

Modelling of stream fishes in the Great Plains, USA

Oakes RM, Gido KB, Falke JA, Olden JD, Brock BL. Modelling of stream fishes in the Great Plains, USA.

Ecology of Freshwater Fish 2005: 14: 361–374. © Blackwell Munksgaard, 2005

Abstract – Predicting species distributions has important implications for the conservation and management of freshwater fishes, particularly in areas such as the Great Plains, USA where human impacts have resulted in extirpations and declines for numerous native species. There are a number of statistical approaches for constructing distributional models; the accuracy of each is likely dependent on the nature of the environmental gradients, species responses to those gradients and the spatial extent of the modelling. Thus, it is important to compare multiple approaches across species and habitats to identify the most effective modelling approach. Using geographical information system (GIS) derived characteristics of stream segments as predictors, we tested the model performance of three methodologies – linear discriminant function analysis, classification trees and artificial neural networks (ANN) – for predicting the occurrence of 38 fish species in a Great Plains river basin. Results showed that all approaches predicted species occurrences with relatively high success. ANN generally were the best models, in that they generated the most significant models (35 of 38 species) and most accurately predicted species presence for the greatest number of species (average correct classification = 81.1%). The importance of GIS variables for predicting stream fish occurrences varied among species and modelling techniques, but were generally strong predictors of species distributions, including the federally endangered Topeka shiner *Notropis topeka*. In summary, predictive models should be viewed as both competitive and complementary methodologies for establishing quantitative linkages between fish species and their environment. Our study demonstrates the potential utility of such an approach for guiding conservation efforts for stream fishes of the Great Plains, USA.

**R. M. Oakes^{1,2}, K. B. Gido¹,
J. A. Falke^{1,3}, J. D. Olden^{4,5},
B. L. Brock⁶**

¹Division of Biology, Kansas State University, Manhattan, KS, ²University of Pennsylvania Law School, Philadelphia, PA, ³Department of Fishery and Wildlife Biology, Colorado State University, Fort Collins, CO, ⁴Graduate Degree Program in Ecology, Colorado State University, Fort Collins, CO, ⁵Center for Limnology, University of Wisconsin, Madison, WI, ⁶Wildlife Conservation Society, Bozeman, MT, USA

Key words: predictive modelling, Great Plains, geographical information system, artificial neural networks, classification trees, linear discriminant function analysis

Keith B. Gido, Division of Biology, Kansas State University, Manhattan, KS 66506, USA; e-mail: kgido@ksu.edu

Accepted for publication June 15, 2005

Un resumen en español se incluye detrás del texto principal de este artículo.

Introduction

The conservation of stream fishes presents particular challenges to managers, because fishes typically use different habitats along the river continuum for spawning, feeding and refugia from predation or environmental disturbances (Schlosser & Angermeier 1995). Consequently, conservation plans must be developed at appropriate scales, which can range from instream habitats to entire riverscapes (Fausch

et al. 2002). Developing appropriate conservation plans takes on particular importance in light of the global decline in freshwater fish biodiversity and the multiple factors responsible for this conservation crisis (Allan & Flecker 1993). With greater numbers of species in need of protection and limited financial resources to devote to surveying stream habitats, predictive models provide scientists with a powerful tool, which can be used to assess the distribution and suitability of habitat for threatened species

across the landscape and facilitate the prioritisation of areas for conservation.

Predictive models have many potential applications for the conservation and management of freshwater fishes. Recent studies have used predictive models to measure impacts of habitat alteration (Oberdorff et al. 2001), examine the influence of scale and geography on relationships between fishes and landscape variables (Townsend et al. 2003), estimate habitat suitability for species re-introductions (Harig & Fausch 2002), predict the likelihood of species invasions (Vander Zanden et al. 2004) and identify areas of persistence for threatened or endangered species (Williams & Araujo 2000). However, an important question that has yet to be addressed is whether, and if so, species predictability is influenced by interspecific differences in species autecologies and the variability of environmental gradients across ecosystems. Clearly, testing the ability of models to successfully predict species occurrence in specific ecosystems is an important step in conservation. Moreover, the degree to which threatened species are predictable can be viewed as an important confirmation of our current understanding (or lack thereof) of the ecology of these imperiled species and can point to critical gaps in our knowledge.

The Great Plains ecosystem of the United States provides a unique template on which to study the predictability of freshwater fish distributions. The prairie biome is recognised as one of the most endangered regions on the continent (Samson & Knopf 1994; Dodds et al. 2004) and Great Plains streams are home to a number of threatened or endangered fishes (e.g., *Notropis topeka*; Haslouer et al. 2005). Modelling freshwater fish distributions in the Great Plains is particularly important for the management of this highly imperiled fauna because little information is known about the ecology of many of these fishes (Matthews 1988; Fausch & Bestgen 1997). Fishes in this region are tolerant of harsh natural conditions, including drastic changes in hydrological regime (Matthews et al. 1988; Fausch & Bramblett 1991; Labbe & Fausch 2000), and fluctuating temperatures and hypoxia (Matthews 1987; Smale & Rabeni 1995). Despite their tolerance, Great Plains fishes have been negatively impacted by human-induced changes in land use (Berkman & Rabeni 1987; Schlosser 1991), hydrology (Eschner et al. 1983; Cross et al. 1985; Winston et al. 1991) and the spread of non-native species (Cross & Moss 1987; Gido et al. 2004), which have collectively resulted in extirpations and drastic declines in a number of native species. The particularly vulnerable status of Great Plains fishes necessitates a predictive understanding of species–habitat relationships to aid the development of conservation plans aimed at preserving the biological integrity of the region.

Ideally, these models will predict the occurrence of species of concern and areas of high indigenous species richness (Wall et al. 2004). This study assesses the ability to develop accurate predictive models for fishes in these dynamic river systems, which to our knowledge, is previously untested.

Herein, we develop predictive models for fishes in the Big Blue River basin in north-eastern Kansas, with the primary goal of providing a management tool to aid in the conservation of fishes in the region. This river system supports a fish fauna that is in critical need of conservation. A comparison of recent surveys of the basin with those conducted in the 1950s suggests that several species have been extirpated (e.g., *Hybognathus argrytus*) or have declined in their range (e.g., *Notropis topeka*; Shrank et al. 2001; Gido et al. 2002). These same studies also document the recent introduction of several nonindigenous species; some within the past 3 years (Gido et al. 2002). Land-use conversion, hydrological modifications and impoundments (i.e., Tuttle Creek Reservoir) have impaired the basin (Falke & Gido in press). In short, the Big Blue River basin typifies many problems that are characteristic of Great Plains ecosystems, and is an appropriate system on which to study the utility of predictive modelling as a conservation tool.

Although there are many different statistical approaches to model ecological data, studies that have directly compared the effectiveness of the different techniques are scarce (Guisan & Zimmermann 2000). Studies that have made direct comparisons across modelling approaches (e.g., Mastrorillo et al. 1997; Manel et al. 1999; Olden & Jackson 2002) have reported significant differences across both statistical techniques and individual taxa. Recent studies have highlighted the utility of nonlinear statistical techniques such as classification trees (CT) (e.g., Magnuson et al. 1998; Rathert et al. 1999; Rejwan et al. 1999) and artificial neural networks (ANN) (e.g., Mastrorillo et al. 1997; Brosse & Lek 1999; Olden & Jackson 2001) in modelling the distributions of freshwater fishes. The advancement of nonlinear statistical approaches not only provides researchers with additional tools with which to model complex ecological relationships, but also presents the problem of choosing the approach that most accurately represents a species–habitat relationship. To address this question, we compared three modelling techniques to assess their performance in modelling the occurrence of Great Plains fishes.

Methods

Study area

The Big Blue River and its tributaries drain approximately 24,900 km² in north-eastern Kansas and

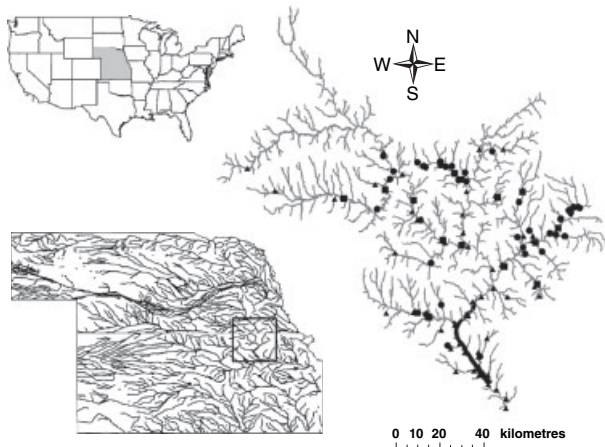


Fig. 1. Map of the Big Blue River Basin, located in south-eastern Nebraska and north-eastern Kansas, USA. Fish sampling locations are represented by circles (KDWP sites), squares (Gido et al. 2002) and triangles (Minckley 1959).

south-eastern Nebraska (Fig. 1). The physical habitats of these streams were influenced by past glaciation events, and are broadly classified into three types: sand-bottomed, mud-bottomed and gravel-bottomed streams (Minckley 1959). Tuttle Creek Dam impounded the Big Blue River in 1962, and formed a large (50 km²) reservoir located approximately 15 km upstream from the confluence of the Big Blue River and the Kansas River. Land cover in the drainage was historically tall and mixed-grass prairie, but is now predominantly row-crop agriculture and pasture.

Fish data

Predictive models were based on 120 fish collections, from which 50 species were captured. The frequency of occurrence of each species in these collections is presented in Table 1. Included in these collections were three species that have been identified as being of special conservation concern: *Hybognathus argrytus*, *Notropis topeka* and *Phoxinus erythrogaster* (Haslouer et al. 2005). Fifty-six sites were sampled between March 1957 and August 1958 (Minckley 1959), prior to the construction of Tuttle Creek Dam using a combination of entrapment devices (hoop and fyke nets), observations by anglers, gill nets, rotenone, electrofishing and predominately seining. The Kansas Department of Wildlife and Parks (KDWP) sampled 49 sites in the basin between September 1995 and August 2000. KDWP sampling followed the protocol of Meador et al. (1993); at each site a reach that was 40 times the average wetted width of the stream (minimum 150 m, maximum 300 m) was sampled using a combination of straight and bag seines (4.7-mm mesh) and DC-pulsed backpack electrofishing. Finally, Gido et al. (2002) sampled 15 sites in the Big Blue River Basin in

autumn 2001, using 4.6 m × 1.2 m (3.2-mm mesh) or 6.1 m × 1.8 m (4.8-mm mesh) straight seines, and backpack electrofishing (DC-pulsed). All major habitats within a stream reach of 100–300 m were sampled, and sampling was ended after two or three seine hauls yielded no new species.

To increase the spatial coverage of the sampling sites, we used fish species records collected in the Big Blue River Basin over approximately 50 years. Although changes in species abundance in the Big Blue River and its tributaries were documented over this time span (Gido et al. 2002), species occurrences within Great Plains river basins have tended to be more stable (Gido et al. 2004). Thus, we assumed that by measuring current habitat conditions, we could capture species–habitat associations based on presence/absence data from historic collections.

Environmental variables

Predictive modelling was conducted at the scale of stream segments. A stream segment was defined as a length of stream from its upstream confluence to its downstream confluence with other tributaries (mean length = 1.6 km). While the segment scale may not encompass all habitats used by freshwater fishes throughout their life stages (Schlosser & Angermeier 1995; Fausch et al. 2002), stream fishes typically move relatively little during most life stages (Skalski & Gilliam 2000) and thus should respond to factors operating at the scale of stream segments used in this study. Twenty-five landscape-scale habitat variables were quantified for each segment and used to model fish distributions (Table 2). These habitat variables were chosen to represent factors that have been linked to fish assemblage structure, including stream size and position in the watershed (Sheldon 1968; Osborne & Wiley 1992; Zorn et al. 2002), surface geology (Matthews & Robison 1988; Nelson et al. 1992; Mandrak 1995), and land use within the catchment (Roth et al. 1996; Wang et al. 1997). Associations among variables were evaluated prior to modelling. As this was partly an exploratory analysis, only highly correlated (i.e., $r > 0.85$) variables were excluded from the data set, and the same variables were used in all three modelling approaches to ensure accurate comparisons across approaches.

Environmental variables were classified at the stream segment scale for use in modelling with ARCGIS software and methods described in Arc-Hydro: Geographical Information System (GIS) for Water Resources (Maidment 2002). Measures of stream size and positioning were calculated from a modified version of the National Hydrography Dataset (USGS 1997), soil measurements were obtained from the STATSGO database (NRCS 1994), and catchment

Table 1. Summary of predictive fish model performances based on artificial neural networks (ANN), discriminant function analysis (DFA) and classification trees (CT).

Species	%	ANN			DFA			CT		
		T	O	Co	T	O	Co	T	O	Co
Lepisosteidae										
<i>Lepisosteus osseus</i>	17.5	17	38	9	19	62	6	22	43	17
Hiodontidae										
<i>Hiodon alosoides</i>	2.5	6	33	4	8	33	6	3	33	3
Clupeidae										
<i>Dorosoma cepedianum</i>	11.7	15	64	6	16	79	5	12	57	6
Cyprinidae										
<i>Campostoma anomalum</i>	65.0	39	23	50	41	27	48	33	21	55
<i>Cyprinella lutrensis</i>	95.8	10	4	100	13	7	100	6	2	100
<i>Cyprinus carpio</i>	36.7	30	36	18	35	41	22	43	55	37
<i>Hybognathus argrytus</i>	5.0	4	17	3	8	33	5	6	50	4
<i>Luxilus cornutus</i>	27.5	22	36	11	23	46	9	29	48	22
<i>Macrhybopsis aestivalis</i>	6.7	7	38	4	13	50	8	11	75	6
<i>Notemigonus crysoleucas</i>	3.3	7	50	4	7	100	3	7	100	3
<i>Notropis atherinoides</i>	7.5	8	56	3	7	33	4	6	33	4
<i>Notropis ludibundus</i>	80.0	25	8	71	30	16	63	22	16	46
<i>Notropis topeka</i>	6.7	10	75	4	15	88	7	9	38	7
<i>Phenacobius mirabilis</i>	65.0	32	15	48	37	15	60	39	27	62
<i>Phoxinus erythrogaster</i>	8.3	10	70	3	13	50	7	12	60	7
<i>Pimephales notatus</i>	50.0	40	27	40	50	38	45	41	37	45
<i>Pimephales promelas</i>	90.8	11	3	73	21	9	100	8	1	73
<i>Pimephales vigilax</i>	3.3	5	75	2	6	50	3	13	100	9
<i>Semotilus atromaculatus</i>	60.8	25	16	28	24	16	26	26	19	36
Catostomidae										
<i>Carpoides carpio</i>	40.0	22	27	13	23	25	15	15	15	15
<i>Catostomus commersoni</i>	22.5	24	56	10	35	74	16	34	70	24
<i>Ictiobus bubalus</i>	10.8	17	69	7	20	77	9	18	62	12
<i>Ictiobus cyprinellus</i>	2.5	3	67	1	7	67	4	5	100	3
<i>Moxostoma macrolepidotum</i>	6.7	9	88	2	15	88	7	8	75	3
Ictaluridae										
<i>Ameiurus melas</i>	35.0	34	40	22	41	55	23	38	45	33
<i>Ameiurus natalis</i>	31.7	32	45	18	32	47	17	28	42	22
<i>Ictalurus punctatus</i>	52.5	16	13	14	26	22	21	24	17	32
<i>Noturus flavus</i>	31.7	36	47	22	39	66	17	37	53	29
<i>Pylodictis olivaris</i>	20.8	26	52	14	28	64	13	22	60	12
Moronidae										
<i>Morone chrysops</i>	4.2	5	60	2	12	60	8	8	80	4
Centrarchidae										
<i>Lepomis cyanellus</i>	67.5	36	25	41	34	21	44	41	30	64
<i>Lepomis humilis</i>	53.3	39	34	30	44	39	34	38	28	48
<i>Lepomis macrochirus</i>	21.7	29	62	14	26	54	13	27	58	18
<i>Micropterus salmoides</i>	17.5	23	71	8	29	100	8	24	71	14
<i>Pomoxis annularis</i>	10.0	12	100	0	23	100	10	21	100	12
Percidae										
<i>Etheostoma nigrum</i>	16.7	19	60	7	32	80	16	18	60	10
<i>Etheostoma spectabile</i>	41.7	37	40	24	35	38	23	42	50	36
Sciaenidae										
<i>Aplodinotus grunniens</i>	13.3	15	56	6	23	75	11	17	50	12

Species are organised by family and arranged by phylogenetic order.

Reported values are the percentage of total error (T), omission error (O), and commission error (Co) for the different modelling approaches, as well as the frequency of species occurrence (%).

Values in bold indicate significant predictions ($P < 0.05$, Cohen's κ statistic).

land use was calculated from the USGS LULC data set (USGS 1994). LULC data were derived from Landsat Thematic Mapper imagery taken between 1988 and 1990, and were quantified at the 1:100,000 scale. This data layer included general land cover types that were less prone to misclassification than other land cover data sets available for the basin.

Statistical analyses

We compared the effectiveness of linear discriminant function analysis (DFA), CT and ANN for predicting the occurrence of 38 different fish species that occurred in greater than 5% of the collections (i.e., in three or more of the 120 collections). Model

Table 2. Segment-scale environmental variables, grouped by category, measured for 120 collection sites in the Big Blue River Basin and used to develop fish-habitat models.

Variable code	Predictor variable
Hydrological parameters	
Strahler	Strahler order of stream segment (Strahler 1957)
Link	Link magnitude (no. of stream segments upstream of a given stream segment, Scheidegger 1965)
Downorder	Strahler order of downstream segment
Dlink	Link magnitude of downstream segment
Drainage	Drainage area above segment (km ²)
Distance	Distance from Tuttle Creek Reservoir (km)
Watershed geology (mean in catchment above segment)	
Kfact	Soil erodibility factor (tonnes-unit ⁻¹ of rainfall erosion index, Wischmeier 1976)
Bd	Bulk density of soils (g·cm ⁻³)
Perm	Soil permeability (cm·h ⁻¹)
Slope	Field slope (%)
Om	Organic matter content of soils (% by weight)
Segment geology (mean of segment)	
Elevation	Elevation (m)
Gradient	Gradient (% change over length of segment)
Se_kfact	Soil erodibility factor (tonnes-unit ⁻¹ of rainfall erosion index, Wischmeier 1976)
Se_perm	Soil permeability (cm·h ⁻¹)
Se_slope	Field slope (%)
Se_om	Organic matter content of soils (% by weight)
Se_wtdep	Water table depth (m)
Land cover (% in catchment above segment)	
Agric	Agricultural land
Barren	Barren land
Forest	Forested land
Range	Rangeland
Urban	Urban land
Wetland	Wetlands
Water	Water

effectiveness was assessed by our ability to correctly classify species presence or absence. In this study, omission error referred to the misclassification of a species as absent when it was observed at a site (i.e., the complement to model sensitivity or false negative), commission error referred to the misclassification of a species as present when it was not observed at a site (i.e., the complement to model specificity or false positive) and total error referred to the total misclassification rate of a species across all sites. All error rates were expressed as percentages. For all model predictions an *a priori* decision threshold of 0.5 was used to classify a species as present or absent.

Linear discriminant function analysis

Discriminant functions were based on linear combinations of predictor variables that maximise the separation between groups (i.e., presence and absence). DFA is expected to perform similarly to logistic regression when parametric assumptions are met (Efron 1975) and therefore was representative of traditional, parametric approaches in general. DFA has been used in ecological studies of fishes (e.g., Joy & Death 2002), but is constrained by the assumptions of linear relationships among variables (e.g., Olden & Jackson 2002). All models were generated using

SPSS, version 11.0. Within-groups covariance matrix was used and prior probabilities were computed from group sizes (i.e., species prevalence). The importance of independent variables in the DFA models were expressed on a scale of negative one to one using standardised discriminant function coefficients.

Classification trees

Classification trees are a powerful nonparametric approach (Breiman et al. 1984) for modelling ecological data (De'ath & Fabricius 2000). Classification trees were constructed using a recursive partitioning algorithm that repeatedly splits the data set into two mutually exclusive groups; where each split sought to minimise the misclassification rate and create groups that were as homogeneous as possible with respect to the response variable (i.e., species presence/absence). Splitting continued with each new group until no new split yielded a significant decrease in misclassification rate, or the group contained a small number of observations (*N* = 10). Prior probabilities were estimated, and goodness-of-fit was calculated using the Gini index, the measure of goodness-of-fit favoured by the developers of CT (Breiman et al. 1984). The importance of each habitat variable in the CT models for each fish species was estimated by summing the

changes in misclassification for each split across all groups, and was expressed on a scale from 0 to 100, with high values indicating greater importance in the model (Breiman et al. 1984). Classification trees were generated using MatLab, version 6.0 (Novi, MI, USA).

Artificial neural networks

Artificial neural networks have shown great promise for modelling nonlinear ecological data and developing prediction or classification rules for many taxa (Lek & Guégan 2000), including fish species distributions (e.g., Olden & Jackson 2001). In this study we used one-hidden-layer feedforward neural networks trained by the backpropagation algorithm (Bishop 1995). The architecture of this network consists of a single input, hidden and output layer, with each layer containing one or more neurones. The input layer contains p neurones, each of which represents one of the 25 environmental variables. The number of hidden neurones in the neural network varies and was chosen to minimise the trade-off between network bias and variance by comparing the performances of different cross-validated networks, with one to 25 hidden neurones and choosing the number that produced the greatest network performance. The output layer contains a single neurone representing the predicted probability of species presence. Additional bias neurones with a constant output (equal to one) are added to the hidden and output layers. Learning rate (η) and momentum (α) parameters (varying as a function of network error) were included during network training to ensure a high probability of global network convergence and a maximum of 1000 iterations for the backpropagation algorithm to determine the optimal axon weights. Prior to training the network, the independent variables were converted to z -scores to standardise the measurement scales of the inputs into the network, and thereby ensure that same percentage change in the weighted sum of the inputs caused a similar percentage change in the unit output.

The explanatory importance of the environmental variables for predicting fish species occurrence was quantified by calculating the product of the hidden-input and hidden-output connection weights between each input neurone and output neurone and then summing the products across all hidden neurones. This procedure was repeated for each environmental variable and the relative contributions of the variables were calculated by dividing the absolute value of each variable contribution by the grand mean (sum of all absolute variable contributions). The relative contributions of each environmental variable were subsequently assessed for their statistical significance using a randomisation test (see Olden et al. 2004 for more details). All ANN were conducted using computer macros written in the MatLab programming language, version 6.0.

Model validation

Species models were evaluated using a cross-validation procedure, in which one site was excluded, a model was constructed using $N-1$ sites, and the excluded site was predicted using this model. This procedure was repeated N times. Cohen's κ statistic (Titus et al. 1984) was used to assess whether the association between predicted and observed prevalence differed from expectations based on chance alone (Olden et al. 2002), because this method is relatively independent of species prevalence (Manel et al. 2001).

Model comparisons and variable importance

We compared the three modelling techniques with Wilcoxon signed ranks tests that examined differences in total, omission and commission error rates among modelling approaches. A Bonferroni corrected alpha was used to account for multiple comparisons. To examine the relative contribution of environmental variables to species predictions generated from the three modelling approaches, we ranked the variables from 25 to 1 (25 = most important variable, 1 = least important). Mean ranks across species models were calculated to identify the environmental variables that exhibited strong and consistent relationships with fish species occurrence. Spearman rank-correlation coefficients were calculated to examine concordance of variable importance across individual modelling techniques for all species.

Results

Comparisons of modelling approaches

Thirty-eight species occurred in three or more collections and were modelled using the different approaches. Cohen's κ statistic indicated that the ANN, DFA and CT approaches produced 35, 29 and 24 species models, respectively, which were significantly better than those expected by chance. Predictability of species presence or absence varied greatly, and as expected was influenced by patterns of species prevalence (Fig. 2). Widely and narrowly distributed species typically had low total error rates, whereas species with intermediate prevalence across sites ranged widely in total error rates. Mean total error rates across the 38 species models ranged from 19.9% for ANN models, 21.4% for CT models and 23.9% for DFA models. Differences in omission and commission error rates across the approaches were much greater but all were strongly correlated with species prevalence; omission error rates increased and commission error rates decreased with declining species prevalence.

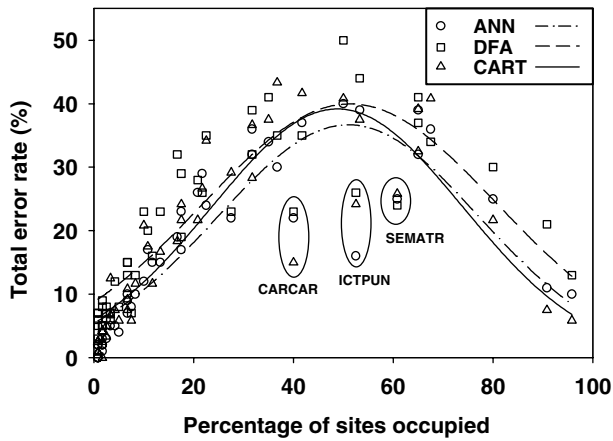


Fig. 2. Comparison of total error rates versus the total number of sites occupied (i.e., species prevalence) for the three different modelling techniques. Lines represent quadratic equations fit to the relation between error rate and species prevalence for 38 species. Examples of species that fall below curves for all individual approaches are circled and labelled with species codes (CARCAR, *Carpionodes carpio*, ICTPUN, *Ictalurus punctatus*, SEMATR, *Semotilus atromaculatus*).

Analysis of error rates among the families of species modelled revealed no clear trends in species predictability related to phylogeny. However, some species models performed better than others with respect to prevalence, across the three modelling approaches. Models generated for species with strict habitat requirements tended to perform better (e.g., had lower error rates than expected based on their prevalence) across all three modelling techniques. Examples of such species include *Carpionodes carpio*, *Ictalurus punctatus* and *Semotilus atromaculatus*, which were present in approximately 50% of the collection records and had average correct classification rates of 80%, 78% and 75%, respectively, across modelling approaches. Conversely, generalist species such as *Lepomis humilus* and *Lepomis macrochirus* were observed in similar percentages of collection records, but generated less predictable models (average correct classification across modelling approaches equalled 60% for both species). Mean Cohen’s κ statistic values also

tended to be higher in models developed for the three species with strict habitat requirements than those developed for the two generalist species.

Total error rates for the three species of special conservation concern in the basin were low across modelling approaches (4–8% for *H. argrytus*, 7–15% for *N. topeka*, 10–13% for *P. erythrogaster*; Table 1). However, omission error rates tended to be high in models generated for *N. topeka* and *P. erythrogaster*, indicating that the models were under-predicting the occurrence of these species. Cohen’s κ statistic values for models developed for these species were generally high (mean κ across models 0.42 for *H. argrytus*, 0.32 for *N. topeka*, 0.29 for *P. erythrogaster*) across modelling approaches, with the exception of the DFA model developed for *N. topeka*, which did not significantly differ from results expected by chance.

Across all species, ANN generally produced the greatest number of models (35 of 38 species) whose results were significantly better than those expected by chance and had the lowest error rates, although the magnitude of differences among approaches was often small. Wilcoxon signed ranks tests showed that total error rates generated by ANN and CT models were significantly (Bonferroni corrected $P < 0.05$) lower than those generated by DFA models (Table 3). Commission error rates were lowest for ANN, intermediate for CT and highest for DFA models. Omission error rates generated by ANN models were significantly lower than those generated by DFA models, but not different from CT models. ANN models most often generated the lowest error rates for a given species, with the lowest total error rate for 14 of the 38 species, the lowest omission error for 18 of 38 species, and the lowest commission error for 19 of 38 species (Table 1). The ANN approach also generated the most accurate models for two (*H. argrytus* and *P. erythrogaster*) of the three species of special conservation concern in the basin. In particular, the ANN approach was able to predict the distribution of *H. argrytus* with a high degree of accuracy (96% species occurrence, 83% species presence, 97% species absence).

Table 3. Pairwise comparisons of total, commission, and omission misclassification errors from the three modelling approaches across species using Wilcoxon signed-rank tests.

Comparison	Total				Commission				Omission			
	Number of positive ranks	Number of ties	Z	P-value	Number of positive ranks	Number of ties	Z	P-value	Number of positive ranks	Number of ties	Z	P-value
ANN versus DFA	5	2	4.58	<0.001	12	1	2.93	0.003	9	3	2.92	0.003
ANN versus CT	16	0	1.41	0.157	26	3	-3.60	<0.001	26	1	-1.40	0.161
DFA versus CT	29	0	-2.85	0.004	33	1	-4.64	<0.001	27	2	-2.33	0.020

ANN, artificial neural networks; DFA, discriminant function analysis; CT, classification trees.

Number of positive ranks indicates the number of species models for which the first variable had greater percent misclassification than the second variable (e.g., total error rates for ANN were greater than DFA for five species models).

Values in bold represent significant differences in error rates between approaches after a Bonferonni correction.

Variable importance for predicting species occurrence

Although there was considerable variation in the importance of environmental variables across the three approaches, measures of stream size and position in the watershed (drainage, link, dlink) and variables reflective of the underlying geology of the basin (kfact, perm) were generally ranked highest across the 38 species models (Fig. 3). However, mean ranked importance of environmental variables in the models ranged between 6.8 and 21.9, and 24 of the 25 variables were ranked as the most important in at least one model. Thus, most variables were important in some species models, and no variable was important in all species models. Moreover, weak associations in variable importance among methods showed that

variable contributions differed among modelling approaches (Table 4).

Discussion

Artificial neural networks, classification trees and linear DFA successfully predicted the presence and absence of the majority of fish species. The high correlation between prevalence and classification success across species and modelling approaches has been previously reported and is partly attributed to insufficient data points available to accurately classify habitats in which species are either present or absent (Scott et al. 2002), and the increased probability of correctly predicting the occurrence of rare and very abundant species (Olden et al. 2002). From a conservation standpoint, models

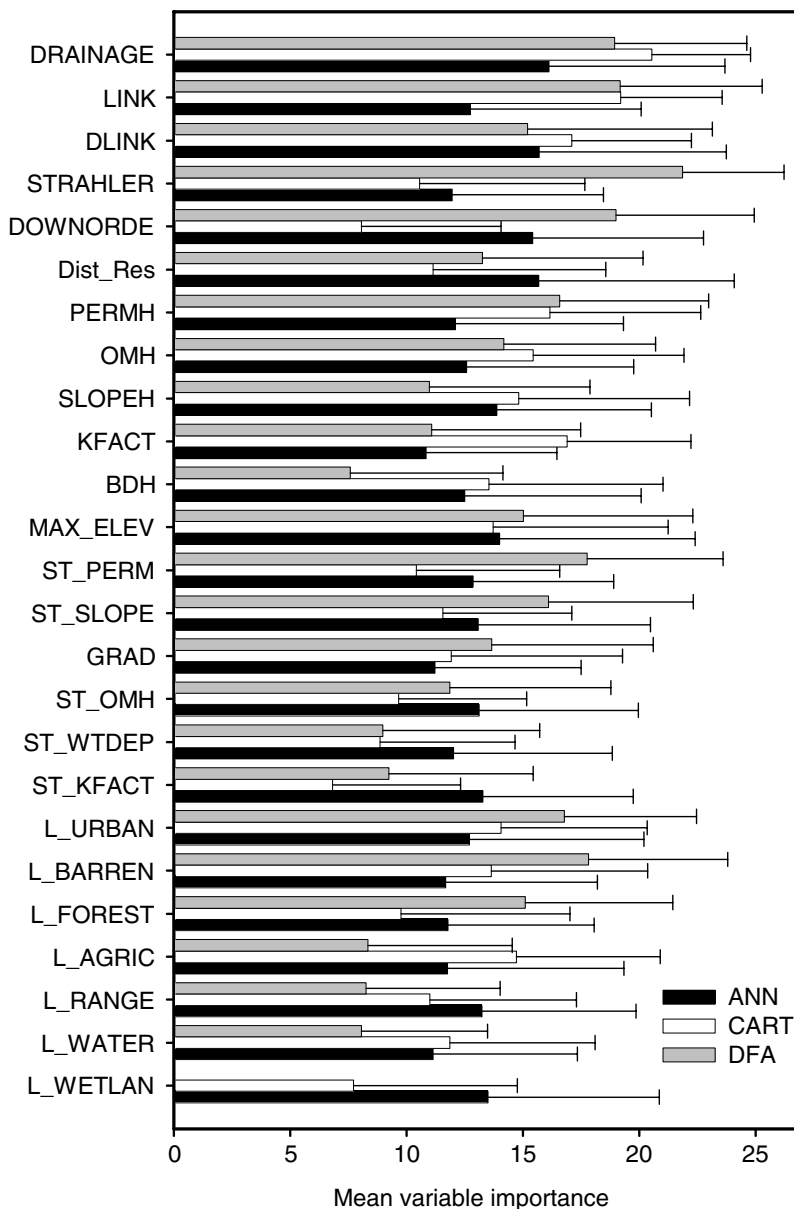


Fig. 3. Mean importance of each environmental variable, based on 38 species models, for ANN (black), CT (white), and DFA (grey) modelling approaches. Environmental variables were ranked for each species, with 25 given to the most important variable in the model, and 1 given to the least important. Variables are grouped by category, listed in Table 2, and then listed in decreasing order by mean rank. The variable WETLAND was not included in DFA models. Whiskers represent 1 SD.

Table 4. Spearman rank correlations comparing importance of environmental variables in artificial neural network (ANN), discriminant function analysis (DFA), and classification tree (CT) modelling approaches.

Species	ANN versus DFA	ANN versus CT	DFA versus CT
Lepisosteidae			
<i>Lepisosteus osseus</i>	-0.100	0.139	-0.179
Hiodontidae			
<i>Hiodon alosoides</i>	-0.143	NA	NA
Clupeidae			
<i>Dorosoma cepedianum</i>	-0.144	-0.030	-0.676
Cyprinidae			
<i>Campostoma anomalum</i>	-0.134	0.157	0.079
<i>Cyprinella lutrensis</i>	0.164	NA	NA
<i>Cyprinus carpio</i>	0.172	-0.428	-0.079
<i>Hybognathus argyritus</i>	0.206	0.005	-0.607
<i>Luxilus cornutus</i>	0.225	-0.156	-0.117
<i>Macrhybopsis aestivalis</i>	0.303	0.048	0.228
<i>Notemigonus crysoleucas</i>	-0.099	NA	NA
<i>Notropis atherinoides</i>	0.194	-0.047	-0.535
<i>Notropis ludibundus</i>	-0.194	0.011	-0.511
<i>Notropis topeka</i>	0.036	0.013	-0.495
<i>Phenacobius mirabilis</i>	0.017	0.198	-0.147
<i>Phoxinus erythrogaster</i>	-0.109	0.245	-0.346
<i>Pimephales notatus</i>	0.283	0.461	-0.218
<i>Pimephales promelas</i>	-0.193	0.054	-0.252
<i>Pimephales vigilax</i>	0.183	NA	NA
<i>Semotilus atromaculatus</i>	0.065	0.003	-0.133
Catostomidae			
<i>Carpiodes carpio</i>	-0.071	-0.362	0.277
<i>Catostomus commersoni</i>	-0.440	0.027	-0.362
<i>Ictiobus bubalus</i>	0.080	-0.244	0.197
<i>Ictiobus cyprinellus</i>	-0.110	NA	NA
<i>Moxostoma macrolepidotum</i>	-0.068	NA	NA
Ictaluridae			
<i>Ameiurus melas</i>	-0.023	0.181	-0.230
<i>Ameiurus natalis</i>	0.010	0.058	-0.236
<i>Ictalurus punctatus</i>	0.105	-0.186	-0.152
<i>Noturus flavus</i>	-0.106	0.279	-0.244
<i>Pylodictis olivaris</i>	-0.050	0.042	-0.083
Moronidae			
<i>Morone chrysops</i>	0.043	NA	NA
Centrarchidae			
<i>Lepomis cyanellus</i>	-0.161	-0.374	-0.472
<i>Lepomis humilis</i>	-0.078	0.392	-0.066
<i>Lepomis macrochirus</i>	-0.028	-0.442	-0.320
<i>Micropterus salmoides</i>	-0.445	-0.255	-0.231
<i>Pomoxis annularis</i>	0.023	NA	NA
Percidae			
<i>Etheostoma nigrum</i>	-0.030	-0.418	0.008
<i>Etheostoma spectabile</i>	0.101	0.038	0.033
Sciaenidae			
<i>Aplodinotus grunniens</i>	-0.084	NA	NA

Species are organised by family and arranged by phylogenetic order. As CT models did not generate variable importance metrics for species where no variable induced a split in the classification tree, no comparison was made for some species, designated NA.

Values in bold indicate significant correlations ($P < 0.05$).

that produce low omission error are of the most interest, because they can be used to identify areas of suitable habitat for rare species across the landscape. Moreover, species presence data is considerably less biased compared with species absence data that can arise solely from the failure to detect a species at a particular site (Scott et al. 2002). However, models that over-predict occurrences (i.e., high commission error) may not be biologically meaningful. Thus, models with the greatest utility for conservation are those that are ecologically interpretable and predict

occurrences of rare species while minimising commission error.

Although the magnitude of differences was often small, ANN models generated the lowest omission error rates. These findings correspond to those of Olden & Jackson (2002), who compared the predictability of four modelling approaches (the same approaches examined here in addition to logistic regression analysis) in predicting fish distributions in 286 temperate lakes in south-central Ontario, Canada. Like this study, Olden & Jackson (2002) found the overall

predictive performance of four modelling approaches was similar across species, but that the ANN modelling approach generated the greatest number of significant models and lower omission error rates than the other approaches. Given the disparity in environmental gradients between prairie streams and temperate lakes, this suggests ANN models may generally outperform other modelling approaches across a wide range of environmental gradients.

Despite the dynamic abiotic characteristics of prairie streams (Dodds et al. 2004) and the prevalence of generalist species (Fausch & Bestgen 1997), we were able to generate significant models for the majority of species at the scale of stream segments. While our approach was similar to that of Olden & Jackson (2002), we expected to observe weaker species–habitat associations because of the inherent differences between hierarchically distributed river basin networks (Frissell et al. 1986) and spatially isolated lake systems. Our results were comparable with those presented by Olden & Jackson (2002), but as expected, our observed error rates for prairie stream fishes were higher than for lake species (notably, species prevalence was similar in the two data sets). For example, using ANN models the mean total error rate across 29 lake species was 15.8%, versus a mean total error rate of 19.9% across the prairie stream species analysed in this study. There also was a greater discrepancy in omission error rates, which averaged 29.2% across lake species versus 44.6% across prairie stream fishes.

Our results suggest that species predictability is associated with the autecology of the species being modelled. Lotic species with strict habitat requirements should more strongly respond to the environmental gradients regulating the distribution of fishes within river basin networks than more wide-ranging species, and consequently may be more likely to co-vary with environmental variables used in predictive modelling. For example, of the three species highlighted in Fig. 2, both *C. carpio* and *I. punctatus* are large-river species that were only collected in main stem rivers of the Big Blue River Basin. The third species, creek chub (*S. atromaculatus*), as its name implies, typically occur in tributaries, and are rarely found in large rivers (Cross & Collins 1995). These species may be classified as ecological specialists in that they respond to a gradient of stream size across the river basin. In contrast, two *Lepomis* species (*L. humilis* and *L. macrochirus*) occurred in approximately the same number of collections, but their models had much greater total error rates. These species are considered to be ecological generalists that tolerate wide-ranging conditions and are present in numerous habitat types (Cross & Collins 1995). Consequently, they may be more difficult to model

than species with strict habitat requirements that can be explained with only one or a few environmental variables (Porter et al. 2000).

Model results were mixed for the species of special conservation concern. The presence of *H. argrytus* was most accurately predicted using the ANN approach, the presence of *N. topeka* was most accurately predicted using the CT approach, and the presence of *P. erythrogaster* was most accurately predicted by the DFA approach. Moreover, the presence of *H. argrytus* was successfully predicted >83% of the time using ANN, but ANN models were only successful at predicting the occurrence of *N. topeka* and *P. erythrogaster* 62% and 50% of the time, respectively. These results illustrate that the predictability of rare species is highly variable and should be interpreted with caution; a finding that is supported by the varying success of recent modelling efforts that used the approaches compared in this study (ANN: Mastrorillo et al. 1997; CT: Rejwan et al. 1999; DFA: Scheller et al. 1999).

Rare species with strict habitat requirements are often among the first species to be adversely impacted by anthropogenic changes (Allan & Flecker 1993). Whereas a species with specific habitat requirements should be more easily modelled, the lack of data points available for threatened or endangered species limits the usefulness of these modelling techniques for rare species (Scott et al. 2002). In this study, *H. argrytus*, *N. topeka* and *P. erythrogaster* occurred in only 5%, 7% and 8%, respectively, of the 120 collection records used in modelling; additional sampling records of *N. topeka* and *P. erythrogaster* likely would have allowed more accurate predictions to be developed for these species. In light of such data limitations, conducting additional surveys prior to beginning modelling efforts may improve predictive accuracy for rare species. Given the often endemic nature of rare species, undertaking a more hierarchical approach to modelling fish species distributions also may assist in predicting the occurrence of rare fishes (Quist et al. 2005). Furthermore, the development of ecologically relevant predictive models that maximise the correct classification of species presence (i.e., minimise omission error), albeit at the expense of increased commission error, will be an important development for modelling of rare species. This can be accomplished by incorporating misclassification costs into the modelling process. Studies typically apply equal costs to omission and commission error; however, in practice it may be advantageous to assign more appropriate costs to the misclassifications if such information is available. For example, more false presences for a rare species may be tolerated to minimise the risk of failing to protect that species; thus greater costs could be assigned to omission errors during the modelling process.

All independent variables used in this study were GIS-derived landscape variables quantified at the stream segment scale. Such variables may either be directly responsible for shaping fish assemblage structure (e.g., Nelson et al. 1992; Osborne & Wiley 1992; Mandrak 1995), or may simply be correlated with other habitat features at finer spatial scales that are known to influence fish distribution and abundance. For example, Zorn et al. (2002) accurately predicted species distributions in Michigan based on soil permeability, slope and drainage area above collection sites, which strongly correlated with hydrology and water temperature. The ability of landscape variables to successfully predict fish distributions provides a useful tool in instances where limited resources do not allow sufficient finer-scale data to be collected. However, researchers must be aware that these variables may be surrogates for the actual factors driving species distributional patterns. Nevertheless, our results demonstrated that significant models could be produced for many fish species across wide ranges of prevalence and habitat preferences using only GIS-derived habitat variables as predictors.

Several highly ranked variables were reflective of either stream size or position in the watershed, or the underlying geology of the basin. These results agree with previous studies (Minckley 1959; Gido et al. 2002; Falke & Gido in press) in the Big Blue River Basin, which linked fish assemblage structure primarily to stream size, location and substrate type. Interestingly, all segment measurements of geological variables had low mean rankings. Whereas segment-measured variables were, to some extent, associated with watershed-level geological variables, the consistent selection of these watershed-level variables highlight the potential utility of large-scale proxies in modelling fish distributions. Such an approach represents the hierarchical manner by which basin geology impacts fish distributions at smaller spatial scales (Fausch et al. 2002). As noted earlier, the Big Blue River Basin was influenced by glaciations, resulting in distinct geology among watersheds, which proved in this study to be key predictors for many species in the basin. Additional studies in areas with less topographic relief and habitat heterogeneity are needed to assess the ability of these modelling techniques to predict fishes in other prairie environments.

Mean rankings of predictor variables suggest some variables were more frequently important across all models, but it appears that almost all variables (24 of 25 in individual species models) were important in at least one species model. Like Olden & Jackson (2002), we found a lack of concordance in variable importance among the different approaches, which indicated different combinations of variables were being chosen to predict species occurrence, despite the

fact that modelling techniques were generating roughly equivalent error rates. Some variance in variable importance was probably because of moderate levels of correlation among variables, as only highly correlated environmental variables were removed from the analysis. Overall, it is difficult to generalise the importance of individual variables because multiple variable interactions were implicitly accounted for in the ANN and CT models (multiple variable interactions were not taken into account in DFA modelling). When modelling large numbers of species, it also was important to consider the trade-off of including enough variables to represent all species responses to environmental gradients, but not so many that the biological relevance of these variables becomes unclear. In instances where specific species are modelled as part of the development of conservation strategies, a more rigorous screening of variables may be appropriate and a consensus of results from multiple models should be used.

The development of significant predictive models for most species using only GIS-derived habitat variables is promising, because of the aforementioned time and monetary advantages of these data, and, as was the case in this study, finer-scale data are often not available. Researchers (Smith & Powell 1971; Tonn 1990; Quist et al. 2005) have hypothesised that environmental filters operate at successive scales, whereby large-scale variables exert an initial influence on aquatic biota, and finer-scale variables have a modifying influence that determine presence-absence at a specific locality. This suggests the inclusion of field data or other finer-scale measurements would have increased our predictive ability. Although in some cases the inclusion of field-measured variables generates only marginal improvements in classification success (e.g., Porter et al. 2000), a combination of landscape and local habitat variables may likely generate the most accurate species models, and finer-scale data should be included if available.

Conclusion

This study has demonstrated the relatively high predictability of Great Plains fish species occurrence, and highlights the utility of ANN over other commonly used modelling approaches. This is encouraging for regional managers responsible for conserving prairie stream fishes that have suffered major declines in abundance and overall range. Although we were less successful in predicting species occurrences than results that have been reported for other systems (e.g., temperate lakes), our models did predict the occurrence of at least two species of special conservation concern in the region (*H. argrytus* and

N. topeka) with a relatively high degree of accuracy (>60% occurrences correctly predicted). Developing meaningful conservation plans for rare species requires knowledge of both the organism and the environmental features that shape its distribution. In instances where these data are not available, measures such as collecting additional data may be required to generate accurate models for rare species.

Future advances in the prediction of stream fishes will likely come from a greater understanding of the processes and scales that regulate fish distributions across the landscape, as well as improvements in the ability of statistical techniques to elucidate these relationships. In particular, a better understanding of spatial and temporal variation in species distributions are needed, as well as an understanding of how the importance of environmental gradients that regulate distributions vary with respect to scale (Angermeier et al. 2002).

Resumen

1. Predecir la distribución de las especies tiene importantes implicaciones para la conservación y gestión de los peces de agua dulce, particularmente en regiones tales como el Great Plains de los Estados Unidos, donde los impactos humanos han resultado en el declive o en la extirpación de numerosas especies nativas.

2. Hay varias aproximaciones estadísticas para construir modelos de distribución; la precisión de cada una probablemente dependa de la naturaleza de los gradientes ambientales, de las respuestas de las especies a esos gradientes y de la extensión espacial de los modelos. Por ello, es importante comparar aproximaciones múltiples a través de especies y hábitats para identificar la modelación más efectiva. Utilizando características derivadas de GIS de segmentos de río como predictores, analizamos la realización del modelo con tres metodologías (análisis de función discriminante lineal, árboles de clasificación y redes neuronales artificiales) para predecir la ocurrencia de 38 especies de peces en la cuenca del río Great Plains.

3. Los resultados mostraron que todas las aproximaciones predicen la ocurrencia de las especies con relativo éxito. En general, las redes neuronales artificiales fueron los mejores modelos ya que generaron los modelos más significativos (35 de 38 especies) y predicen de forma más precisa la presencia de especies para un mayor número de especies (clasificación correcta corregida = 81.1%). La importancia de las variables GIS para predecir la ocurrencia de peces de río varió entre especies y técnicas de modelado pero fueron generalmente buenos predictores de las distribuciones de las especies incluyendo la especie amenazada *Notropis topeka*.

4. En resumen, los modelos predictivos deberían ser vistos como metodologías competitivas y complementarias para establecer enlaces cuantitativos entre las especies de peces y sus ambientes. Nuestro estudio demuestra la utilidad potencial de esta aproximación como guía para los esfuerzos de conservación de los peces de río de la región del Great Plains (USA).

Acknowledgements

We thank Kristen Hase from the Kansas Department of Wildlife and Parks for providing data used in this study. Scott Sowa and Gust Annis of the Missouri Resource Assessment Partnership provided assistance in developing GIS variables. Comments by Walter Dodds, John Baker, and anonymous reviewers greatly improved this manuscript. This work was supported by a grant from the U.S. Geological Survey National Gap Analysis Program to KBG.

References

- Allan, J.D. & Flecker, A.S. 1993. Biodiversity conservation in running waters. *BioScience* 43: 32–43.
- Angermeier, P.L., Krueger, K.L. & Dolloff, C.A. 2002. Discontinuity in stream-fish distributions: implications for assessing and predicting species occurrence. In: Scott, J.M., Heglund, P.J. & Morrison, M.L., eds. *Predicting species occurrences: issues of accuracy and scale*. Washington, DC: Island Press, pp. 519–527.
- Berkman, H.E. & Rabeni, C.F. 1987. Effect of siltation on stream fish communities. *Environmental Biology of Fishes* 18: 285–294.
- Bishop, C.M. 1995. *Neural networks for pattern recognition*. New York, NY: Oxford University Press.
- Breiman, L., Friedman, J.H., Olshen, A. & Stone, C.G. 1984. *Classification and regression trees*. Belmont, CA: Wadsworth International Group.
- Brosse, S. & Lek, S. 1999. Predicting fish distribution in a mesotrophic lake by hydroacoustic survey and artificial neural networks. *Limnology and Oceanography* 44: 1293–1303.
- Cross, F.B. & Collins, J.T. 1995. *Fishes in Kansas*, 2nd edn. Lawrence, KS: University of Kansas Natural History Museum.
- Cross, F.B. & Moss, R.E. 1987. Historic changes in fish communities and aquatic habitats in plains streams of Kansas. In: Matthews, W.J. & Heins, D.C., eds. *Community and evolutionary ecology of North American stream fishes*. Norman, OK: University of Oklahoma Press, pp. 155–165.
- Cross, F.B., Moss, R.E. & Collins, J.T. 1985. Assessment of dewatering impacts on stream fisheries in the Arkansas and Cimarron Rivers. Pratt, KS: Kansas Fish and Game Commission, Nongame Contract 46.
- De'ath, G. & Fabricius, K.E. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81: 3178–3192.
- Dodds, W.K., Gido, K.B., Whiles, M.R., Fritz, K.M. & Matthews, W.J. 2004. Life on the edge: ecology of prairie streams. *BioScience* 54: 205–216.
- Efron, B. 1975. The efficiency of logistic regression compared to normal discriminant analysis. *Journal of the American Statistical Association* 70: 892–898.
- Eschner, T.R., Hadley, R.F. & Crowley, K.D. 1983. Hydrologic and morphologic changes in channels of the Platte River basin in Colorado, Wyoming, and Nebraska: a historical perspective. Washington, DC: US Geological Survey, Professional Paper 1277-A.
- Falke, J.A. & Gido, K.B. (in press). Spatial effects of reservoirs on fish assemblages in Great Plains streams in Kansas, USA. *River Research and Applications*.

- Fausch, K.D. & Bestgen, K.R. 1997. Ecology of fishes indigenous to the central and southwestern Great Plains. In: Knopf, F.L. & Samson, F.B., eds. Ecology and conservation of Great Plains vertebrates. New York, NY: Springer-Verlag, pp. 131–166.
- Fausch, K.D. & Bramblett, R.G. 1991. Disturbance and fish communities in intermittent tributaries of a western Great Plains river. *Copeia* 1991: 659–674.
- Fausch, K.D., Torgersen, C.E., Baxter, C.V. & Li, H.W. 2002. Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes. *BioScience* 52: 483–498.
- Frissell, C.A., Liss, W.J., Warren, C.E. & Hurley, M.D. 1986. A hierarchical framework for stream habitat classification: viewing streams in a watershed context. *Environmental Management* 10: 199–214.
- Gido, K.B., Guy, C.S., Strakosh, T.R., Bernot, R.J., Hase, K.J. & Shaw, M.A. 2002. Long-term changes in the fish assemblages of the Big Blue River basin 40 years after the construction of Tuttle Creek Reservoir. *Transactions of the Kansas Academy of Science* 105: 193–208.
- Gido, K.B., Schaefer, J.F. & Pigg, J. 2004. Patterns of fish invasions in the Great Plains of North America. *Biological Conservation* 118: 121–131.
- Guisan, A. & Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147–186.
- Hartig, A.L. & Fausch, K.D. 2002. Minimum habitat requirements for establishing translocated cutthroat trout populations. *Ecological Applications* 12: 535–551.
- Haslouer, S.G., Eberle, M.E., Edds, D.E., Gido, K.B., Mamoliti, C.S., Triplet, J.R., Collins, J.T., Distler, D.A., Huggins, D.G. & Stark, W.J. 2005. Current status of native fish species in Kansas. *Transactions of the Kansas Academy of Science* 108: 32–46.
- Joy, M.K. & Death, R.G. 2002. Predictive modeling of freshwater fish as a biomonitoring tool in New Zealand. *Freshwater Biology* 47: 2261–2275.
- Labbe, R. & Fausch, K.D. 2000. Dynamics of intermittent stream habitat regulate persistence of a threatened fish at multiple scales. *Ecological Applications* 10: 1774–1791.
- Lek, S. & Guégan, J.F. 2000. Artificial neuronal networks, applications to ecology and evolution. New York, NY: Springer-Verlag.
- Magnuson, J.J., Tonn, W.M., Banerjee, A., Toivonen, J., Sanchez, O. & Rask, M. 1998. Isolation vs. extinction in the assembly of fishes in small northern lakes. *Ecology* 79: 2941–2956.
- Maidment, D.R. 2002. Arc Hydro: GIS for Water Resources. Redlands, CA: ESRI Press.
- Mandrak, N.E. 1995. Biogeographic patterns of fish species richness in Ontario lakes in relation to historical and environmental factors. *Canadian Journal of Fisheries and Aquatic Sciences* 52: 1462–1474.
- Manel, S., Dias, J. & Ormerod, S.J. 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling* 120: 337–347.
- Manel, S., Williams, H.C. & Ormerod, S.J. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* 38: 921–931.
- Mastrorillo, S., Lek, S., Dauba, F. & Belaud, A. 1997. The use of artificial neural networks to predict the presence of small-bodied fish in a river. *Freshwater Biology* 38: 237–246.
- Matthews, W.J. 1987. Physiochemical tolerance and selectivity of stream fishes as related to their geographic ranges and local distributions. In: Matthews, W.J. & Heins, D.C., eds. Community and evolutionary ecology of North American stream fishes. Norman, OK: University of Oklahoma Press, pp. 111–120.
- Matthews, W.J. 1988. North American prairie streams as systems for ecological study. *Journal of the North American Benthological Society* 7: 387–409.
- Matthews, W.J. & Robison, H.W. 1988. The distribution of the fishes of Arkansas: a multivariate approach. *Copeia* 1988: 358–374.
- Matthews, W.J., Cashner, R.C. & Gelwick, F.P. 1988. Stability and persistence of fish faunas and assemblages in three mid-western streams. *Copeia* 1988: 947–955.
- Meador, M.R., Cuffney, T.F. & Gurtz, M.E. 1993. Methods for sampling fish communities as part of the National Water-Quality Assessment Program. Denver, CO: US Geological Survey, Open-File Report, pp. 93–104.
- Minckley, W.L. 1959. Fishes of the Big Blue River basin, Kansas. Kansas University Museum of Natural History Publications 11: 401–442.
- Nelson, R.L., Platts, W.S., Larsen, D.P. & Jensen, S.E. 1992. Trout distribution and habitat in relation to geology and geomorphology in the North Fork Humboldt River drainage, northeastern Nevada. *Transactions of the American Fisheries Society* 121: 405–426.
- NRCS 1994. State soil geographic (STATSGO) database for Kansas. Fort Worth, TX: Natural Resources Conservation Service.
- Oberdorff, T., Pont, D., Hugueny, B. & Chessel, D. 2001. A probabilistic model characterizing fish assemblages of French rivers: a framework for environmental assessment. *Freshwater Biology* 46: 399–415.
- Olden, J.D. & Jackson, D.A. 2001. Fish-habitat relationships in lakes: gaining predictive and explanatory insight by using artificial neural networks. *Transactions of the American Fisheries Society* 130: 878–897.
- Olden, J.D. & Jackson, D.A. 2002. A comparison of statistical approaches for modeling fish species distributions. *Freshwater Biology* 47: 1976–1995.
- Olden, J.D., Jackson, D.A. & Peres-Neto, P.R. 2002. Predictive models of fish species distributions: a note on proper validation and chance predictions. *Transactions of the American Fisheries Society* 131: 329–336.
- Olden, J.D., Joy, M.K. & Death, R.G. 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling* 178: 389–397.
- Osborne, L.L. & Wiley, M.J. 1992. Influence of tributary spatial position on the structure of warmwater fish communities. *Canadian Journal of Fisheries and Aquatic Sciences* 49: 671–681.
- Porter, M.S., Rosenfeld, J. & Parkinson, E.A. 2000. Predictive models of fish species distribution in the Blackwater drainage, British Columbia. *North American Journal of Fisheries Management* 20: 349–359.

- Quist, M.C., Rahel, F.J. & Hubert, W.A. 2005. Hierarchical faunal filters: an approach to assessing effects of habitat and nonnative species on native fishes. *Ecology of Freshwater Fish* 14: 24–39.
- Rathert, D., White, D., Sifneos, J.C. & Hughes, R.M. 1999. Environmental correlates of species richness for native freshwater fish in Oregon, USA. *Journal of Biogeography* 26: 257–273.
- Rejwan, C., Collins, N.C., Brunner, J., Shuter, B.J. & Ridgway, M.S. 1999. Tree regression analysis on the nesting habitat of smallmouth bass. *Ecology* 80: 341–348.
- Roth, N.E., Allan, J.D. & Erickson, D.L. 1996. Landscape influences on stream biotic integrity assessed at multiple spatial scales. *Landscape Ecology* 11: 141–156.
- Samson, F. & Knopf, F. 1994. Prairie conservation in North America. *BioScience* 44: 418–421.
- Scheidegger, A.E. 1965. The algebra of stream-order numbers. Washington, DC: US Geological Survey, Professional Paper 525B.
- Scheller, R.M., Snarski, V.M., Eaton, J.G. & Oehlert, G.W. 1999. An analysis of the influence of annual thermal variables on the occurrence of fifteen warmwater fishes. *Transactions of the American Fisheries Society* 128: 257–264.
- Schlosser, I.J. 1991. Stream fish ecology: a landscape perspective. *BioScience* 41: 704–712.
- Schlosser, I.J. & Angermeier, P.L. 1995. Spatial variation in demographic processes in lotic fishes: conceptual models, empirical evidence, and implications for conservation. In: Nielsen, J.L. & Powers, D.A., eds. *Evolution and the aquatic ecosystem: defining unique units in population conservation*. Bethesda, MD: American Fisheries Society, Symposium 17, pp. 392–401.
- Scott, J.M., Heglund, P.J. & Morrison, M.L. (eds) 2002. *Predicting species occurrences: issues of accuracy and scale*. Island Press, Washington, DC.
- Sheldon, A.L. 1968. Species diversity and longitudinal succession in stream fishes. *Ecology* 49: 193–198.
- Shrank, S.J., Guy, 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.
- Skalski, G.T. & Gilliam, J.F. 2000. Modeling diffusive spread in a heterogeneous population: a movement study with stream fish. *Ecology* 81: 1685–1700.
- Smale, M.A. & Rabeni, C.F. 1995. Hypoxia and hyperthermia tolerances of headwater stream fishes. *Transactions of the American Fisheries Society* 124: 698–710.
- Smith, C.L. & Powell, C.R. 1971. The summer fish communities of Brier Creek, Marshall County. *Oklahoma American Museum Novitates* 2458: 1–30.
- Strahler, A.N. 1957. Quantitative analysis of watershed geomorphology. *Transactions of the American Geophysical Union* 38: 913–920.
- Titus, K., Mosher, J.A. & Williams, B.K. 1984. Chance-corrected classification for use in discriminant analysis: ecological applications. *American Midland Naturalist* 111: 1–7.
- Tonn, W.M. 1990. Climate change and fish communities: a conceptual framework. *Transactions of the American Fisheries Society* 119: 337–352.
- Townsend, C.R., Doledec, S., Norris, R., Peacock, K. & Arbuckle, C. 2003. The influence of scale and geography on relationships between stream community composition and landscape variables: description and prediction. *Freshwater Biology* 48: 768–785.
- USGS 1994. USGS land use and land cover (LULC) data. Sioux Falls, SD: US Geological Survey.
- USGS 1997. National hydrography dataset. Reston, VA: US Geological Survey.
- Vander Zanden, M.J., Olden, J.D., Thorne, J.H. & Mandrak, N.E. 2004. Predicting the occurrence and impact of bass introductions on temperate lake food webs. *Ecological Applications* 14: 132–148.
- Wall, S.S., Berry, C.R., Blausey, C.M., Jenks, J.A. & Kopplin, C.J. 2004. Fish-habitat modeling for gap analysis to conserve the endangered Topeka shiner (*Notropis topeka*). *Canadian Journal of Fisheries and Aquatic Sciences* 61: 954–973.
- Wang, L., Lyons, J., Kanehl, P. & Gatti, R. 1997. Influences of watershed land use on habitat quality and biotic integrity in Wisconsin streams. *Fisheries* 22: 6–12.
- Williams, P.H. & Araujo, M.B. 2000. Using probability of persistence to identify important areas for biodiversity conservation. *Proceedings of the Royal Society of London* 267: 1959–1966.
- Winston, M.R., Taylor, C.M. & Pigg, J. 1991. Upstream extirpation of four minnow species due to damming of a prairie stream. *Transactions of the American Fisheries Society* 120: 98–105.
- Wischmeier, W.H. 1976. Use and misuse of the universal soil loss equation. *Journal of Soil and Water Conservation* 31: 5–9.
- Zorn, T.G., Seelbach, P.W. & Wiley, M.J. 2002. Distributions of stream fishes and their relationship to stream size and hydrology in Michigan's lower peninsula. *Transactions of the American Fisheries Society* 131: 70–85.