rare occurrence or limited spatial ranges. Less expected was the variety in the direction of slopes for the covariates. For example, the Highback Chub and Piedmont Darter exhibited positive slopes in response to increasing forest land cover, while the Roanoke Darter exhibited a negative slope as forest land cover increased. Responses to developed and agricultural lands also included positive and negative slopes (slope coefficients and associated 95% CIs are presented in Table S.1).

We also report on three select intermediate species (Bluegill *Lepomis macrochirus*, Bluehead Chub *Nocomis leptocephalus*, and Tessellated Darter *Etheostoma olmstedi*) and three tolerant species (Redbreast Sunfish *Lepomis auritus*, Creek Chub, and White Sucker *Catostomus commersonii*). In both of these tolerance groupings, the species showed clear differences in occurrence probability relative to the same covariate,
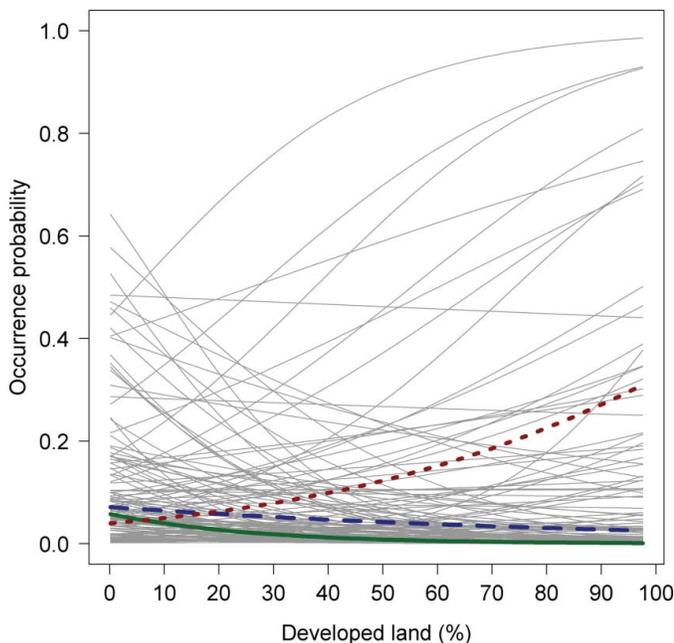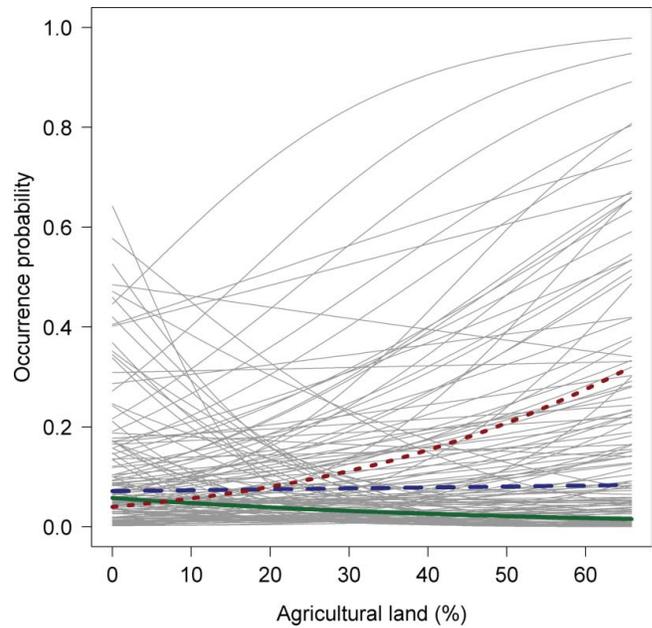

FIGURE 3. Species-specific occupancy probabilities (thin gray lines) in response to percentage developed land in the upstream network catchment; mean responses for intolerant (solid green line), intermediate (dashed blue line), and tolerant (dotted red line) fish species are also presented. [Color figure available online.]
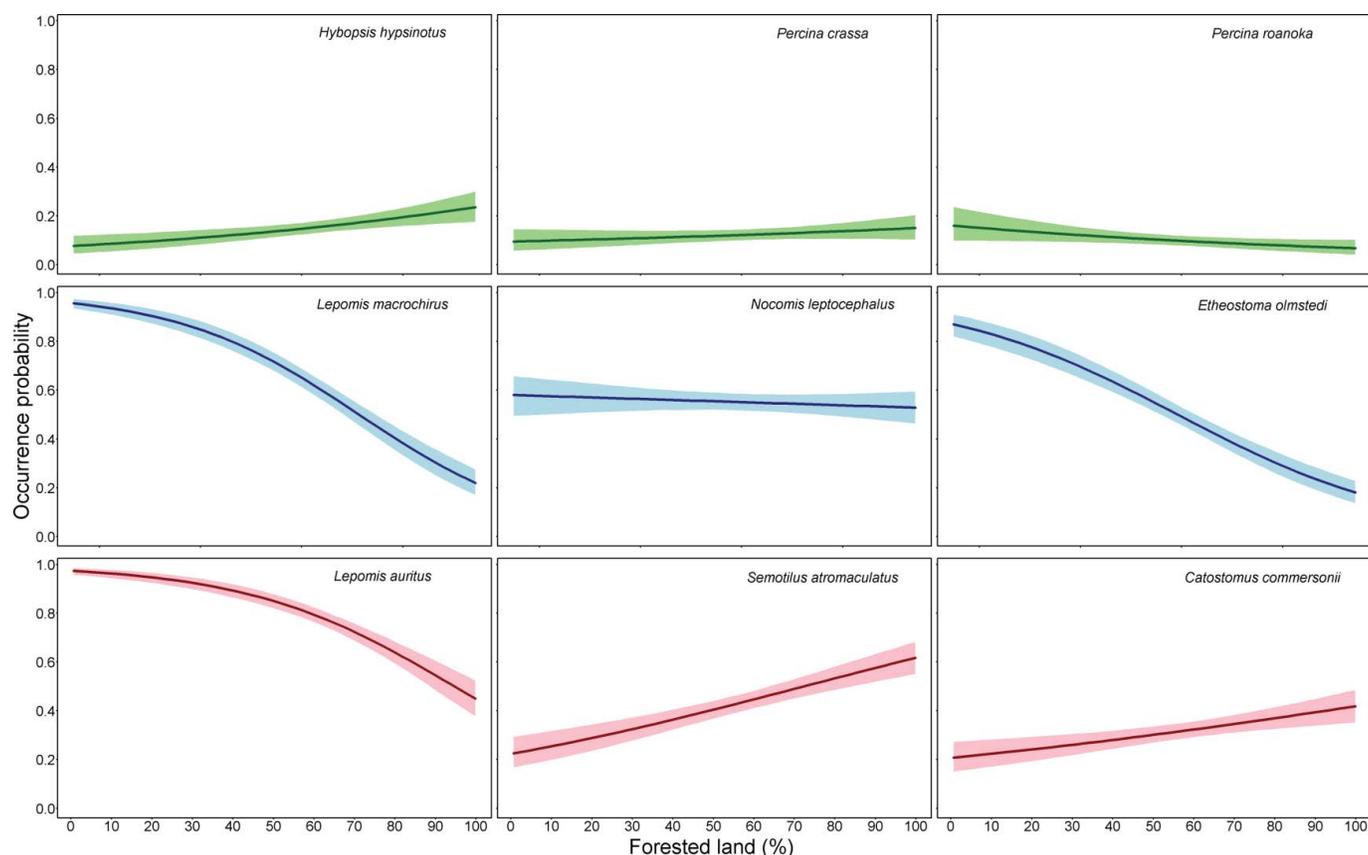
Figure 5. Responses of select intolerant (top row), intermediate (middle row), and tolerant (bottom row) fish species to percentage forested land in the upstream network catchment. Solid lines are posterior means; shaded regions are 95% credible intervals (*Hybopsis hypsinotus* = Highback Chub; *Percina crassa* = Piedmont Darter; *Percina roanoka* = Roanoke Darter; *Lepomis macrochirus* = Bluegill; *Nocomis leptocephalus* = Bluehead Chub; *Etheostoma olmstedi* = Tessellated Darter; *Lepomis auritus* = Redbreast Sunfish; *Semotilus atromaculatus* = Creek Chub; *Catostomus commersonii* = White Sucker). [Color figure available online.]

although some of this can be attributed to both larger sample sizes and occurrence across a greater range of covariates than was observed for most of the intolerant species. The HBMM's influence on low sample sizes was largely absent for the well-sampled intermediate and tolerant species we highlight; therefore, the heterogeneity in slope direction and magnitude within the groupings represents well-estimated results (as opposed to interpreting heterogeneity in uncertain results).

DISCUSSION

Although we expected heterogeneity in occurrence probability among tolerance groups, we found a substantial amount of heterogeneity within tolerance groups in response to all three landscape covariates. This information can be used to improve both species-level and assemblage-level responses to land use. It may be unrealistic to think that management and conservation efforts will soon place equal priority on all stream species; therefore, our occurrence models for rare and intolerant species can provide information for re-valuation of surrogate species management—the strategy currently used by

the U.S. Fish and Wildlife Service. Occurrence patterns for intermediate species in our study were more similar to those of intolerant species than to those of tolerant species. Although our model does not suggest why this might be the case, we can begin to hypothesize that land use impacts tolerant species differently than intermediate and intolerant species. For example, tolerant species could comprise a greater proportion of introduced species, which include more generalists than other groups. Other studies have shown that introduced species can be linked to landscape-level characteristics (Lapointe and Light 2012) and that invasive species may be more suited to anthropogenically disturbed habitat (but for insects: Grez et al. 2013). Either explanation or both explanations could support the unique occurrence probability patterns we found.

We also found that it took a relatively small amount of land use change for tolerant species to become the dominant species group (i.e., most likely to occur). Tolerant species were projected to be the most common species by the time 20% of catchment cover was either agricultural land or developed land. Particularly concerning was the fact that tolerant species became more common than intolerant species when
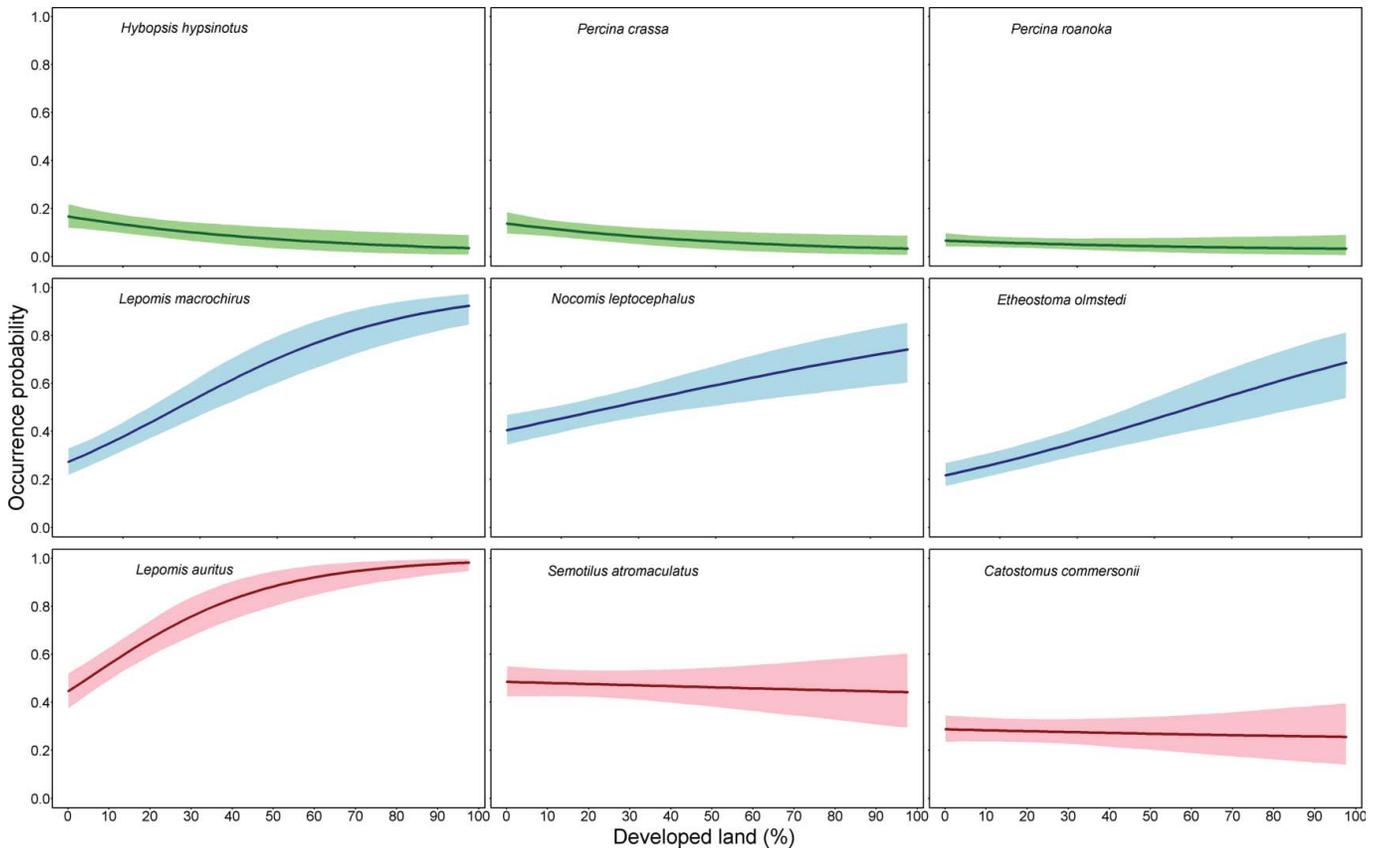
FIGURE 6.   Responses of select intolerant (top row), intermediate (middle row), and tolerant (bottom row) fish species to percentage developed land in the upstream network catchment. Solid lines are posterior means; shaded regions are 95% credible intervals. Common names of species are defined in Figure 5. [Color figure available online.]

anthropogenic land use in the catchment was as little as 5%. These occurrence probability curves are reporting on a variety of species from different sites; however, they do permit us to better quantify the expected relative occurrences of different tolerance groups. Our species groupings may also be responding in relation to a landscape threshold, which has been a demonstrated response in other studies.

Model Strengths and Limitations

There is often no shortage of environmental covariates from which to choose when modeling species distributions. Due to the scale of our study and the available data, land cover and land use covariates were the most appropriate, and we benefited from the fact that land cover should be particularly robust, as all streams have terrestrial catchments with varying amounts of forest, agriculture, and development. By investigating basic covariates, we also did not risk overfitting our models, and therefore our results should generalize to systems outside of those we investigated. Of course, basic land use covariates may not capture all of the complex mechanisms structuring fish assemblages, but they are known to affect the distribution of many fish species (regardless of mechanism),

and they represent landscape characteristics that can be managed to some degree in many systems and thus are useful for rare species conservation and management.

The use of hierarchical models allowed for the inclusion of species for which there was little information about their response to anthropogenic alterations in the landscape; we were able to include these species by borrowing information from the entire community through the use of a species random effect. This capability, referred to as partial pooling, is important because limited information on distribution or response to landscape alterations is commonplace for many species; however, conservation decision-making often cannot be postponed until more data become available (Leung and Steele 2013). In contrast, estimating species-specific model parameters using only species-specific data (i.e., no pooling) would be difficult if not impossible because a wide range of land use values is required to reveal the effects of land use on species occurrences. Our results can also be a hypothesis-generating tool for identifying and prioritizing species for which little information is known or that are of interest for other reasons. Additionally, the use of Bayesian estimation provides probabilistic interpretation of results, making it a better tool for ranking management alternatives. This may be useful in a
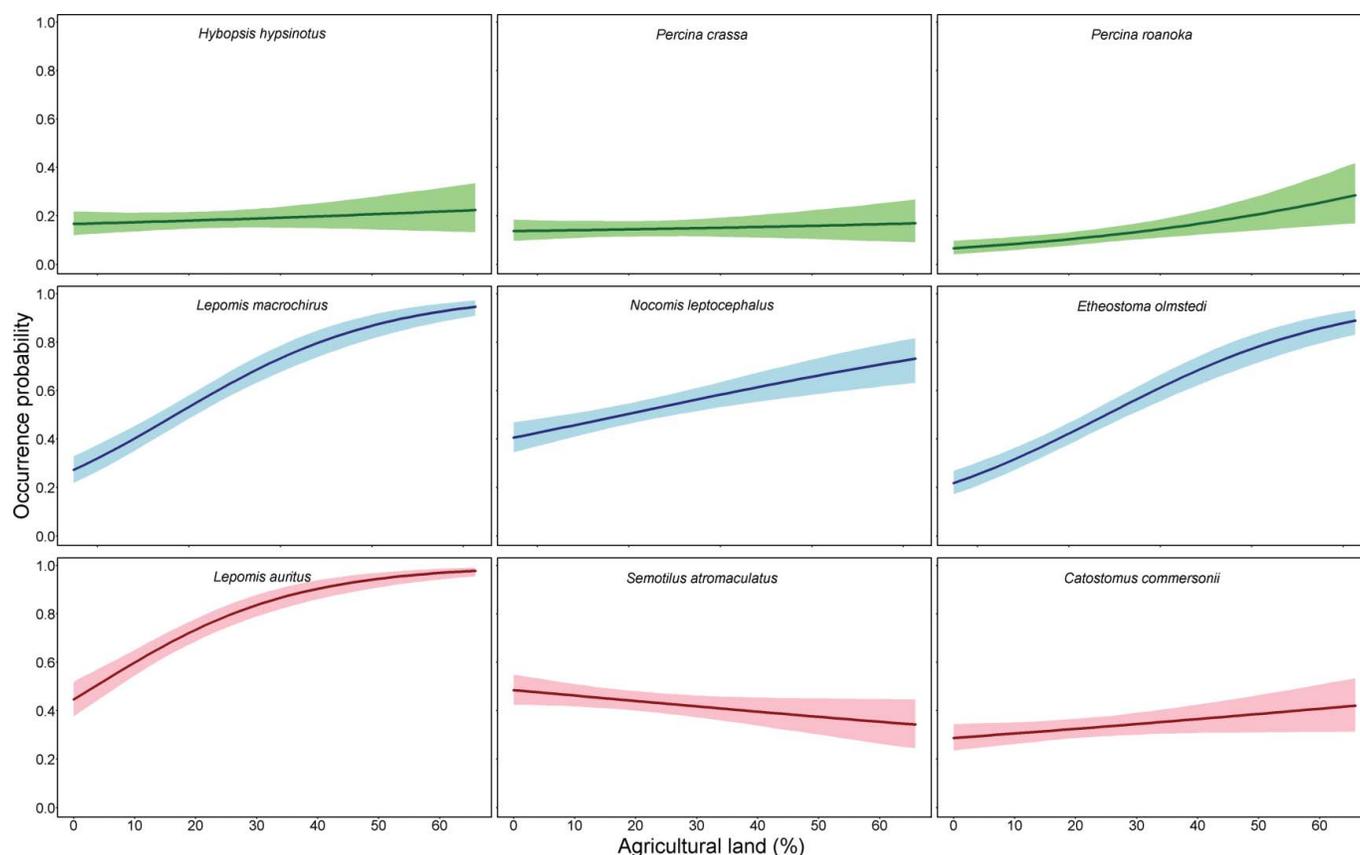
Figure 7.   Responses of select intolerant (top row), intermediate (middle row), and tolerant (bottom row) fish species to percentage agricultural land in the upstream network catchment. Solid lines are posterior means; shaded regions are 95% credible intervals. Common names of species are defined in Figure 5. [Color figure available online.]

management scenario where only a fixed number of species or habitats will receive management or conservation resources. The probabilistic results of our model create a situation in which species can be ranked based on estimates of effects and uncertainty, regardless of statistical significance.

In addition to the borrowed strength that informs the species-specific occurrence probabilities, the second level of the HBMM can be modified for any covariate that is appropriate for a specific conservation question. We used tolerance levels from the North Carolina index of biotic integrity due to their wide applicability and inclusion of a variety of information, whereas future applications of this model could explore a range of covariates and would be adaptable to those covariates that are most appropriate for the species or system in question. Ricketts et al. (1999) pointed out that no taxon or indexing is without error but that continual improvement of species information and geographic patterns will produce more informative combinations of assemblages and perhaps better surrogate species. In this way, our study contributes to the growing list of efforts to better describe communities and surrogate species through developments in species distribution models (e.g., Azeria et al. 2009; Meador and Carlisle 2009; Ovaskainen and Soininen 2011).

One limitation to our approach is that land use correlations and issues of scale prevent explicit conclusions about causal mechanisms. Frequently, the effects of agriculture and development are similar in streams (e.g., increased sediment load, greater nutrient and pollution loads, and changes in stream morphology; Allan 2004). Further compounding this is the variable effect of scale on a stream reach. Because streams and catchments are hierarchies of smaller units and because large-scale features may constrain the development of smaller features (particularly within stream), it can be difficult to partition the influence of different factors at different scales (Lammert and Allan 1999). Despite these challenges, hierarchical models are useful in characterizing some of the spatial variability that presently confounds the covariate effects of species distribution models (Wagner et al. 2006).

Another factor possibly influencing our results is the presence of introduced species. Introduced species are often generalists (Olden et al. 2004), although this attribute is somewhat accounted for in the tolerance framework we adopted. In fact, many of the tolerant species we modeled were introduced species; however, the statewide scale of our investigation would not have been appropriate for categorizing introduced species because the native–introduced factor operates at much smaller

scales. Perhaps more importantly, conservation typically needs to address what is present in a community; therefore, modeling the factor of native versus introduced would have addressed a different question. We sought to provide estimates that reflect current communities rather than introduction potential.

Despite the overall quality of our data, we recognize that without detection probabilities, our results are limited to producing apparent species distributions (Kéry et al. 2010). Based on the consistency of our sampling methods and the inferred high detection rates, we still caution against the use of our results to infer true occupancy. Additionally, caution must be taken when interpreting species-specific patterns from multi-species models (Zipkin et al. 2010). To this end, we suggest that the primary use of our models should be to highlight the broad-scale relationships between species presence and landscape features rather than to predict a given species' occurrence at a single site. In fact, in some cases of single-species occurrence prediction, we might anticipate low or variable accuracy based on the relatively few predictors used in our models and the inherent complexities required for fine-scale prediction (Wagner et al. 2014). Although we could have chosen to fit a model with a large number of covariates in order to maximize prediction (even if certain predictors' effects could not be explained), we opted to limit predictors and thus maximize our understanding of species–landscape relationships (Kuhn and Johnson 2013).

Species distribution models have a long history in ecology and continue to be developed and improved (Elith and Leathwick 2009). Our investigation was focused on streams in North Carolina; however, not only is the model easily adapted for other species, covariates, and groupings, but many of the estimates we calculated could likely be applied in other southeastern U.S. stream habitats for which species data are unavailable. Our approach posits a useful combination of existing groups with quantified species information. For example, rather than developing conservation strategies solely on a single imperiled (e.g., surrogate) species or purely on a biodiversity index (e.g., species richness), our method provides the opportunity to quantify responses for nearly all species in an assemblage. Existing frameworks like tolerance levels, conservation status, or reproductive guild can be implemented into our approach, and the resulting estimates can be used to develop new conservation priorities. For many species, we currently await conservation efforts that require the best possible information, knowing that such information will be incomplete. Our approach uses existing information to better quantify and reduce uncertainty in predicting the presence of a limitless list of species over a range of possible predictors.

REFERENCES

Allan, D., D. Erickson, and J. Fay. 1997. The influence of catchment land use on stream integrity across multiple spatial scales. Freshwater Biology 37:149–161.

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.

Azeria, E. T., D. Fortin, C. Hébert, P. Peres-Neto, D. Pothier, and J.-C. Ruel. 2009. Using null model analysis of species co-occurrences to deconstruct biodiversity patterns and select indicator species. Diversity and Distributions 15:958–971.

Blevins, Z., E. Effert, D. Wahl, and C. Suski. 2013. Land use drives the physiological properties of a stream fish. Ecological Indicators 24:224–235.

Brooks, D. R., R. L. Mayden, and D. A. McLennan. 1992. Phylogeny and biodiversity: conserving our evolutionary legacy. Trends in Ecology and Evolution 7:55–59.

Caro, T. M., and G. O'Doherty. 1999. On the use of surrogate species in conservation biology. Conservation Biology 13:805–814.

Carpenter, S. R., E. H. Stanley, and M. J. Vander Zanden. 2011. State of the world's freshwater ecosystems: physical, chemical, and biological changes. Annual Review of Environment and Resources 36:75–99.

Chittick, B., M. Stoskopf, N. Heil, J. Levine, and M. Law. 2001. Evaluation of Sandbar Shiner as a surrogate for assessing health risks to the endangered Cape Fear Shiner. Journal of Aquatic Animal Health 13:86–95.

Dudgeon, D., A. H. Arthington, M. O. Gessner, Z.-I. Kawabata, D. J. Knowler, C. Lévêque, R. J. Naiman, A.-H. Prieur-Richard, D. Soto, and M. L. Stiassny. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biological Reviews 81:163–182.

Elith, J., and J. R. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics 40:677–697.

Foley, J. A., R. DeFries, G. P. Asner, C. Barford, G. Bonan, S. R. Carpenter, F. S. Chapin, M. T. Coe, G. C. Daily, and H. K. Gibbs. 2005. Global consequences of land use. Science 309:570–574.

Gevrey, M., F. Sans-Piche, G. Grenouillet, L. Tudesque, and S. Lek. 2009. Modeling the impact of landscape types on the distribution of stream fish species. Canadian Journal of Fisheries and Aquatic Sciences 66:484–495.

Grez, A. A., T. A. Rand, T. Zaviezo, and F. Castillo-Serey. 2013. Land use intensification differentially benefits alien over native predators in agricultural landscape mosaics. Diversity and Distributions 19:749–759.

Gu, W., and R. K. Swihart. 2004. Absent or undetected? Effects of non-detection of species occurrence on wildlife–habitat models. Biological Conservation 116:195–203.

Homer, C., J. Dewitz, J. Fry, M. Coan, N. Hossain, C. Larson, N. Herold, A. McKerrow, J. N. VanDriel, and J. Wickham. 2007.

Completion of the 2001 National Land Cover Database for the conterminous United States. Photogrammetric Engineering and Remote Sensing 73:337–341.

Jelks, H. L., S. J. Walsh, N. M. Burkhead, S. Contreras-Balderas, E. Diaz-Pardo, D. A. Hendrickson, J. Lyons, N. E. Mandrak, F. McCormick, J. S. Nelson, S. P. Plantania, B. A. Porter, C. B. Renaud, J. J. Schmitter-Soto, E. B. Taylor, and M. L. Warren Jr. 2008. Conservation status of imperiled North American freshwater and diadromous fishes. Fisheries 33:372–407.

Karr, J. R. 1981. Assessment of biotic integrity using fish communities. Fisheries 6(6):21–27.

Kéry, M., B. Gardner, and C. Monnerat. 2010. Predicting species distributions from checklist data using site-occupancy models. Journal of Biogeography 37:1851–1862.

Kéry, M., and M. Schaub. 2012. Bayesian population analysis using WinBUGS: a hierarchical perspective. Academic Press, San Diego, California.

Kuhn, M., and K. Johnson. 2013. Applied predictive modeling. Springer, New York.

Lammert, M., and J. D. Allan. 1999. Assessing biotic integrity of streams: effects of scale in measuring the influence of land use/cover and habitat structure on fish and macroinvertebrates. Environmental Management 23:257–270.

Lapointe, N. W., and T. Light. 2012. Landscape-scale determinants of non-native fish communities. Diversity and Distributions 18:282–293.

Leung, B., and R. J. Steele. 2013. The value of a datum–how little data do we need for quantitative risk analysis? Diversity and Distributions 19:617–628.

Lu, Y., J. E. Bauer, E. A. Canuel, Y. Yamashita, R. Chambers, and R. Jaffé. 2013. Photochemical and microbial alteration of dissolved organic matter in temperate headwater streams associated with different land use. Journal of Geophysical Research: Biogeosciences 118:566–580.

Malmqvist, B., and S. Rundle. 2002. Threats to the running water ecosystems of the world. Environmental Conservation 29:134–153.

Meador, M. R., and D. M. Carlisle. 2009. Predictive models for fish assemblages in eastern U.S. streams: implications for assessing biodiversity. Transactions of the American Fisheries Society 138:725–740.

NCDENR (North Carolina Department of Environment and Natural Resources). 2006. Standard operating procedure for stream fish communities. NCDENR, Raleigh.

NOAA (National Oceanic and Atmospheric Administration). 2010. Development sprawl impacts on the terrestrial carbon dynamics of the United States: data download. NOAA, Silver Spring, Maryland. Available: www.ngdc.noaa.gov/dmsp/download_sprawl.html. (June 2014).

Olden, J. D., L. N. Poff, M. R. Douglas, M. E. Douglas, and K. D. Fausch. 2004. Ecological and evolutionary consequences of biotic homogenization. Trends in Ecology and Evolution 19:18–24.

Ovaskainen, O., and J. Soininen. 2011. Making more out of sparse data: hierarchical modeling of species communities. Ecology 92:289–295.

Paller, M. H. 1995. Relationships among number of fish species sampled, reach length surveyed, and sampling effort in South Carolina coastal plain streams. North American Journal of Fisheries Management 15:110–120.

Paul, M. J., and J. L. Meyer. 2001. Streams in the urban landscape. Annual Review of Ecology and Systematics 32:333–365.

Ricketts, T. H., E. Dinerstein, D. M. Olson, and C. Loucks. 1999. Who's where in North America? Bioscience 49:369–381.

Royle, J. A., and R. M. Dorazio. 2008. Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities. Academic Press, New York.

Spiegelhalter, D., A. Thomas, N. Best, and D. Lunn. 2004. WinBUGS user manual. Medical Research Council, Biostatistics Unit, Cambridge, UK.

Sutherland, A. B., J. L. Meyer, and E. P. Gardiner. 2002. Effects of land cover on sediment regime and fish assemblage structure in four southern Appalachian streams. Freshwater Biology 47:1791–1805.

Tyre, A. J., B. Tenhumberg, S. A. Field, D. Niejalke, K. Parris, and H. P. Possingham. 2003. Improving precision and reducing bias in biological surveys: estimating false-negative error rates. Ecological Applications 13:1790–1801.

Wagner, T., J. T. Deweber, J. Detar, D. Kristine, and J. A. Sweka. 2014. Spatial and temporal dynamics in Brook Trout density: implications for population monitoring. North American Journal of Fisheries Management 34:258–269.

Wagner, T., D. B. Hayes, and M. T. Bremigan. 2006. Accounting for multilevel data structures in fisheries data using mixed models. Fisheries 31:180–187.

Walsh, C. J., A. H. Roy, J. W. Feminella, P. D. Cottingham, P. M. Groffman, and R. P. Morgan II. 2005. The urban stream syndrome: current knowledge and the search for a cure. Journal of the North American Benthological Society 24:706–723.

Warren, M. L., B. M. Burr, S. J. Walsh, H. L. Bart Jr., R. C. Cashner, D. A. Etnier, B. J. Freeman, B. R. Kuhajda, R. L. Mayden, H. W. Robinson, S. T. Ross, and W. C. Starnes. 2000. Diversity, distribution, and conservation status of the native freshwater fishes of the southern United States. Fisheries 25(10):7–31.

Zipkin, E. F., J. A. Royle, D. K. Dawson, and S. Bates. 2010. Multispecies occurrence models to evaluate the effects of conservation and management actions. Biological Conservation 143:479–484.