Modeling freshwater fish distributions using multiscale landscape data: A case study of six narrow range endemics

Robert L. Hopkins II*, Brooks M. Burr

Department of Zoology, Southern Illinois University, 1125 Lincoln Drive, LS II, Carbondale, IL 62901, United States

A B S T R A C T

Species distribution models (SDMs) have become integral tools in scientific research and conservation planning. Despite progress in the assessment of various statistical models for use in SDMs, little has been done in way of evaluating appropriate ecological models. In this paper, we evaluate the multiscale filter framework as a suitable theoretical model for predicting freshwater fish distributions in the upper Green River system (Ohio River drainage), USA. The spatial distributions of six fishes with contrasting biogeographies were modeled using boosted regression trees and multiscale landscape data. Species biogeography did not appear to affect predictive performance and all models performed well statistically with receiver operating characteristic area under the curve (AUC) ranging from 0.87 to 0.98. Predictive maps show accurate estimations of ranges for five of six species based on historical collections. The relative influence of each type of environmental feature and spatial scale varied markedly with between species. A hierarchical effect was detected for narrowly distributed species. These species were highly influenced by soil composition at larger spatial scales and land use/land cover (LULC) patterns at more proximal scales. Conversely, LULC pattern was the most influential feature for widely distributed at all spatial scales. Using multiscale data capable of capturing hierarchical landscape influences allowed production of accurate predictive models and provided further insight into factors controlling freshwater fish distributions.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Species distribution models (SDMs) have become a fundamental component of scientific research and are used for investigating species–environment relationships (Austin, 2002; Guisan and Thuiller, 2005), ecological forecasting (Clark et al., 2001), and conservation planning (Moilanen et al., 2008). Principally, SDMs are correlational studies in which species occurrences are modeled to a set of environmental variables (Austin, 2007) with the ultimate objective of maximizing predictive accuracy and predictive mapping (Joy and De'Ath, 2004; Franklin et al., 2009; Marmion et al., 2008). Species distribution modeling consists of three chief components, these are: an ecological model, a data model, and finally a statistical model (Austin, 2002; Guisan and Thuiller, 2005). The apex of this framework is the ecological model (Huston, 2002). However, the theoretical underpinnings of SDMs are rarely given adequate consideration (Guisan and Thuiller, 2005; Austin, 2007) and largely consist of a loosely defined environmental niche concept—an implicit assumption that environmental variables, in part, control species distributions (Leibold, 1995). Moreover, environmental predictors are often selected haphaz-

ardly, usually on the basis of availability instead of ecological relevance. The result is, as Guisan and Thuiller (2005) state, "a weakening of the overall approach". Anchoring SDM research in an explicit conceptual framework is necessary to improve quantification of species–environment relationships, allow descriptions of functional linkages, and increase applicability to conservation planning.

Hynes (1975) first articulated the importance of landscape-level attributes in shaping stream conditions and, ultimately, faunal distributions. Vannote et al. (1980), using an energy equilibrium framework and macroinvertebrate communities as the exemplar, demonstrated the predictability of biological assemblages along the stream continuum. Later, Poff (1997) presented the multiscale filtering framework which posits that aquatic fauna distributions are regulated by environmental features (e.g., land use/land cover (LULC), geology, soil composition, climate) operating at different spatial and temporal scales. Subsequent studies (e.g., Allan et al., 1997; Lammert and Allan, 1999; Schrank et al., 2001; Scott et al., 2002; Weigel et al., 2003) provided empirical evidence demonstrating multiscale environmental influences on biotic assemblages and physicochemical features of stream ecosystems. Similarly, MacKey and Lindenmayer (2001) present a conceptual framework for such hierarchical modeling of terrestrial species. Using Leadbeater's Possum as the exemplar, the authors demonstrate how climatic and geologic factors control the broad-scale geographical limits of a

---

* Corresponding author. Tel.: +1 618 453 4113; fax: +1 618 453 2806.
E-mail address: rhopkins@siu.edu (R.L. Hopkins II).

species while vegetation composition and structure influence the distributional patterns at more local scales.

More recently, researchers (e.g., Olden and Jackson, 2002; Joy and De'Ath, 2004; Leathwick et al., 2005; Oakes et al., 2005) have built predictive models of fish distributions using multiscale environmental variables. However, usually only predictive performance of various statistical models are analyzed with minimal consideration of model structure or species–environment relationships. Moreover, the environmental predictors used are often an eclectic combination of putative environmental correlates chosen purely on the basis of availability. Despite much evidence suggesting hierarchical control of aquatic species distributions (Poff, 1997), this approach has not been used in predictive modeling. In recognition of the necessity for a hierarchical approach, Hopkins (in press) evaluated the use of multiscale landscape data in aquatic species distribution modeling. Using a case study of a freshwater mussel, the author demonstrated that the influences of land cover and geomorphic variables are often manifested at different spatial scales. Furthermore, the use of spatial pattern metrics as predictor variables was also examined. In a study of a California watershed, Kearns et al. (2005) showed that measures of LULC spatial pattern (e.g. patch density, mean patch size, contagion, etc.) explained more variation in water quality measures than LULC composition. Similarly, Hopkins (in press) observed that measures of LULC pattern were the most influential variables in modeling the distribution of a freshwater mussel. However, geomorphic variables were also highly influential and interrogation of the model revealed that these two types of variables were explaining different components of the species distribution. The aforementioned study examined only a single species—to our knowledge there has been no studies evaluating of the use of multiscale landscape data for predictive modeling of freshwater fish using multiple species.

In this paper, we evaluate the use of the multiscale filter framework as a valid ecological model for predicting fish distributions.

For the data model, we focus on landscape variables quantified at three spatial scales because, as Austin (2002) notes, landscape data are easily quantified in a geographical information system (GIS) and permit predictive mapping across large spatial extents—an important application for conservation planning. Also, we employ environmental variables that capture historical and contemporary attributes of the landscape. The incorporation of geomorphic variables enables models to capture aspects of historical constraints on species distributions (Burr and Page, 1986). Moreover, these variables exhibit a high degree of positive spatial-autocorrelation and, when quantified at larger scales, provide spatially explicit information and function as a surrogate for stream size and network position (Joy and De'Ath, 2004). The use of environmental predictors that are static over long periods of time will help address the requisite assumption that modeled species are in pseudo-equilibrium with the environment (Austin, 2007). In particular, we examine if predictive modeling anchored within the multiscale filter framework is applicable to fish species with contrasting biogeographies. Using boosted regression trees (Elith et al., 2008) we developed and analyzed models for six fish species endemic to the study area with focus on two questions: (i) How is predictive performance affected by species biogeography? and (ii) How is the relative influence of each spatial scale and environmental feature affected by species biogeography?

2. Methods

2.1. Study area

The upper Green River system (Ohio River drainage) is located in the central-eastern United States (Fig. 1) and drains an area of approximately 13,000 sq. km. with elevation ranging from 110 m to 550 m above sea level. Three main subsystems comprise the upper Green River system – the Barren River, Nolin River, and Green River
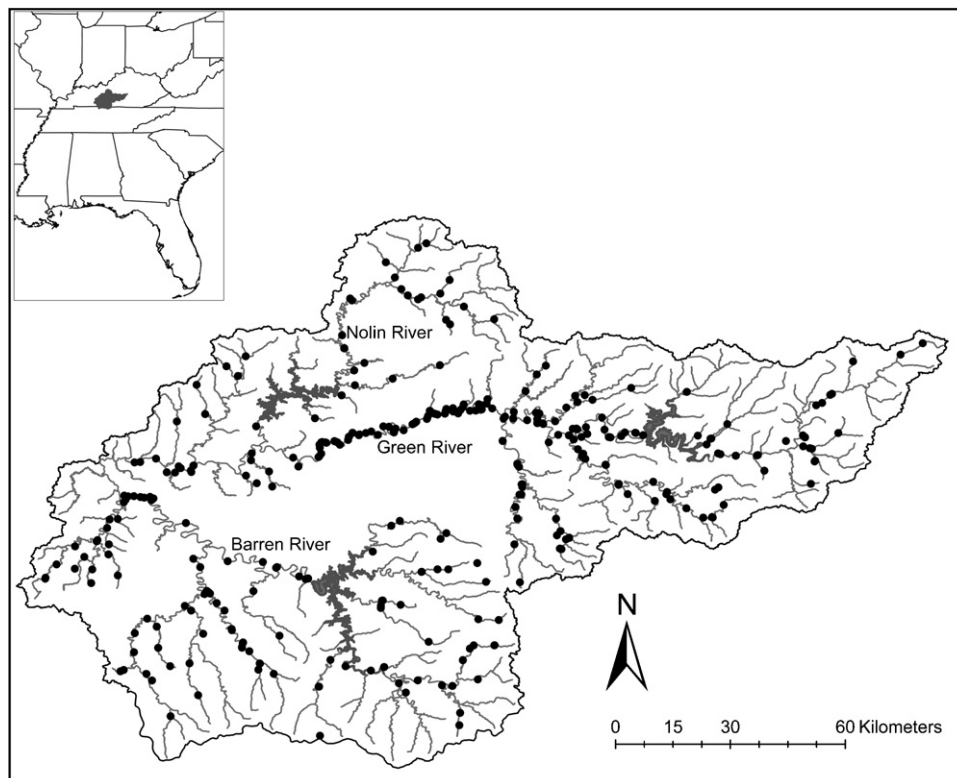

Fig. 1. Map of study area with sample sites. Only streams of 2nd order size or greater are shown. Shaded area in inset maps shows location of study area in eastern United States.

proper – and each is impounded by a reservoir. Physiography is varied, with dissected uplands in the southern, eastern, and northern areas of the watershed. In contrast, central and western portions of the drainage have extensive karsting and much lower relief topography.

2.2. Fish data

Occurrence data for six fish species endemic to the upper Green River system were extracted from an existing dataset consisting of museum-based information (Hopkins et al., 2009; Hopkins, in press). Records were available for 304 sites sampled between 1995 and 2008. Multiple locations received repeated visits during the 13-year period—sites were coded as positive if the respective species was detected. Sites in which a species was not captured were treated as probable absences. Survey efforts were not standardized and a variety of sampling techniques were used. The potential confounding effects of variable sampling efficiencies is recognized, however, the repeated sampling and temporal span of the dataset reduces the extent of these problems. Because absence records are never completely reliable, a binomial deviance loss function which is robust to noisy occurrence data (e.g. false negative observations) (Elith et al., 2008) was employed in model fitting.

The six species selected have extensive collection histories and the broad-scale distributions are well known. Each species is endemic to the study area, these were: *Etheostoma barrenense* (Splendid Darter), *E. bellum* (Orangethroat Darter), *E. barbouri* (Teardrop Darter), *E. rafinesquei* (Kentucky Snubnose Darter), *E. kantuckeense* (Highland Rim Darter), and *Thoburnia atripinnis* (Blackfin Sucker). We focus on species endemic to the study area because they afford a more robust analysis of species–environment relationships. Each species varied in terms of prevalence and geographical range. Prevalence of each species was estimated as the proportion of sample sites occupied by the species. Geographic range of each species was estimated by the area (km^2) of the smallest polygon enclosing all positive occurrences of the species on the map. Species were ranked by prevalence and range to facilitate comparison of influence of environmental predictors in the models.

2.3. Environmental predictors

Following the suggestions of Allan (2004), three spatial scales were analyzed for each sample point, these were: (1) the subcatchment, (2) a 100 m subcatchment riparian buffer, (3) and a 100 m buffer for a 1000 m upstream reach. These scales are referred to as the subcatchment, riparian buffer, and reach buffer, respectively. Subcatchments for each point were delineated from a 30 m digital elevation model and streams within the subcatchment were then buffered using ArcGIS 9.2 (Environmental System Research Institute, 2006) to create riparian buffers. Reach segments were delineated by hand, then buffered using AcrGIS 9.2 and the resulting polygons were used as masks for data extraction.

For each sampling point, 75 variables were calculated – 25 for each scale – comprising four broad types of landscape features: LULC pattern, LULC composition, soil composition, and geology composition (Table 1). Landscape metrics were calculated using FRAGSTATS (McGarigal and Marks, 1995): data for each spatial scale for each point were extracted and analyzed as a unique landscape. Geology and soil composition data were extracted from 1:250,000 maps. Land use/cover data were extracted from a ca. 2000 map created for a related project using Landsat ETM+ images from September and October 1999–2001. Five classes were used in the LULC map: agriculture, pasture, forest, water, and developed/exposed. Urban and barren land covers, both of which were present in very low proportions and had a high degree of spatial association, were combined to form the developed/exposed

Table 1
Description of metrics included in the distributional models and the respective categories. Each metric was calculated at three spatial scales for a total of 75 metrics.

Variable type	Description
Land use/land cover composition	
Pct_agr	Percent proportion of agricultural land use
Pct_pas	Percent proportion of pasture land use
Pct_for	Percent proportion of forest land cover
Pct_dev	Percent proportion of developed/exposed land use
Land use/land cover pattern	
Pd_agr	Number of agricultural patches per 100 hectares
Pd_pas	Number of pasture patches per 100 hectares
Pd_for	Number of forest patches per 100 hectares
Pd_dex	Number of developed/exposed patches per 100 hectares
Ps_agr	Mean agriculture patch size in hectares
Ps_pas	Mean pasture patch size in hectares
Ps_for	Mean forest patch size in hectares
Ps_dex	Mean developed/exposed patch size in hectares
Pd	Total patch density per 100 hectares
Contagion	Probability of adjacent cells being the same LULC type
Soil composition	
Pct_fl	Percent proportion of fine-loamy soils
Pct_f	Percent proportion of fine soils
Pct_fs	Percent proportion of fine-silty soils
Pct_c	Percent proportion of clayey soils
Pct_cs	Percent proportion of coarse-silty soils
Pct_ls	Percent proportion of loamy-skeletal soils
Geology composition	
Pct_ac	Percent proportion of alluvium/coal measure
Pct_fly	Percent proportion of flysch
Pct_mol	Percent proportion of molasse
Pct_shl	Percent proportion of shale carbonate
Pct_she	Percent proportion of shelf carbonate

class. The map had an overall accuracy of 88.04%. A ca. 2000 land cover map was chosen because greater than 70% of the records in the dataset were accumulated between 1997 and 2002; thus, the time interval between sampling and acquisition of LULC data was decreased. A basic assumption of the study was limited changes in LULC during the sampling period, especially between 1997 and 2002. A general comparison of the ca. 2000 map to 1992 National Land Cover Data showed no marked changes in LULC within the study area.

2.4. Statistical modeling and predictive mapping

Relationships between species occurrence and environmental predictors were analyzed using boosted regression trees (BRT), a form of logistical regression using decision trees and a boosting algorithm. Boosting is an optimization technique that reduces predictive deviance in a stagewise fashion by combining numerous trees into a single model. Each successive tree focuses on modeling as-yet-unexplained response variance. The effects of successive trees within a model on predictions are reduced in relation to the learning rate applied to the model fitting process to avoid overfitting the model. For an excellent overview of BRTs see Elith et al. (2008).

Each model was fitted in R (R Development Core Team, 2006) using the gbm package (Ridgeway, 2006) and code provided by Elith et al. (2008). Optimal model parameters were determined by trial and error—altering tree complexity (the number of split points in each tree) and learning rate until predictive deviance was minimized without overfitting. The degree of overfitting was assessed by observing the change in predictive error produced when additional trees are added to the models (e.g. sharp increases in predictive error with the addition of trees indicates an overfit model). Models in which the predictive error remained stable with the addition of trees beyond the optimum were chosen. Each

optimization trial consisted of a 10-fold cross-validation using a training fraction of 0.90. All variables were included in the model fitting process; however, BRT, being a machine-learning method, only incorporates important variables into the final model. A recursive feature elimination procedure was used to test for overfitting arising from a high number of predictor variables. In this test, a model is simplified by removing variables in a stepwise process until the resulting predictive error exceeds the predictive error of the full model.

Once the optimal combination of tree complexity and learning rate was found, predictive performance was further evaluated using a ten-fold cross-validation procedure with re-substitution. For each cross-validation trial, 70% of the dataset was randomly selected for model fitting and the excluded 30% was used for testing. The predictive error, percent response variance explained, and area under the receiver operator characteristic curve (AUC) were calculated for each species. The relative influence of each environmental variable in the model was determined using scripts from the gbm library. To aid comparison, the total influences of variables in the models were summed according to general feature type (e.g. geology composition, LULC composition, LULC pattern, and soil composition) and spatial scale (e.g. subcatchment, riparian buffer, and reach buffer). Patterns of influence of each feature type and spatial scale were then compared in terms of species prevalence and geographic range. Fitted functions of select environmental variables are provided to demonstrate species responses.

Predictive mapping was performed by dividing all streams 2nd order and above into 3000 m segments and quantifying the environmental predictors for the segment midpoint. In total, 1108 stream segments were delineated. Probability of occurrence for each species was then calculated using the optimal models found and mapped using ArcGIS 9.2.

3. Results

3.1. Model performance, parameters, and predictive mapping

Species prevalence varied from 0.04 to 0.46 and range varied from 965 to 8219 km^2 (Table 2). Rankings of species in terms of prevalence and range were the same; thus, the two attributes are hereafter considered concurrently. The Teardrop Darter was the most narrowly distributed (965 km^2) and least prevalent (0.04) species and was detected only in small streams in eastern portions of the study area. In contrast, the Orangefin Darter was detected at 46% of the sample sites and occupies portions of all major river systems in the watershed with the exception of the Nolin River drainage. The Blackfin Sucker (range = 1137 km^2; prevalence = 0.06), Highland Rim Darter (range = 2436 km^2; prevalence = 0.07), and Splendid Darter (range = 2631 km^2; prevalence = 0.19) are all restricted to the upper Barren River drainage. The Kentucky Darter (range = 7049 km^2; prevalence = 0.46) is the only species which occupies portions of the Nolin River drainage and is not found in the Barren River drainage.

Thus, each of the species represents a unique biogeography in terms of range and prevalence.

Despite marked differences, species biogeography did not affect model performance. All models provided good levels of discrimination with AUCs ranging from 0.87 to 0.98 (Table 2). The Splendid Darter model performed best (AUC = 0.98 ± 0.02) with 92.4% of the response variance explained and a low predictive error (0.09 ± 0.02). The Highland Rim Darter model had the lowest AUC of 0.87; however, the Teardrop Darter model explained the lowest proportion of the response variance (71.8%). The highest predictive errors were detected in the Kentucky Darter (0.36 ± 0.04) and Orangefin Darter (0.34 ± 0.05) models. A high proportion of the response variance was explained by all models (range = 71.8–92.4%)–demonstrating strong associations between landscape variables and species distributions.

In contrast, differences in optimal model parameters were related to disparate biogeographical attributes. Narrowly distributed, uncommon species (e.g. Teardrop Darter and Blackfin Sucker) required lower tree complexities (tc = 3) and slow learning rates (lr = 0.005). Species with moderate ranges (e.g. Highland Rim Darter and Splendid Darter) required comparable tree complexities but faster learning rates. Lastly, broadly distributed and more common species (e.g. Kentucky Darter and Orangefin Darter) also required higher tree complexities (tc = 5 and 7, respectively).

Predictive maps (Figs. 2–4) showing probability of occurrences for each species provides further insight into the performance and application of the statistical models. The predicted distributions for five species (Blackfin Sucker, Highland Rim Darter, Splendid Darter, Kentucky Darter, and Orangefin Darter) align closely with expectations based on historical collection records (Burr and Warren, 1986). Examination of predictive maps suggests overfitting was not a problem and models generalized quite well by predicting distributions outside the range of observed occurrences. Conversely, the predicted distribution of Teardrop Darter was markedly different from expectations based on historical records; however, the predicted range was reasonable based on observed occurrences in the dataset. There is mounting evidence that the Teardrop Darter has experienced a sharp range-contraction during the last 30 years (Hopkins et al., 2009). Once widespread, the Teardrop Darter now appears restricted to isolated streams in the extreme eastern portions of the watershed.

3.2. Relative influence of environmental features and spatial scale

In each model, with the exception of Teardrop Darter, LULC pattern and soil composition were the most influential landscape features (Fig. 5). The influence of soil composition was particularly high for those species with narrow ranges and endemic to the Barren River drainage. Soil composition was not associated with the distribution of Teardrop Darter. Also, in the Kentucky Darter and Orangefin Darter models, influence of soil composition was only slightly higher than geology composition and LULC composition. Overall patterns of influence were correlated with species range.

Table 2
Summary of biogeography, optimal parameters, and performance measures for each species model.

|	Teardrop Darter	Blackfin Sucker	Highland Rim Darter	Splendid Darter	Kentucky Darter	Orangefin Darter
Range (km^2)	965	1137	2436	2631	7049	8219
Prevalence	0.04	0.06	0.07	0.19	0.23	0.46
Tree complexity	3	3	3	3	5	7
Learning rate	0.005	0.005	0.01	0.01	0.01	0.01
Number of trees	1250	2250	1700	1400	2000	1800
% Variance explained	71.8	91.2	83.4	92.4	73.7	75.9
Predictive error	0.12 ± 0.02	0.08 ± 0.03	0.12 ± 0.04	0.09 ± 0.02	0.36 ± 0.04	0.34 ± 0.05
AUC	0.94 ± 0.05	0.95 ± 0.03	0.87 ± 0.05	0.98 ± 0.02	0.89 ± 0.05	0.91 ± 0.03
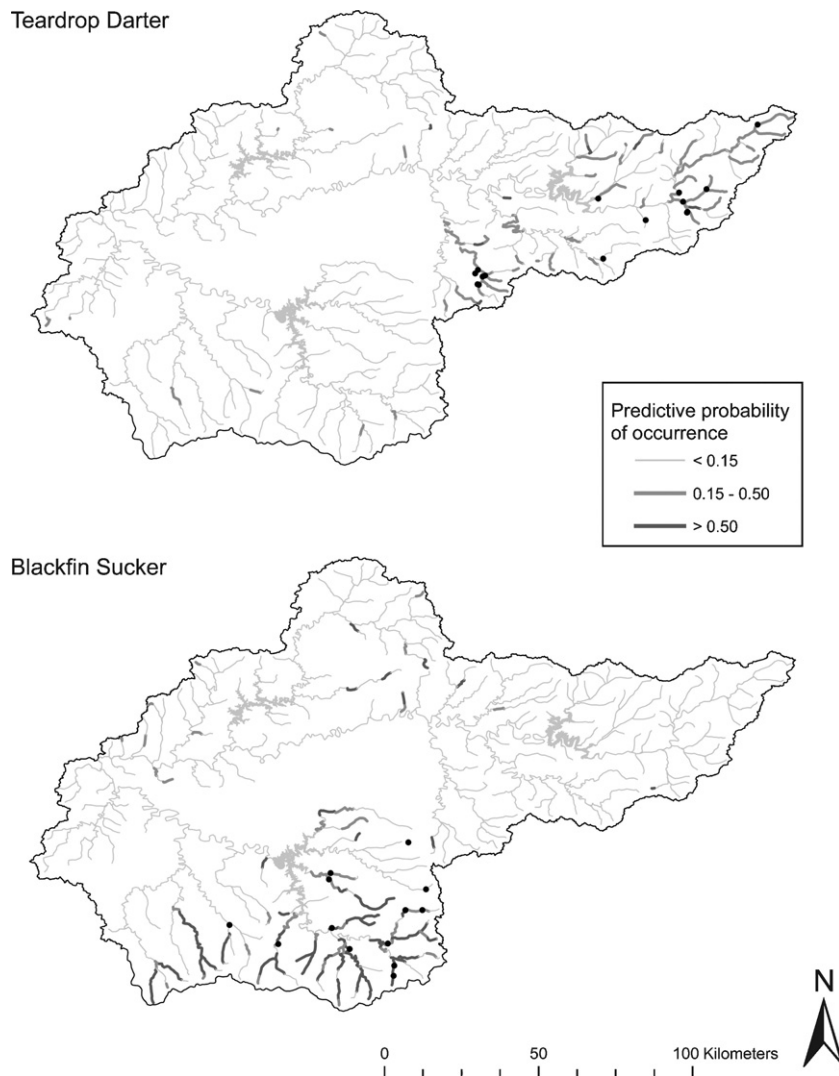
Fig. 2. Probability of occurrence map for Teardrop Darter and Blackfin Sucker. Three different probability classes are represented. Black dots are observed occurrences.

However, the relative influence (RI) for type of predictor varied markedly as did the RIs of each spatial scale.

For Splendid Darter, the influence of each feature increased across spatial scales with subcatchment being most important (RI = 58.79%): at each scale, soil composition was the most influential, followed closely by LULC pattern. The pattern of influence for Blackfin Sucker and Highland Rim Darter, which are also endemic to the Barren River, was similar to the Splendid Darter model. There was one notable difference, LULC pattern in the riparian buffer exhibited very high influence (RI = 29.66%) in the Highland Rim Darter model. Comparatively, the pattern of influence for each environmental feature was similar for Kentucky Darter and Orangefin Darter and approximately conserved across spatial scales. For Kentucky Darter and Orangefin Darter, which are more widespread, LULC pattern was consistently the most influential variable at each scale with a cumulative influence of 65.39 and 53.67% in the models, respectively.

The pattern of influence observed for Teardrop Darter was strikingly different. Spatial pattern of LULC exhibited at cumulative influence of 74.98% in the model, most of which was attributable to the subcatchment (61.84%). Composition of LULC in the riparian buffer had an RI of 12.70% in the model. It is interesting to note that this disparate pattern is observed in the single species in which a severe range-contraction is suspected. Unlike other species

with narrow ranges (e.g. Blackfin Sucker, Highland Rim Darter, and Splendid Darter), geology and soil composition had essentially no effect on the Teardrop Darter model.

Despite some congruence in overall importance of LULC pattern and soil composition, each species model consisted of distinct subsets of environmental variables: precise metrics incorporated, their relative influence, and response curves varied with each species. Boosted Regression Trees are designed to select (use) only important variables in modeling species distributions (Elith et al., 2008). Unimportant variables are technically left in the model, but have no or very little influence. The number of important (RI ≥ 3%) variables in each model ranged from four (Kentucky Darter model) to nine (Blackfin Sucker model). The number of important variables was not correlated with biogeographical attributes. Even though only a small subset of variables was important for each species, using the full set of predictors did not result in overfit models. Recursive feature elimination tests revealed that, depending on the model, anywhere from 61 to 72 variables could be removed without resulting in increases in predictive error.

Fitted functions of the most influential LULC pattern metric and soil composition metric demonstrate differences in models. When the effect on logit(p) exceeds zero, the corresponding variable measures are associated with the presence of the species. In contrast, values below zero indicate a negative effect on species occurrence.
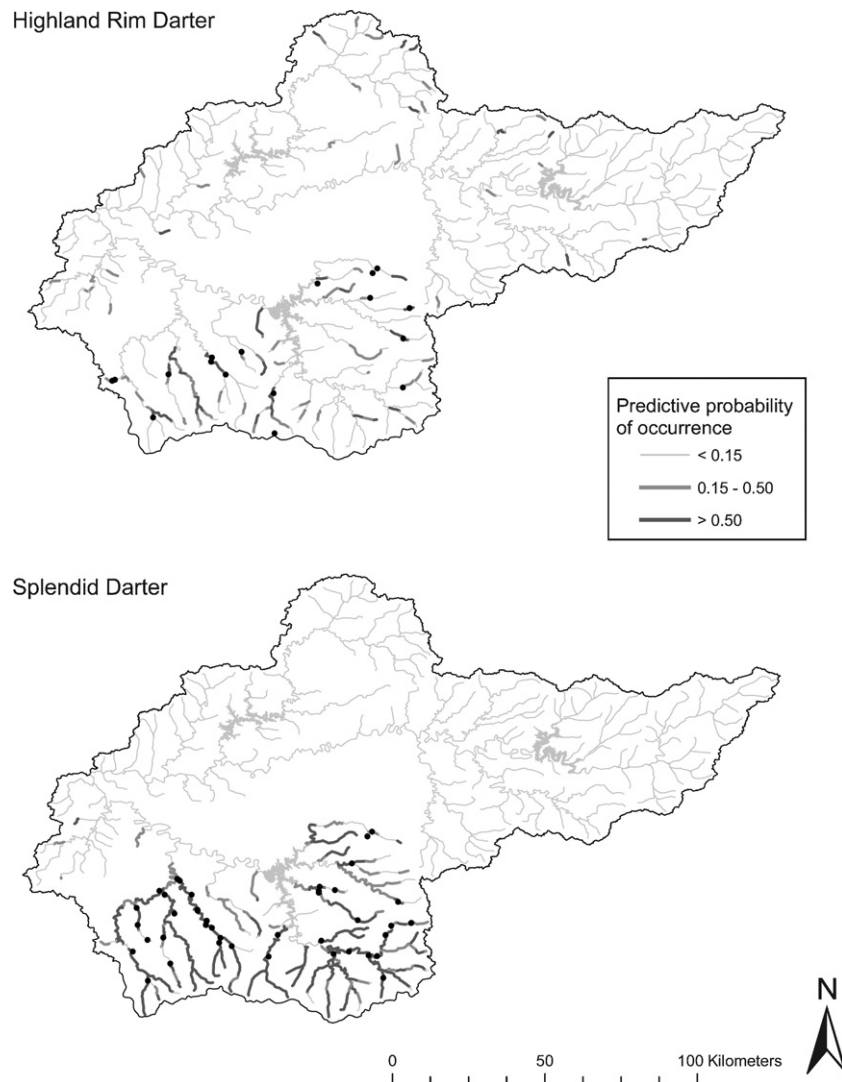
Fig. 3. Probability of occurrence map for Highland Rim Darter and Splendid Darter. Three different probability classes are represented. Black dots are observed occurrences.

The most influential LULC pattern metric and the respective spatial scale were different in each species model (Fig. 6). There was a degree of commonality among three models with spatial pattern of developed/exposed land cover being the most influential LULC pattern metric in the Teardrop Darter, Blackfin Sucker, and Highland Rim Darter models. The species responded negatively to increases in mean patch size and patch density of develop/exposed land within the riparian buffer and subcatchment. In contrast, Splendid Darter was highly influenced by the patch density of pasture in the subcatchment, being positively associated with low patch densities. Kentucky Darter responded strongly to mean patch size of agriculture in the subcatchment. Lastly, Orangefin Darter was positive correlated with the measure of contagion in the reach buffer.

Responses to soil composition were very different for each species (Fig. 7). Generally speaking, the influence of soil composition was lower for widely distributed species, however, the Teardrop Darter proved to be an exception. Both the Teardrop Darter and Kentucky Darter showed essentially no association with soil composition, with the most influential variables having RIs of 1.3 and 1.8%, respectively. Orangefin Darter was only marginally associated with the percent of fine-loamy soils in the subcatchment (RI = 3.3%); moreover, the relationship between this predictor and species distribution varied. In contrast, species endemic to the Barren River were highly influenced by soil composition. Blackfin

Sucker responded positively to the percent of clayey soils in the subcatchment (RI = 15.7%), with a noticeable threshold around 9%. Highland Rim Darter and Splendid Darter were positively correlated with the percent of loamy-skeletal soils in the subcatchment (RIs = 25.0 and 12.5%, respectively).

4. Discussion

Our results provide evidence that the multiscale filter framework is an appropriate ecological model for predicting fish species distributions when using landscape data. Contrasting biogeographies did not affect model performance and all models offered high levels of discrimination for species occurrences with AUCs ranging from 0.87 to 0.98. Predictive maps suggest distributions were well approximated based on observed occurrences. Comparisons of predicted distributions to historical collection records also indicate accurate estimations of distributions for five species. However, the predicted distribution of the Teardrop Darter was truncated when compared to the historical range. As aforementioned, a range-contraction has been documented for this species: the Teardrop Darter only occupies about 20% of the historical range (Hopkins et al., 2009). There are anecdotal reports suggesting that other darters in the subgenus Catonotus (which have similar ecologies) have also
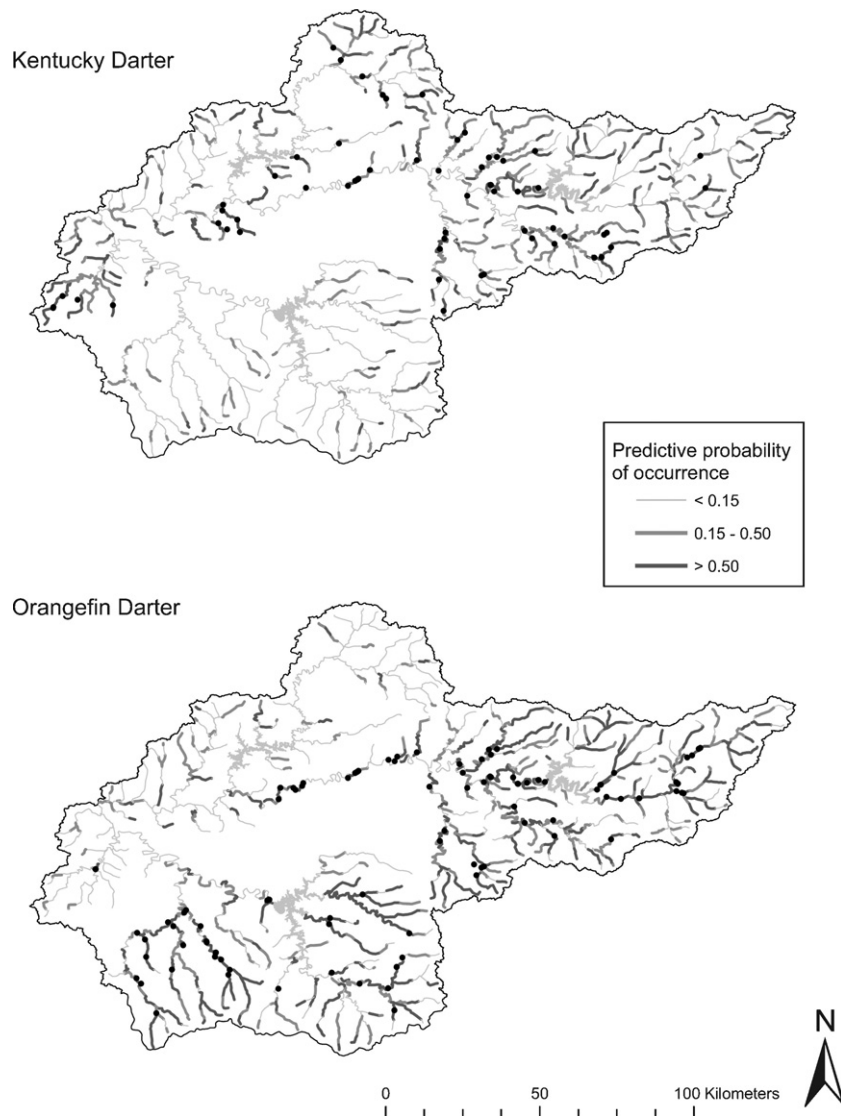
Fig. 4. Probability of occurrence map for Kentucky Darter and Orangefin Darter. Three different probability classes are represented. Black dots are observed occurrences.

experienced sharp reductions in overall ranges (Hopkins and Fisk, 2008). Causes of decline have not been formally investigated; however, coal mining activities and agricultural land use have been implicated. In the current study, the Teardrop Darter was strongly influenced by the patch density of develop/exposed land within the subcatchment and was strongly associated with subcatchments having low densities of develop/exposed patches. Teardrop Darter was also positively associated with riparian buffers having less than 7% agricultural land use (result not shown). Each of these land cover types (develop/exposed and agriculture) has been shown to have negative effects on stream systems by increasing sedimentation and modifying flow regimes (Allan, 2004). Teardrop Darter, and Catonotus darters in general, may be especially sensitive to such perturbations within the landscape.

Using a suite of multiscale landscape predictors proved vital as each species distribution was modeled with markedly different subsets of environmental predictors. For each model, a high proportion of response variance was explained with relatively low numbers of predictors—providing confidence in using the models to investigate species–environment relationships. The relative contribution of each type of landscape feature and spatial scale was contingent upon species biogeography. The spatial pattern of LULC was important in modeling the distribution of all species.

However, as aforementioned, the influence of specific variables varied greatly between models. Three of the species responded most strongly to measures of developed/exposed land in the riparian buffer and subcatchment. Species responded negatively to increasing patch densities and patch sizes of develop/exposed land. Blackfin Sucker and Highland Rim Darter responded negatively when mean patch sizes of developed/exposed land surpassed 0.1 ha in the riparian buffer and subcatchment, respectively. This is a very small patch size, however, Allan (2004) notes that the effects of develop/exposed land on stream ecosystems are disproportionately large compared to the prevalence in the landscape. Similarly, Kentucky Darter responded negatively when mean patch size of agricultural lands in the subcatchment exceeded 1.9 ha. Analogous to develop/exposed land, the effects of agricultural land on stream systems are also accentuated when patch sizes are large (Turner et al., 2001).

Comparatively, Splendid Darter responded most strongly to the density of pasture patches in the subcatchment. Native conditions in the study area consist of forest, thus, higher patch densities are indicative of increasing levels of disturbance (e.g. human activities) within the watershed which results in degraded stream conditions (Turner et al., 2001). Given that the Barren River consists predominantly of pasture and forested lands, Splendid Darter appeared to
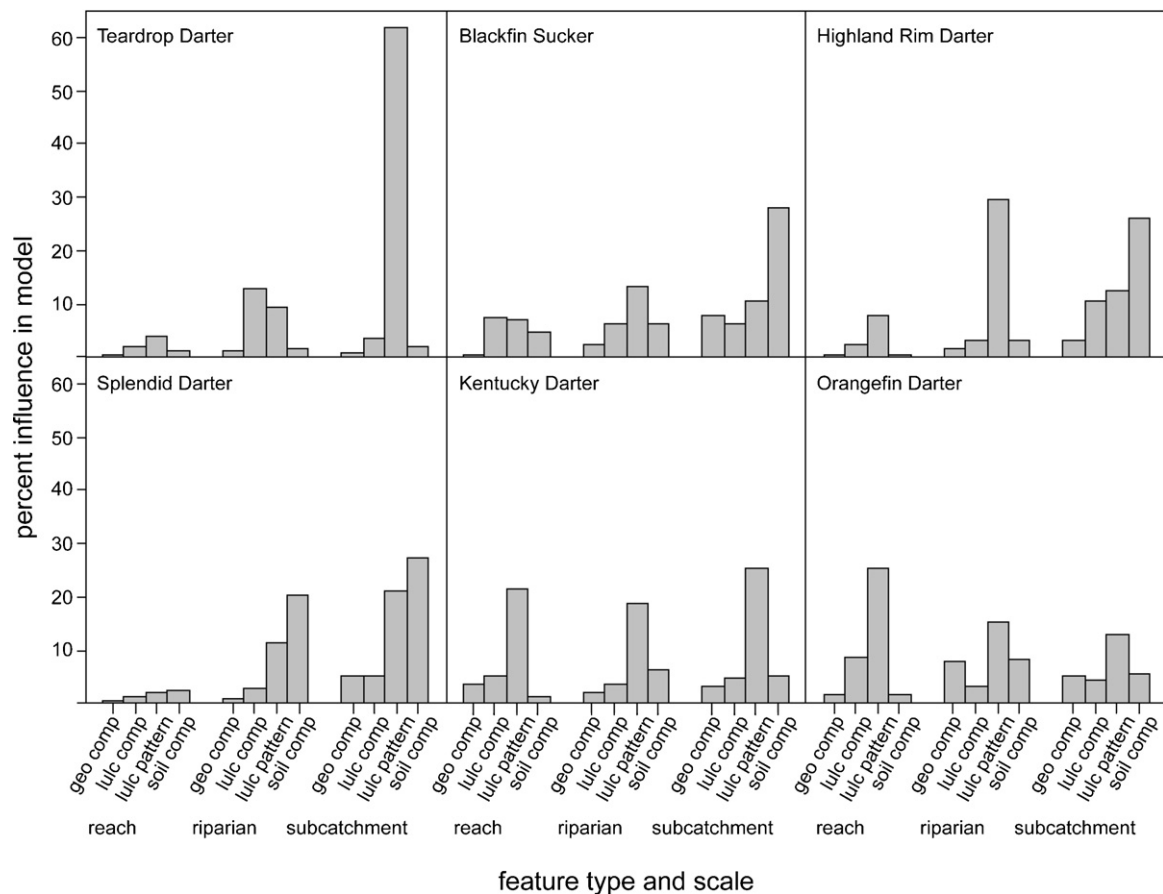
Fig. 5. Summary of relative contributions of each variable type for each spatial scale for each species distribution model.

occupy regions in which the subcatchment was less fragmented. Likewise, Orangefin Darter responded positive to increasing contagion in the reach buffer. Interrogation of the model indicated a strong interaction of contagion with percent forest in the reach buffer. Optimal conditions for Orangefin Darter occurred with high contagion and high percent forest in the reach buffer.

For the narrowly distributed species, LULC pattern was most influential at the riparian buffer and subcatchment scales. As species range increased, the influence of LULC pattern in the reach buffer also increased strikingly with RIs exceeding 20%. Notably, only the two most widely ranging species – Kentucky Darter and Orangefin Darter – were observed in areas downstream of major reservoirs. It is possible that conditions in the reach buffer are more influential for these species because upstream linkages, at least for some populations, are disrupted by the presence of reservoirs (Bunn and Arthington, 2002). In contrast, none of the narrow ranging species were located downstream of reservoirs: for these species, it seems that associations with riparian buffer and subcatchment conditions are preserved (Hynes, 1975; Vannote et al., 1980). This shift in relative influence was only observable with a multiscale approach.

For species endemic to the Barren River drainage, we detected a multiscale filtering effect (Poff, 1997). Within the subcatchment, the soil composition was the most influential variable for each of these species. According to Poff (1997) and MacKey and Lindenmayer (2001), effects of geomorphic variables are manifest at larger spatial scales. In particular, these species responded positively to increasing proportions of loamy-skeletal and clayey soils in the subcatchment. Major deposits of both these soil types are only found in headwater areas of the Barren River drainage—which is where these species are restricted to. At more proximal scales

(e.g. riparian buffer and reach buffer), the influence of LULC pattern increased substantially and influence of soil composition decreased. There was a slight deviance from this pattern which may have been related to species prevalence. For the Splendid Darter, the influence of all variables decreased with spatial scale. The Splendid Darter is ecologically labile and very common throughout its range (Etnier and Starnes, 1993). Thus, variables quantified at larger scales worked well for modeling the distribution of this species. In contrast, the Blackfin Sucker and Highland Rim Darter, which are more sensitive to local environmental conditions (Etnier and Starnes, 1993), have sporadic distributions within their range and appear to respond more strongly to landscape conditions, especially LULC pattern, within reach buffers and riparian buffers. Proximal landscape conditions, especially LULC pattern and composition, strongly affect sediment and flow regimes which are the primary drivers of physical habitat in stream systems (Bunn and Arthington, 2002). Other aquatic fauna, such as freshwater mussels, have also been documented to respond strongly to landscape conditions in the reach buffer (Poole and Downing, 2004; Hopkins, in press).

Observations for the Blackfin Sucker and Highland Rim Darter provide cogent evidence of a hierarchy effect and align closely with patterns of influence expected within the multiscale filter framework. Broad-scale aspects of the distributions are correlated with soil composition within the subcatchment. Nested within this broadly defined range, more local-scale aspects of distributions appear regulated by LULC conditions within the reach buffer and riparian buffer.

The ability to resolve this hierarchical constraint is perhaps an artifact of spatial scale, being related to the ratio of the species range to the study area extent. As this ratio increases, the probability of detecting factors determining broad-scale distributions decreases
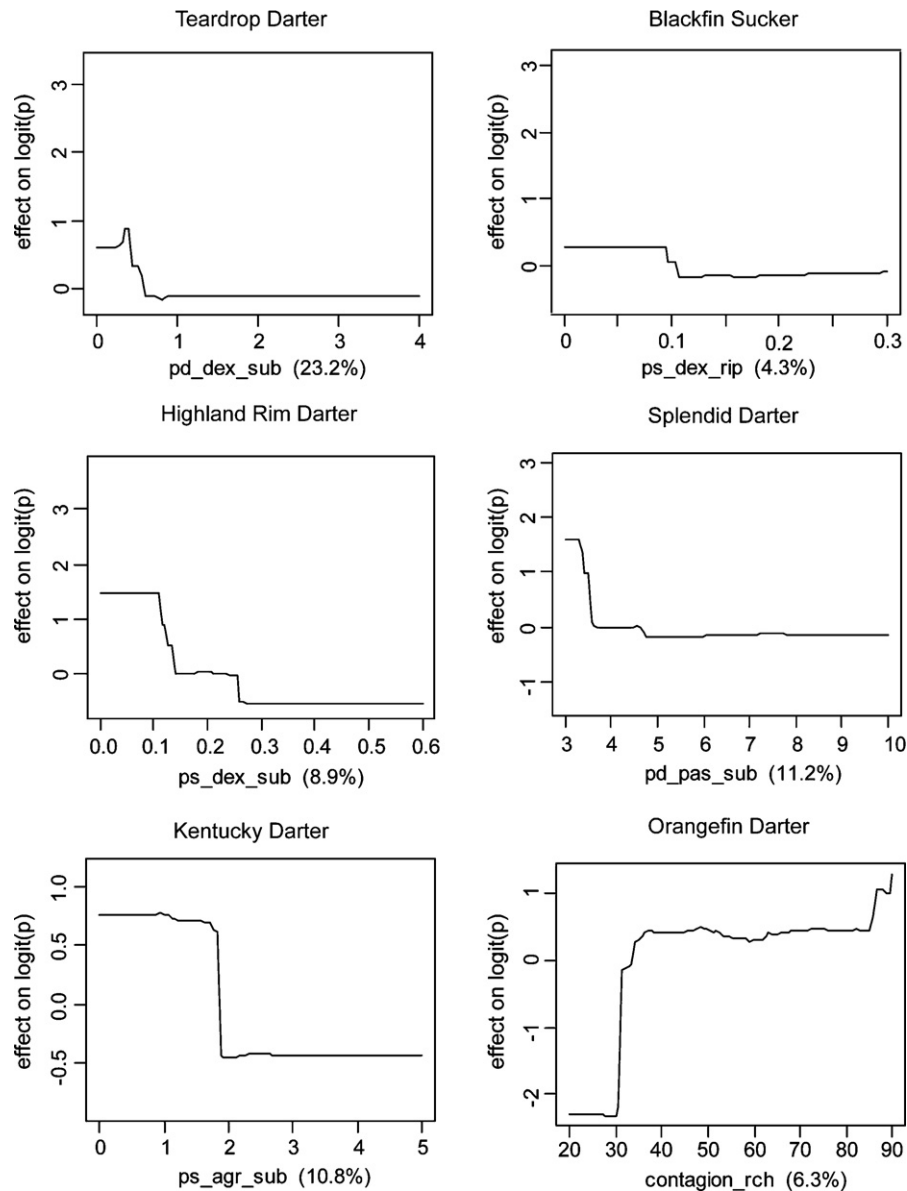
Fig. 6. Fitted functions for the most influential LULC pattern variable in the BRT model for each species examined. Note differences in logit scale (vertical axes) for each plot.

(Levin, 1992). This could explain why the more broadly distributed species showed no or weak hierarchy effects, each having a relatively consistent pattern of influence of variables across spatial scales with each scale having approximately equal magnitudes of influence.

On the contrary, the Teardrop Darter was narrowly distributed and but showed no evidence of a hierarchy effect. We speculate that the special circumstances surrounding the Teardrop Darter (e.g. range reduction) were responsible for producing the rather enigmatic patterns of influences for a narrow ranging species. Teardrop Darter was once widespread throughout the study area and, given the pattern detected for Kentucky Darter and Orangefin Darter, we would expect weak relationships to geomorphic conditions. As expected, no relationship with soil or geology composition is detected for Teardrop Darter. However, unlike Kentucky Darter and Orangefin Darter, there are no contemporary records of this species located below reservoirs. As a result – and, similar to more narrow ranging species – relationships with LULC conditions at larger spatial scales are more apparent. Certainly, further research is needed to elucidate the precise relationships

between this species and landscape conditions to identify causes of decline.

The incorporation of multiscale data capable of capturing different attributes of species biogeography is essential to developing robust species distribution models (MacKey and Lindenmayer, 2001). In the current study, the exclusion of either spatial scale or environmental feature would have certainly compromised model performance. For comparison, Joy and De'Ath (2004) used only subcatchment predictors in modeling the distribution of 14 New Zealand fishes and had lower AUCs ranging from 0.63 to 0.88. Also, when modeling aquatic species distributions, careful attention must be given when choosing the study area extent and also the extent of the area for predictive mapping. Predictive mapping outside of the study area will introduce some degree of certainty because, as demonstrated in this study, species–environment relationships are highly contingent upon species ranges and study area extents.

Lastly, a multiscale approach to species distribution modeling should not be viewed as a panacea. Additional factors not included in the present study are surely involved in regulating the distri-
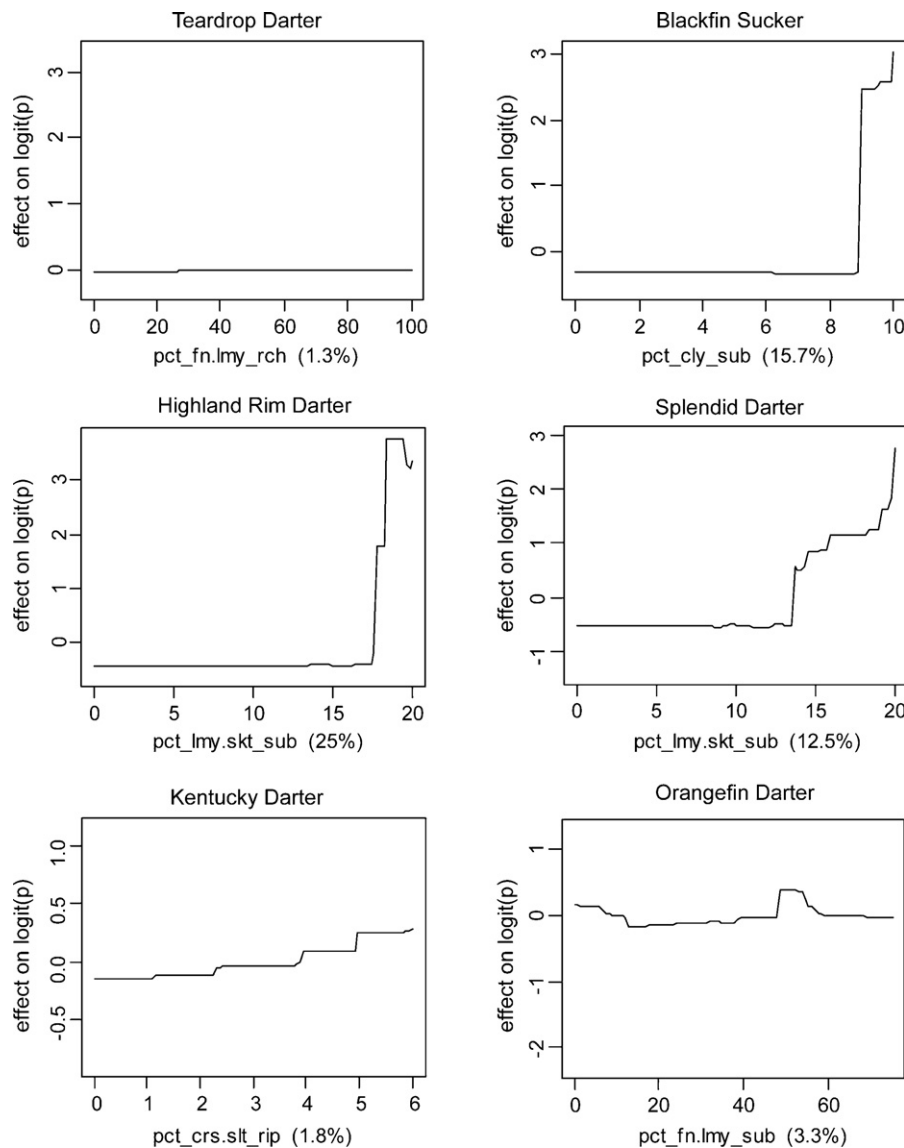
Fig. 7. Fitted functions for the most influential soil composition variable in the BRT model for each species examined. Note differences in logit scale (vertical axes) for each plot.

bution of the focal species. Other researchers (Jones et al., 1999; Weigel et al., 2003) have noted aquatic communities responding strongly to subtle variations in substrate conditions–variation not captured using landscape data. In a study of a Lake Ontario tributary, Singkran and Meixler (2008) found that a combination of microhabitat and landscape variables provided the best explanation of variation in fish assemblages along a longitudinal gradient. While the incorporation of microhabitat variables may provide increased explanatory and predictive power, the use of such variables would preclude predictive mapping across large spatial extents and limit the applicability of the model for conservation planning.

Conversely, biotic factors such as predation or competition may also greatly influence spatial distribution of species. Schrank et al. (2001) observed that distribution and persistence of Topeka Shiner populations in Kansas streams was highly influenced by the presence of predators. Similarly, Anderson et al. (2002) employed a GIS-based niche model to demonstrate how inclusion of variables approximating competition can greatly increase predictive performance of South American pocket mice distribution models.

5. Conclusions

The results of this study indicate that using landscape data applied under the multiscale filter framework is an appropriate method for modeling fish distributions and works well across species with contrasting biogeographies. All species were modeled well statistically and had a high AUC value. Predictive maps suggested that each model generalized well and provided a good approximation of species distribution based on observed occurrences. Incorporating geomorphic and LULC variables permitted the model the capture hierarchical effects on species distributions. The interplay of species range and prevalence determine the relative influence of spatial scale and landscape features in SDMs. In the case of species endemic to the upper Barren River drainage, strong hierarchical effects were detected with geomorphic variables exerting notable influence at the riparian buffer and subcatchment scales. Species with sporadic distributions with their ranges were also strongly influenced by LULC pattern in the reach buffer. Exclusion of multiscale data would have certainly resulted in decreased model performance and reduced the ability to formulate functional linkages between landscape variables and species distributions.

The pattern of environmental influence detected for the Teardrop Darter was in sharp contrast to that of other narrowly distributed species–essentially no geomorphic influences were detected. We speculate that putative range reduction of Teardrop Darter contributed to this enigmatic pattern of influence. In summary, we provide support for use of multiscale landscape data in modeling fish distributions for ecological studies and conservation planning.

References

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

Allan, J.D., 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems. Annu. Rev. Ecol. Syst. 35, 257–284.

Anderson, R.P., Peterson, A.T., Gomez-Laverde, M., 2002. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. Oikos 98, 3–16.

Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. Ecol. Model. 157, 101–118.

Austin, M.P., 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecol. Model. 200, 1–19.

Bunn, S.E., Arthington, A.H., 2002. Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. Environ. Manage. 30, 492–507.

Burr, B.M., Page, L.M., 1986. Zoogeography of fishes of the lower Ohio-upper Mississippi basin. In: Hocutt, C.H., Wiley, E.O. (Eds.), The Zoogeography of North American Freshwater Fishes. John Wiley & Sons, New York, NY, USA, pp. 287–324.

Burr, B.M., Warren Jr., M.L., 1986. A distributional atlas of Kentucky fishes. Kentucky Nat. Preserves Commiss. Sci. Techn. Ser. 4, 1–398.

Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dobson, A., Foley, J.A., Lodge, D.M., Pascual, M., Pielker Jr., R., Pizer, W., Pringle, C., Reid, W.V., Rose, K.A., Sala, O., Schlesinger, W.H., Wall, D.H., Wear, D., 2001. Ecological forecasts: an emerging imperative. Science 293, 657–660.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77, 802–813.

Environmental System Research Institute, 2006. ArcGIS/ArcInfo for Windows XP Version 9.2.0. Redlands, CA, USA.

Etnier, D.A., Starnes, W.C., 1993. The Fishes of Tennessee. The University of Tennessee Press, Knoxville, TN, USA, p. 689.

Franklin, J., Wejnert, K.E., Hathaway, S.A., Rochester, C.J., Fisher, R.N., 2009. Effect of species rarity on the accuracy of species distribution models for reptiles and amphibians in southern California. Diversity Distrib. 15, 167–177.

Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8, 993–1009.

Hopkins, R.L., in press. Use of landscape pattern metrics and multiple spatial scales in aquatic species distribution models. Landscape Ecol.

Hopkins II, R.L., Burns, M.D., Burr, B.M., Hopman, L.J., 2009. Building a centralized database for Kentucky fishes: progress and future applications. J. Kentucky Acad. Sci. 69, 73–78.

Huston, M.A., 2002. Introductory essay: critical issues for improving predictions. In: Scott, J.M., Heglun, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Covelo, CA, USA, pp. 7–21.

Hynes, H.B.N., 1975. The stream and its valley. Verhandlungen der Internationale Vereinigung für Theoretische und Angewandte Limnologie 19, 1–15.

Jones III, E.B.D., Helfman, G.S., Harper, J.O., Bolstad, P.V., 1999. Effects of riparian forest removal on fish assemblages in southern Appalachian streams. Conserv. Biol. 13, 1454–1465.

Joy, M.K., De'Ath, R.G., 2004. Predictive modeling and spatial mapping of freshwater fish and decapods assemblages using GIS and neural networks. Freshwater Biol. 49, 1036–1502.

Kearns, F.P., Kelly, N.M., Carter, J.L., et al., 2005. A method for the use of landscape metrics in freshwater research and management. Landscape Ecol. 20, 113–125.

Lammert, M., Allan, J.D., 1999. Assessing biotic integrity of streams: Effects of scale in measuring the influence of land use/cover and habitat structure on fish and macroinvertebrates. Environ. Manage. 23, 254–270.

Leathwick, J.R., Rowe, D., Richardson, J., Elith, J., Hastie, T., 2005. Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. Freshwater Biol. 50, 2034–2052.

Leibold, M.A., 1995. The niche concept revisited: mechanistic model and community context. Ecology 76, 1371–1382.

Levin, S.A., 1992. The problem of pattern and scale in ecology. Ecology 73, 1943–1967.

MacKey, B.G., Lindenmayer, D.B., 2001. Towards a hierarchical framework for modelling the spatial distribution of animals. J. Biogeogr. 28, 1147–1166.

Marmion, M., Luoto, M., Heikkinen, R.K., Thuiller, W., 2008. The performance of state-of-the-art modelling techniques depends on geographical distribution of species. Ecol. Model., doi:10.1016/j.ecolmodel.2008.10.019.

McGarigal, K., Marks, B.J., 1995. FRAGSTATS. Spatial analysis program for quantifying landscape structure. USDA Forest Service General Technical Report PNW-GTR-351.

Moilanen, A., Leathwick, J., Elith, J., 2008. A method for spatial freshwater conservation prioritization. Freshwater Biol. 53, 577–592.

Oakes, R.M., Gido, K.B., Falke, J.A., Olden, J.D., Brock, B.L., 2005. Modelling of stream fishes in the Great Plains USA. Ecol. Freshw. Fish. 14, 361–374.

Olden, J.D., Jackson, D.A., 2002. A comparison of statistical approaches for modeling fish species distributions. Freshwater Biol. 47, 1976–1995.

Poff, N.L., 1997. Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. J. N. Am. Benthol. Soc. 16, 391–409.

Poole, K.E., Downing, J.A., 2004. Relationship of declining mussel biodiversity to stream-reach and watershed characteristics in an agricultural landscape. J. N. Am. Benthol. Soc. 23, 114–125.

R Development Core Team, 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ridgeway, G., 2006. Generalized boosted regression models. Available from ftp://ftp.auckland.ac.nz/pub/software/CRAN/doc/packages/gbm.pdf.

Schrank, S.J., Gyu, C.S., Whiles, M.R., Brock, B.L., 2001. Influence of instream and landscape-level factors on the distribution of Topeka Shiners Notropis topeka in Kansas streams. Copeia 2001, 413–421.

Scott, M.C., Helfman, G.S., McTammany, M.E., Benfield, E.F., Bolstad, P.V., 2002. Multiscale influences on physical and chemical stream conditions across Blue Ridge landscape. J. Am. Wat. Res. Assoc. 38, 1379–1392.

Singkran, N., Meixler, M.S., 2008. Influences of habitat and land cover on fish distributions along a tributary to Lake Ontario, New York. Landscape Ecol. 23, 539–551.

Turner, M.G., Gardner, R.H., O'Neill, R.H., 2001. Landscape Ecology in Theory and Practice: Pattern and Process. Springer-Verlag, New York, NY, USA, p. 401.

Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R., Cushing, C.E., 1980. The river continuum concept. Can. J. Fish. Aquat. Sci. 37, 130–137.

Weigel, B.M., Wang, L., Rasmussen, P.W., Butcher, J.T., Stewart, P.M., Simon, T.P., Wiley, M.J., 2003. Relative influence of variables at multiple spatial scales on stream macroinvertebrates in the Northern Lakes and Forest ecoregion USA. Freshwater Biol. 48, 1440–1461.