

Comparison between Model-Predicted and Field-Measured Stream Habitat Features for Evaluating Fish Assemblage–Habitat Relationships

TRAVIS O. BRENDEN,*¹ LIZHU WANG, RICHARD D. CLARK JR., AND PAUL W. SEELBACH

*Institute for Fisheries Research, University of Michigan and Michigan Department of Natural Resources,
212 Museums Annex, Ann Arbor, Michigan 48109-1084, USA*

JOHN LYONS

Wisconsin Department of Natural Resources, 1350 Femrite Drive, Monona, Wisconsin 53716-3736, USA

Abstract.—The use of model-predicted, local-scale habitat data as inputs in analyses intended to evaluate multiscale fish assemblage–habitat relationships in streams has become increasingly common as the scale at which such studies are conducted has increased. We used fish assemblage and habitat data from 208 Wadeable streams in Wisconsin and Michigan to determine whether model-predicted habitat data would yield results similar to those of field-measured data in multiscale analyses of fish assemblage–habitat relationships. Predictions of local habitat features from landscape variables were generated by means of generalized additive modeling with likelihood-based boosting. Relationships between fish assemblage measures and landscape and local habitat features were studied via partial constrained multivariate ordination analyses. The total variation explained in the fish assemblage data sets was similar for model-predicted local habitat features and field-measured data, as was the proportion of variation explained that was due independently to local and regional (i.e., landscape) habitat features. We observed dissimilar results in the magnitude of ordination scores for local habitat features and the directional relationships between local habitat ordination scores and individual species and assemblage metric scores. Our findings indicate that model-predicted, local-scale habitat data can be useful for evaluating the relative strengths of local and regional habitat features in structuring fish assemblages, but caution may be necessary when evaluating species–habitat or assemblage metric–habitat relationships.

For fish and aquatic macroinvertebrates to colonize and persist within streams, species must possess traits that allow individuals to pass through environmental “filters” that operate from continental to microhabitat spatial scales (Tonn 1990; Poff 1997). As a result, studies relating fish and macroinvertebrate assemblages to habitat features often find that both regional-scale

(i.e., climate and landscape) and local-scale (i.e., instream) measures explain significant variation in assemblage structure (Richards et al. 1996; Allan et al. 1997; Zorn et al. 2004; Weigel et al. 2006). Although increased availability of regional-scale geographic information systems (GIS) data sets has facilitated the use of large-scale data for studying fish–habitat relationships, the ability to cost-efficiently measure local habitat features across large areas remains a challenge. Technological advances (e.g., infrared videography, high spatial resolution hyperspectral imagery) have facilitated remote monitoring of local habitat conditions; however, it remains to be seen whether such technologies can be cost-effectively applied across entire ecoregions or states, which is the extent at which many fish–habitat research studies are commonly conducted (Angermeier and Winston 1999; Baker et al. 2005; Creque et al. 2005; Heitke et al. 2006; Steen et al. 2006).

One promising approach for inventorying local-scale habitat features across large areas is predictive modeling based on landscape data (Wang et al. 2006a). Such an approach for stream habitat inventory is premised on landscapes constraining local-scale habitat conditions (Frissell et al. 1986). Empirical studies indeed have shown that substantial variation in local-scale habitats can be explained through regional landscape features (Richards et al. 1996; Zorn and Wiley 2006), suggesting that this type of modeling may be beneficial for inventorying local-scale habitats. Local-scale habitat features that frequently are predicted from regional landscape and climate data include stream temperature (Creque et al. 2005; Wehrly et al. 2006), physical habitat (Baker et al. 2005), hydrology (Hastenrath 1990; Wiley et al. 1997; Hsieh et al. 2003; Creque et al. 2005), water chemistry (Hunsaker and Levine 1995; Daly et al. 2002), and groundwater loading (Baker et al. 2001).

Although predictive modeling of local-scale habitat features is believed to be beneficial for some purposes

* Corresponding author: brenden@msu.edu

¹ Current address: Quantitative Fisheries Center, Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan 48824-1101, USA.

Received November 10, 2005; accepted December 15, 2006
Published online April 19, 2007

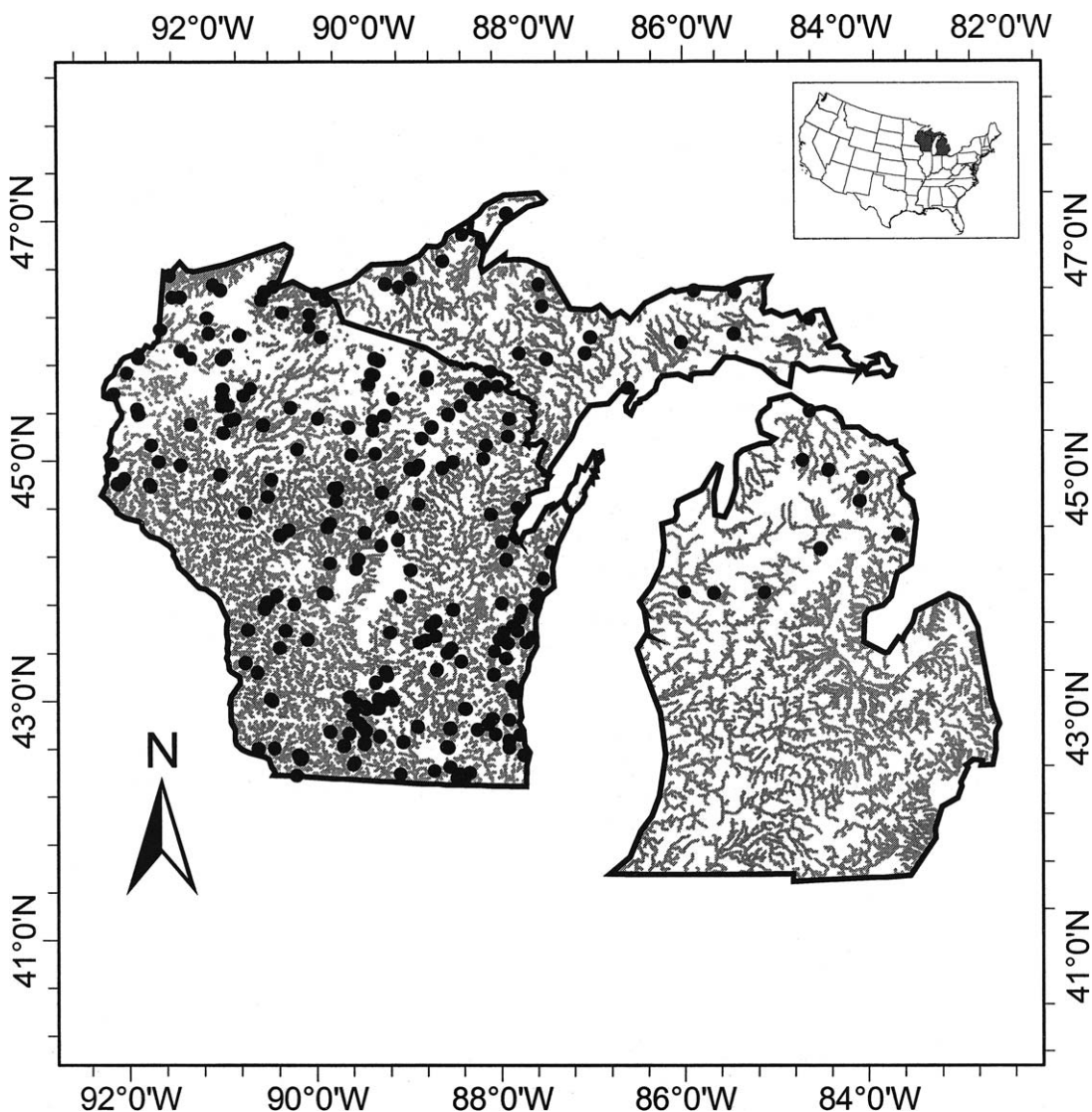


FIGURE 1.—Locations of the 208 stream sites in Wisconsin and Michigan from which data on fish and local-scale habitat features were obtained for the purpose of modeling multiscale fish assemblage–habitat relationships.

(Seelbach et al. 2002; Wang et al. 2006a), it remains unclear whether predictive modeling of local habitat features would be useful when researching fish–habitat relationships. Because models are inherent simplifications of reality, it is reasonable to expect that some explainable variability would be lost if model predictions were used as independent variables in analyses meant to evaluate fish–habitat relationships. What remains less certain, though, is whether interpretable fish–habitat relationships would remain similar for both model-predicted and field-measured habitat data. For example, stream ecologists often study whether

regional-scale or local-scale habitat features have a stronger structural influence on fish assemblages (Wang et al. 2003, 2006b; Weigel et al. 2006). If local-scale habitat data are predicted from regional landscape features, it is possible that the fraction of total variation explained in the fish assemblages due independently to local habitat may decline, leading researchers to perhaps falsely conclude that landscape features have a stronger influence on fish assemblages. The purpose of our research was to address this question by comparing results when both model-predicted and field-measured local habitat data were

TABLE 1.—Mean and coefficient of variation (CV = 100 × SD/mean) for the fish assemblage variables at 208 stream sampling sites in Michigan and Wisconsin.

Variable	Description	Mean	CV
FISHNB	Fish abundance (individuals/100 m)	133.86	112.74
PCCARNIV	Top carnivore individuals (%)	12.88	156.80
PCCLCDSP	Cool- and coldwater species (%)	23.93	96.43
PCCOLCD	Cool- and coldwater individuals (%)	23.98	128.12
PCINSECT	Insectivore individuals (%)	51.67	47.48
PCINTONB	Intolerant individuals (%)	17.05	126.98
PCLITHO	Lithophilous individuals (%)	34.15	67.53
PCOMNIV	Omnivore individuals (%)	14.80	110.51
PCSALMNB	Salmonid individuals (%)	9.49	211.89
PCSALMSP	Salmonid species (%)	7.27	177.58
PCTOLER	Tolerant individuals (%)	39.05	70.12
SPCNT	Number of fish species	12.51	48.11

used to conduct multivariate, multiscale analyses of fish assemblage–habitat relationships for Michigan and Wisconsin streams.

Methods

Data collection.—We used data from 208 wadeable stream sites located in Wisconsin and the Upper and northern Lower peninsulas of Michigan (Figure 1). Fish assemblages at these sites were sampled during the summers of 1997 to 2003. Fish sampling was conducted once at each site using either two backpack electrofishing units operated in tandem or a single tow-barge electrofishing unit. At each site we sampled stream distances approximately 35 times the mean wetted stream width. Fish assemblages at the sites were described according to species abundance (number of individuals per 100 m of sampled stream); species presence or absence; and 12 assemblage metrics describing thermal, feeding, tolerance, and reproduction classifications presented in Lyons et al. (1996; Table 1). Wang et al. (2003) provides additional details regarding sampling and description of fish assemblages at the sites.

Local-scale habitat conditions (hereafter referred to as local habitat) at the stream sites were measured on the same day that fish sampling occurred. Conductivity; dissolved oxygen; discharge; lengths of riffles, pools, and runs; and distances between bends and between riffles were measured at the sites. Channel morphology (e.g., bank-full width, mean depth at thalweg), bottom substrate (percent stream bottom covered with detritus, cobble, gravel, sand, or silt), log availability (density of logs >1.5 m in length and >0.1 m in diameter), fish cover (percent undercut bank, woody debris, or overhanging vegetation), bank stability (percent stable bank), and riparian condition (percent of bank within 5 m of stream that is developed) were measured along 12 transects estab-

lished at the stream sites (Simonson et al. 1994). Water temperature during the month of July for the year in which fish sampling was conducted was recorded via temperature data loggers deployed at the stream sites.

Regional-scale climate and landscape features (hereafter referred to as regional habitat) for the sampled stream sites were summarized at four levels (reach catchment, network catchment, reach buffer, and network buffer) via a standardized GIS process (Brenden et al. 2006). Reach catchments included all upstream areas that drained directly to the stream reaches; network catchments included all upstream areas that drained to the stream reaches by either overland or waterway routes (Figure 2). Similarly, reach buffers included only those areas bounded within a certain distance of the stream reaches, and network buffers included all reach buffers located upstream from the individual stream reaches. The distance used to delineate riparian buffers was 60 m on either side of the stream channels (Brenden et al. 2006). Summarized regional habitat variables included percentages of surficial geology, land cover, and bedrock geology classifications within the delineated areas. Other summarized regional habitat variables included mean July air temperature, number of growing degree-days (10°C baseline temperature), mean precipitation, surface area, slope, soil permeability, and potential groundwater delivery. Brenden et al. (2006) provides additional details regarding regional habitat data sources, delineation of regional habitat levels, and methods used for habitat feature attribution.

Variable selection.—We attributed 117 regional and 60 local habitat features to each stream site. To reduce the large number of candidate variables to a smaller set (Nelson 2001), we used PROC VARCLUS in SAS (SAS Institute 2004), which is conceptually similar to principal components analysis and implements an iterative procedure for clustering variables. We thereby clustered the regional and local data into 20 clusters each that explained approximately 70% of the total observed variation in each data set. One variable from each of the clusters was then selected to use for analyzing fish assemblage–habitat relationships (Nelson 2001). Selected variables generally had high correlations with their own cluster components and low correlations with other clusters or were selected based on a hypothesized relationship with at least one of the fish assemblage data sets.

After selecting the representative variables from the habitat clusters, we used forward variable selection in CANOCO (ter Braak and Smilauer 1998) to identify those variables with significant conditional effects on the data sets for species abundance, species presence or absence, or assemblage metrics. Forward selection of

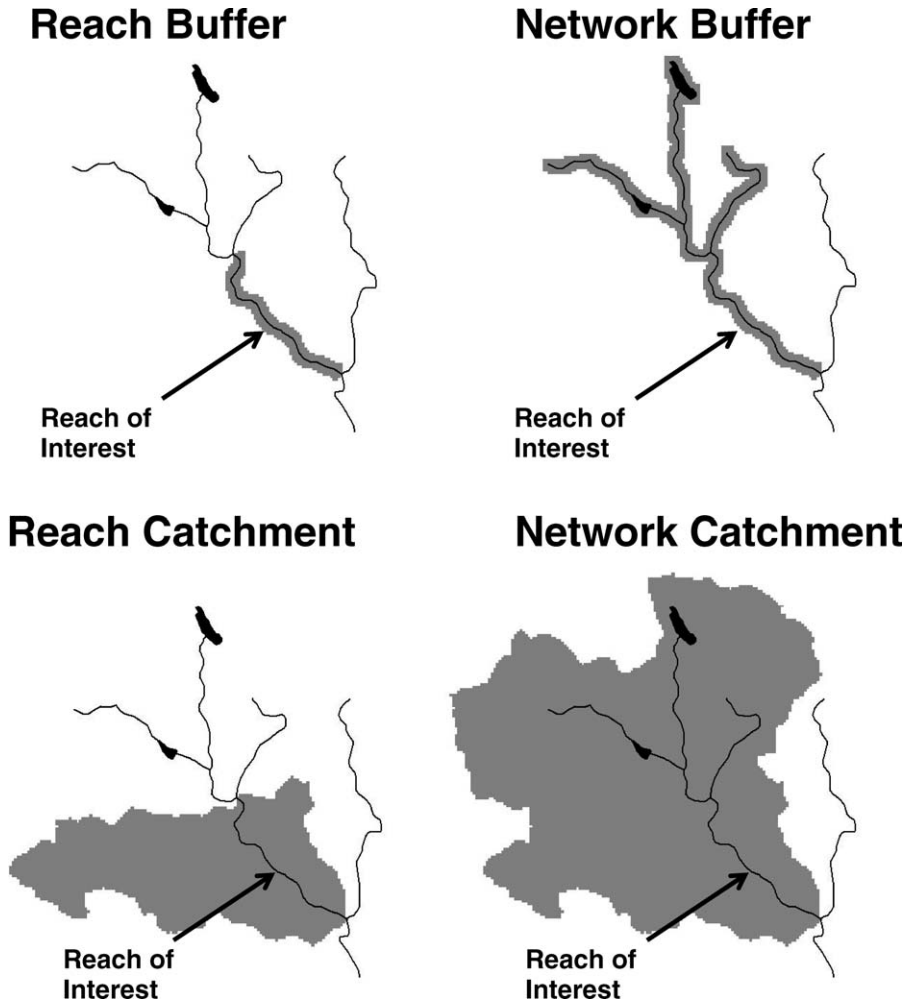


FIGURE 2.—Example of the differences, for a particular reach of interest, between the riparian buffers and catchments for a reach (two left panels) and those for a network (two right panels). The areas encompassed by each of the levels are outlined in gray. In this example, the direction of water flow is generally from the northwest (upper left) to the southeast (lower right). Figure modified from Brendan et al. (2006).

the habitat variables was conducted separately for regional and local habitats. Based upon gradient lengths determined through detrended correspondence analysis, we used canonical correspondence analysis (CCA) for the forward variable selection for the species abundance and species presence–absence data sets and redundancy analysis (RDA) for the forward variable selection for the fish assemblage data set (ter Braak and Prentice 1998). Monte Carlo permutations (number of permutations = 199) were used to determine the statistical significance of the conditional effects of the habitat variables. Only those variables with statistically significant ($\alpha = 0.05$) conditional effects were used in the multiscale analyses of fish–habitat relationships.

Model prediction of local habitat features.—We developed models to predict those local habitat features that were found to have statistically significant conditional effects on at least one of the fish assemblage data sets. Because of the possibility that regional variables could have nonlinear effects on local habitat variables and because of the large number of regional habitat variables that were available for model fitting, we chose to use generalized additive modeling with likelihood-based boosting to fit the prediction models (Tutz and Binder 2006). Generalized additive modeling is a semiparametric regression approach for generating nonlinear response curves between dependent and independent variables, and because users do not need to specify a particular model

equation it can be a useful modeling procedure in cases where there is little available information indicating how independent variables may influence dependent variables (Venables and Dichmont 2004). Likelihood-based boosting is a method for selecting variables and the degree of variable smoothing when fitting generalized additive models; it is efficient with large-dimension data sets (Tutz and Binder 2006). Prediction models for the local habitat features were fit in R (R Development Core Team 2006) using the GAMBoost package (Tutz and Binder 2006), which uses penalized B-splines to model nonlinear relationships between variables. All of the 117 regional habitat variables that were attributed to the stream sites were used as candidate variables for fitting the prediction models for the local habitat features. When fitting models to percentage data, a binomial distribution was specified as the family distribution; for all other local habitat features a Gaussian distribution was specified as the family distribution. The number of boosting steps used in fitting the generalized additive models was determined by AIC minimization, whereas the penalty parameter for the likelihood-based boosting was selected using a coarse line search such that the number of required boosting steps was at least 50 (Tutz and Binder 2006).

Multiscale analyses of fish–habitat relationships.—Partial constrained ordination, which is a commonly used multivariate statistical technique for exploring the relative importance of multiscale habitat components in aquatic environments (Wang et al. 2003, 2006b; Sandin and Johnson 2004; Weigel et al. 2006), was used to partition the variation in the data for species abundance, species presence–absence, and assemblage metrics between regional and local habitat components. Canonical correspondence analysis was used to partition variation for the species abundance and species presence–absence data sets, and RDA was used to partition variation for the assemblage metrics data set. Partial constrained ordinations were conducted in CANOCO (ter Braak and Smilauer 1998) using methods described in Borcard et al. (1992). Ordinations were conducted using both observed and model-predicted local habitat data sets. Ordinations were also conducted using a joint local habitat data set, which consisted of a combination of model-predicted and field-measured local habitat variables. For the joint data set, we included as predicted data only those local habitat variables that, when compared with observed data, had Pearson's product-moment correlation coefficients greater than 80%. For each ordination that was conducted, Monte Carlo permutation tests (number of permutations = 199) were used to determine the

statistical significance of the observed relationship between the fish data set and the local habitat features.

Comparisons in the partial constrained ordination results among the observed, model-predicted, and joint local habitat data sets were based on differences in the total variation explained and the fraction of explained variation attributable independently to local habitat features for the data sets for species abundance, species presence–absence, and assemblage metrics. Although total variation explained is generally not comparable across different data sets, the correspondence among the observed, model-predicted, and joint data sets should result in this being a valid comparison (Økland 1999). To determine whether the relationships were similar across the different local habitat data sets, we also examined the magnitude of ordination scores for the local habitat features and the agreement in the directional relationships by comparing the ordination scores for the local habitat features and the individual species with assemblage metrics for the two ordination axes that explained the largest proportions of observed variance in the fish assemblage–habitat relationships.

Results

We collected 54 species from the stream sites. Average fish abundance ranged from a minimum of 0.1 fish/100 m for walleye *Sander vitreus* to a maximum of 17.7 fish/100 m for common shiner *Luxilus cornutus*. Of the 20 regional and locale habitat variables that were initially selected through the variable clustering, 17 regional and 12 local habitat variables had statistically significant conditional effects for at least one of the fish assemblage data sets (Tables 2, 3).

Prediction accuracy varied widely for the fitted generalized additive models for the 12 local habitat variables with statistically significant conditional effects for at least one of the fish data sets (Figure 3). Pearson's product-moment correlation coefficients of model predictions in relation to observed data ranged from 37% to 93% (Table 3). Model predictions for bankfull width, conductivity, maximum daily mean water temperature, mean thalweg water depth, and stream gradient had Pearson correlations greater than 80% compared with observed data. Thus, model predictions for these five variables were included along with the observed data for the other local habitat variables when conducting ordinations with the joint local habitat data set.

The total variation explained in the species abundance data set decreased from 30% when observed local habitat data were used in the partial constrained ordinations to 28% when model-predicted data were used. Total variation explained was 29% when joint

TABLE 2.—Mean and coefficient of variation (CV) for regional-scale stream habitat variables having significant ($P < 0.05$) conditional effects on fish assemblage measures, as determined by forward selection in canonical correlation and redundancy analyses. Significant conditional effects for variables include effects for the species abundance data set (AB), effects for the species presence–absence dataset (PA), and effects for the fish assemblage metrics data set (AM).

Variable	Effect	Description	Mean	CV
AGRICU	AB, PA	Agriculture land cover in network catchment (%)	30.29	97.18
AREA	AB, PA, AM	Network catchment area (km ²)	186.82	154.42
BARRENWL	AM	Barren land cover in reach catchment (%)	1.77	226.17
BCARBONB	AB, PA	Carbonate bedrock geology in network buffer (%)	24.02	162.68
BDEP051B	AB, PA	Bedrock depth 15.48–30.48 m in network buffer (%)	25.35	118.39
BDEPL1HB	AB, PA, AM	Bedrock depth <30.48 m in network buffer (%)	64.77	62.11
BDEPMENW	AB, PA	Mean bedrock depth in network catchment (m)	33.77	33.02
BSHALEB	PA	Shale bedrock geology in network buffer (%)	13.60	194.11
DARCYW	AB, PA, AM	Groundwater velocity in network catchment (m/d)	9.95	95.89
GCOARSEB	AB, PA	Coarse-texture surficial geology in network buffer (%)	65.73	63.14
GFINEW	AB, PA, AM	Fine-texture surficial geology in network catchment (%)	7.91	296.41
GMEDIMB	AB	Medium-texture surficial geology in network buffer (%)	6.97	314.59
GRASS	AB, PA	Grassland land cover in network catchment (%)	9.99	83.71
NFWETLND	AB, PA, AM	Nonforested wetland land cover in network catchment (%)	2.93	138.43
SLOPE	PA, AM	Mean network catchment slope	5.44	54.78
SOILPERB	AB, PA, AM	Mean soil permeability in network buffer (cm/100 h)	21.11	707.34
URBANWL	AB, PA, AM	Urban land cover in reach catchment (%)	5.13	285.88

local habitat data were used in the ordination analysis with the species abundance data set. The fraction of variation explained independently by local habitat declined from 36% for observed data to 30% for model-predicted data. The fraction of variation explained independently by local habitat was 33% for the joint data set. Declines in the fraction of variation explained independently by local habitats when model-predicted data were used in the ordinations largely were offset by increases in the fraction of variation explained jointly by regional and local habitats (Figure 4). Statistically significant relationships between the species abundance and local habitat features were detected for the observed ($F = 2.78$; $P < 0.005$), model-predicted ($F = 2.11$; $P < 0.005$), and joint ($F = 2.38$; $P < 0.005$) local habitat data sets.

For the species presence–absence data set, the total variation explained decreased from 30% when observed local habitat data were used in the partial constrained ordinations to 28% when model-predicted data were used. Total variation explained in the species presence–absence data was 29% when joint local habitat data were used in the ordination analysis. The fraction of variation explained independently by local habitat features declined from 30% for observed data to 25% for model-predicted data. The fraction of variation explained independently by local habitat was 28% for the joint local habitat data. As was observed with the species abundance data set, declines in the fraction of variation explained independently by local habitat features with model-predicted data were largely offset by increases in the amount of variation explained

TABLE 3.—Mean and coefficient of variation (CV) for local-scale stream habitat variables having significant ($P < 0.05$) conditional effects on the different fish data sets, as determined by forward selection in canonical correlation and redundancy analyses. Also shown are Pearson's product-moment correlation coefficients (r) calculated between field-measured observations and predictions resulting from fitting generalized additive models by likelihood-based boosting. Significant conditional effects for variables include effects for the species abundance data set (AB), effects for the species presence–absence data set (PA), and effects for the assemblage metrics data set (AM).

Variable	Effect	Description	Mean	CV	r
BNKFWID	AB, PA, AM	Bank-full width (m)	12.90	76.36	92.9
COND	AB, PA, AM	Conductivity ($\mu\text{S}/\text{cm}$)	387.54	83.20	91.1
DAY	AB, PA, AM	Maximum daily mean water temperature in July ($^{\circ}\text{C}$)	23.40	13.41	81.4
DEPTH	AB, PA	Mean water depth at thalweg (m)	0.53	46.62	80.1
DETRTS	AB, PA, AM	Stream bottom covered with detritus (%)	4.57	187.58	37.4
DSTRB	AB, PA	Bank with agriculture or urban land cover (%)	7.93	233.17	55.2
EMB	AB, PA, AM	Rocky substrate covered by silt or sand (%)	57.32	55.65	70.1
GRAD	AM	Stream gradient (m/1,000 m)	2.94	139.05	88.6
POOL	AB, PA	Stream reach that is pool (%)	10.75	164.12	58.3
SBMAC	AB	Stream bottom covered with macrophytes (%)	17.36	126.77	40.8
SINUOUS	AM	Sinuosity of stream reach	1.38	25.52	65.1
WETLAND	AB	Wetland land cover within 5 m of stream bank (%)	8.31	274.48	61.1

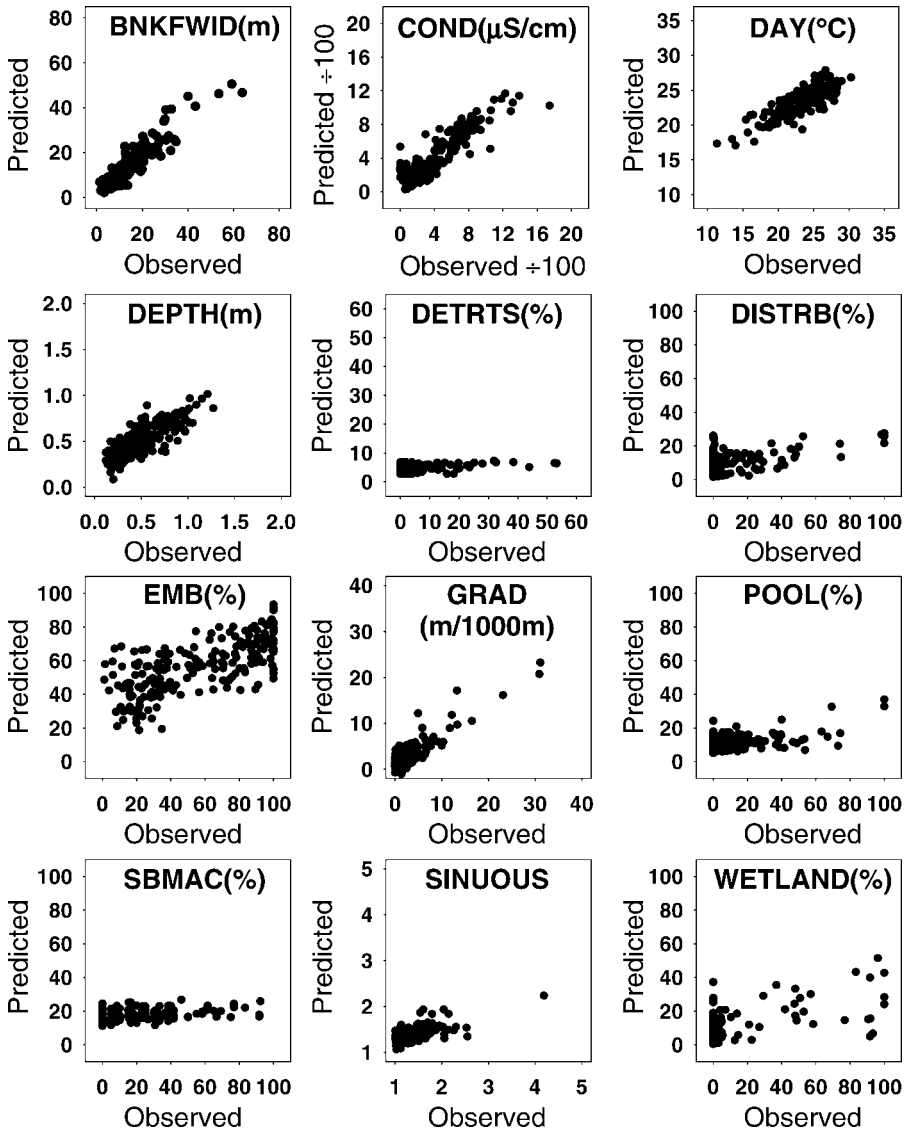


FIGURE 3.—Scatterplots of model predictions and field-measured observations for the local-scale stream variables used in multivariate analyses of fish assemblage–habitat relationships (see Table 3 for variable descriptions). Predictions were generated through generalized additive modeling in which variable selection and the degree of smoothing were determined by likelihood-based boosting (Tutz and Binder 2006).

jointly by regional and local habitats (Figure 4). Statistically significant relationships between the species presence–absence and local habitat features were again detected for the observed ($F = 2.90$; $P < 0.005$), model-predicted ($F = 2.30$; $P < 0.005$), and joint ($F = 2.61$; $P < 0.005$) local habitat data sets.

The total variation explained in the fish assemblage data set decreased from 29% for the observed local habitat data set to 26% for the model-predicted data set. Total variation explained in the fish assemblage data

was 27% when the joint local habitat data were used in the ordination analysis. The fraction of variation explained due independently to local habitat features was 46% for the observed and 39% for the model-predicted data. For the joint local habitat data, the fraction of variation explained independently by local habitat features was 41%. As was found for the species abundance and species presence–absence data, statistically significant relationships between the assemblage metrics and local habitat features were detected for the

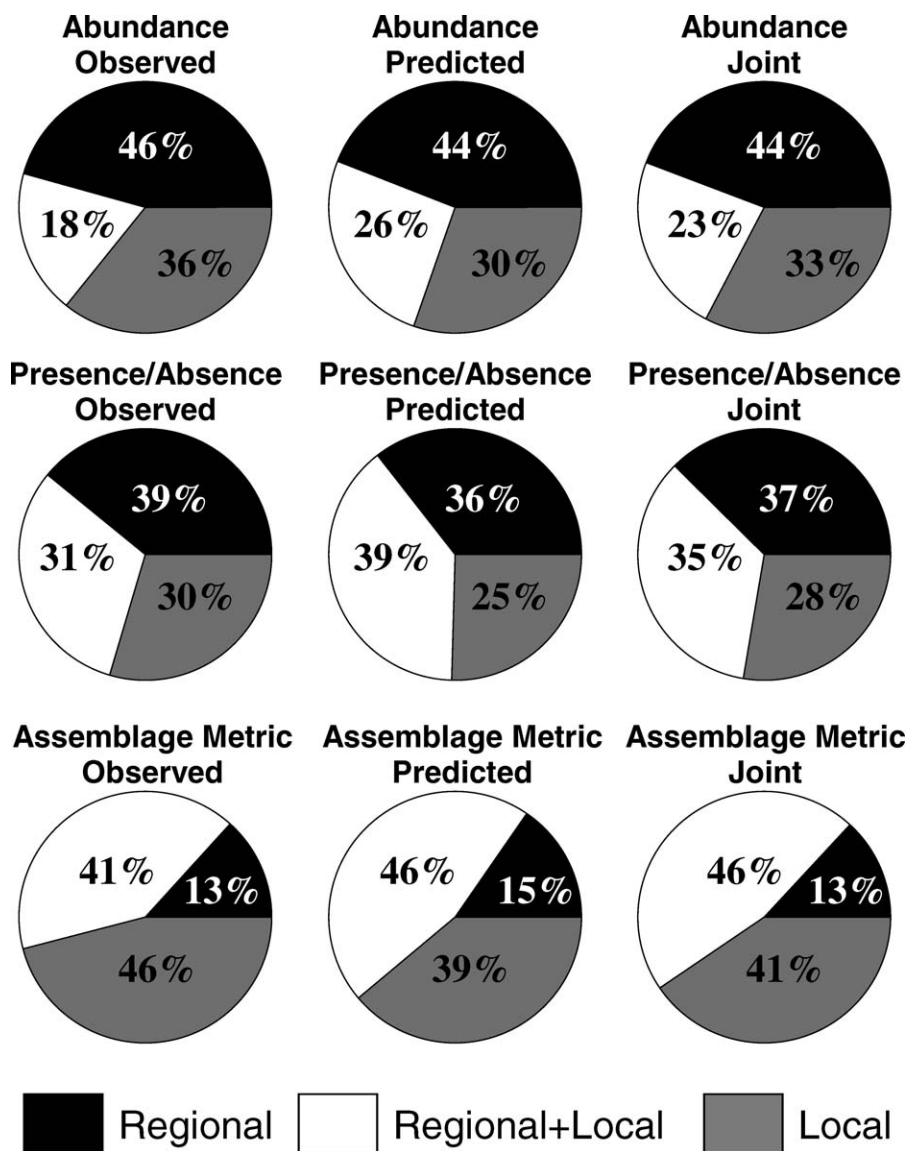


FIGURE 4.—Fractions of the explained variation in species abundance that were attributable to regional, local, and joint (regional plus local) habitat features for observed, model-predicted, and joint local habitat data sets, as determined by partially constrained ordinations. For the species abundance and presence–absence data sets, canonical correspondence analysis was used to partition the variation; for the assemblage metric data sets, redundancy analysis was used. The actual proportions explained by the data sets are indicated on each graph.

observed ($F = 5.19$; $P < 0.005$), model-predicted ($F = 3.99$; $P < 0.005$), and joint ($F = 3.72$; $P < 0.005$) local habitat data sets.

Most of the variables that had large scores for the first ordination axis for the observed data set also had large scores for the model-predicted data set. Bank-full width and maximum daily mean water temperature were among the variables with the largest ordination

scores for the fish abundance and presence–absence data sets, whereas conductivity, maximum daily mean water temperature, and percent embeddedness had the largest ordination scores for the observed and model-predicted local habitat data sets for the assemblage data set (Table 4). There also was general agreement in the magnitude of the local habitat feature scores for the second ordination axis for the observed and model-

TABLE 4.—Habitat variable scores for the first two ordination axes for the fish abundance, species presence–absence, and fish assemblage metrics data. Scores in bold italics are for the variables with the three highest scores for that ordination axis. Asterisks indicate that model predictions for a local-scale variable were used in the joint (combination of observed and model-predicted data) local data set analysis. See Table 3 for descriptions of variables.

Local-scale variable	Ordination axis 1			Ordination axis 2		
	Observed	Predicted	Joint	Observed	Predicted	Joint
Species abundance						
BNKFWID*	<i>-0.159</i>	<i>-0.131</i>	-0.112	-0.094	-0.017	-0.056
COND*	0.048	0.011	0.054	-0.009	0.038	-0.021
DAY*	<i>-0.534</i>	<i>-0.406</i>	-0.263	-0.065	-0.070	<i>-0.301</i>
DEPTH*	-0.071	<i>-0.094</i>	-0.047	0.166	0.057	-0.051
DETRTS	-0.039	0.012	<i>0.377</i>	<i>0.460</i>	<i>0.178</i>	<i>-0.246</i>
DSTRB	-0.130	-0.084	-0.113	-0.054	0.038	-0.099
EMB	0.046	0.050	<i>0.404</i>	<i>0.448</i>	<i>0.203</i>	-0.165
POOL	0.022	0.054	0.017	-0.002	<i>-0.153</i>	0.026
SBMAC	<i>0.197</i>	0.086	0.225	0.152	0.030	0.125
WETLAND	-0.087	0.045	<i>0.276</i>	<i>0.368</i>	0.110	<i>-0.235</i>
Species presence–absence						
BNKFWID*	<i>-0.136</i>	<i>-0.132</i>	-0.130	-0.101	0.057	-0.029
COND*	-0.031	-0.048	-0.015	-0.001	0.010	-0.079
DAY*	<i>-0.467</i>	<i>-0.305</i>	<i>-0.207</i>	0.085	<i>-0.161</i>	<i>-0.271</i>
DEPTH*	-0.071	<i>-0.161</i>	-0.154	0.017	0.009	-0.039
DETRTS	<i>0.099</i>	0.058	<i>0.297</i>	<i>0.335</i>	<i>-0.108</i>	<i>-0.137</i>
DSTRB	-0.014	-0.027	-0.110	<i>-0.145</i>	0.019	0.100
EMB	0.097	0.106	<i>0.273</i>	<i>0.298</i>	<i>-0.075</i>	<i>-0.124</i>
POOL	0.069	0.068	0.130	0.110	-0.018	-0.009
Fish assemblage metrics						
BNKFWID*	-0.016	-0.008	0.009	-0.072	-0.047	-0.046
COND*	<i>0.277</i>	<i>-0.213</i>	<i>0.217</i>	<i>-0.126</i>	<i>-0.114</i>	<i>-0.089</i>
DAY*	<i>-0.170</i>	<i>0.086</i>	-0.067	<i>-0.344</i>	<i>-0.284</i>	<i>-0.293</i>
DETRTS	0.038	-0.085	0.042	0.001	-0.048	0.008
EMB	<i>0.258</i>	<i>-0.115</i>	<i>0.269</i>	0.011	0.066	0.048
GRAD*	0.016	0.070	-0.077	<i>0.155</i>	<i>0.155</i>	<i>0.146</i>
SINUOUS	0.094	-0.081	<i>0.098</i>	0.002	-0.040	0.015

predicted data sets. Percent of stream bottom covered with detritus and percent embeddedness were among the variables with the largest ordination scores for both the species abundance and species presence–absence data. Conductivity, maximum daily mean water temperature, and stream gradient had the largest ordination scores with fish assemblages in the second ordination axis for both the observed and model-predicted data.

The magnitude of variable scores for ordination between the observed and joint habitat data sets did not agree as well as for the observed and model-predicted data sets. For the species abundance data, none of the local habitat features with the largest ordination scores for the observed data were among the variables with the largest ordination scores for the joint data for the first ordination axis (Table 4). Percent of stream bottom covered with detritus and percent wetland land type had large ordination scores for both observed and joint data sets for the second ordination axis when ordinations were conducted on the species abundance data. For the species presence–absence data set, maximum daily mean water temperature and percent of stream

bottom covered with detritus for the first ordination axis and percent embeddedness and percent of stream bottom covered with detritus for the second axis were among the variables with the largest ordination scores for both the observed and joint data (Table 4). For the assemblage metrics data, conductivity, maximum daily mean water temperature, percent embeddedness, and stream gradient had large ordination scores for either the first or second ordination axis for both the observed and joint local habitat data (Table 4).

For the species abundance data set, directional agreement between local habitat ordination scores and individual species scores for the first ordination axis equaled 71% for the observed and model-predicted local habitat data. For the observed and joint data, directional agreement between local habitat ordination scores and individual species scores was 53% for the first ordination axis. For the second ordination axis, directional agreement between local habitat ordination scores and individual species scores was 59% for the observed and model-predicted data and was 50% for the observed and joint data.

For the species presence–absence data set, direc-

tional agreement between local habitat ordination scores and individual species scores for the first ordination axis equaled 86% for the observed and model-predicted local habitat data. Directional agreement between local habitat ordination scores and individual species scores for the observed and joint data was 83% for the first ordination axis. For the second ordination axis, directional agreement between local habitat ordination scores and individual species scores was 61% for the observed and model-predicted data and was 59% for observed and joint data.

For the fish assemblage data set, directional agreement between local habitat ordination scores and assemblage metric scores for the first ordination axis was 68% for the observed and model-predicted local habitat data. A 68% agreement in directional relationships between local habitat ordination scores and fish assemblage scores for the first ordination axis also was found for the observed and joint data sets. For the second ordination axis, directional agreement between local habitat ordination scores and assemblage scores was 71% for the observed and model-predicted data and was 100% for the observed and joint local habitat data.

Discussion

Predictive modeling of local-scale habitat features is an appealing method for stream habitat inventory because of its potential cost savings relative to traditional field sampling (Wang et al. 2006a). In practice, however, accuracy of models developed for predicting local-scale habitat features can be somewhat questionable. Coefficients of determination for models predicting instream habitat features (e.g., water chemistry, physical habitat, and stream temperature) often are less than 60% (Baker et al. 2005; Wehrly et al. 2006). Additionally, prediction models are often not properly validated (Olden et al. 2002), nor are model assumptions (e.g., normality, linearity) properly checked. Such factors raise the question as to whether results obtained with model-predicted local habitat data would be similar to those that might otherwise have been obtained with field-measured data.

Our study found that the model-predicted local habitat data did, to a certain extent, provide results that were similar to those obtained with field-measured data when analyzing multiscale fish assemblage-habitat relationships in Michigan and Wisconsin streams. In particular, we found that model-predicted local habitat data resulted in only modest declines (2–3%) in the total variation explained in three fish assemblage data sets relative to that of field-measured data. Declines in total variation explained were even less when local habitat data consisted of a combination of observed and

model-predicted variables. We also found that the relative strengths of regional and local habitat features in structuring fish assemblages remained similar, regardless of whether observed or model-predicted data were used in our ordinations. For example, observed, model-predicted, and joint local habitat data sets all indicated that regional features accounted for a larger fraction of total variation explained in the species presence-absence data set. Further, all three data sets indicated that shared and local-scale habitat features accounted for the largest fractions of total explained variation for the fish assemblage data set. In general, we found that declines in the fraction of variation explained independently by local habitat features were offset by increase in the fraction of variation explained jointly by regional and local habitat features. Given that our model predictions for local habitat variables were generated using regional data, this result was somewhat expected.

The major issue that we found with using model-predicted data concerned the identification of environmental gradients for local habitats and the identifiable relationships between local habitat features and individual species or assemblage metrics. We found that environmental gradients that were identified with field-measured data were not necessarily the same as those that were identified with model-predicted data. This was the case even when only the most accurate model predictions were incorporated in the local habitat data set. Additionally, we found that directional relationships between individual species and fish assemblage ordination scores and local habitat scores oftentimes switched directions when model-predicted data were used in the ordinations; this could lead to incorrect conclusions as to how a species or group of species is affected by a particular habitat feature. As a result, any conclusions that we might have tried to draw regarding how individual species or fish assemblage metrics were affected by any of the local-scale habitat features for Michigan and Wisconsin streams would have been tentative at best.

Our intent for this research was not to explicitly study fish-habitat relationships, nor was it to construct the best possible model for predicting local habitat components. We thus intentionally have avoided discussing results related to either of these research aspects; rather, we have focused exclusively on the similarity of results obtained through field-measured and model-predicted data. Regarding the former research aspect, Wang et al. (2003, 2006b) previously used some of the data in the present study to identify the relative influences of catchment, reach, and riparian buffer habitat features on stream fish assemblages (these publications discuss how fish in Michigan and

Wisconsin streams are structured by multiscale habitat features). Regarding the latter research aspect, we encourage study of the potential benefits of generalized additive modeling through likelihood based boosting as a technique for fitting prediction models. In particular, its ability to deal efficiently with large data sets (e.g., hundreds of explanatory variables) makes it a conceptually appealing model-fitting technique. Research comparing this modeling approach to other commonly used model approaches potentially could be beneficial to the entire fisheries science and management field.

In conclusion, the use of regional data to predict local habitat conditions in streams has proliferated recently largely because of increased use of GIS in stream research and management (Fisher and Rahel 2004). Predictive modeling of local habitat features through regional landscape and climate conditions has been advocated as a useful approach for hindcasting predevelopment habitat conditions and for forecasting how changes in regional land use may affect instream habitat (Seelbach et al. 2002; Kilgour and Stanfield 2006; Wang et al. 2006a). Additionally, it is becoming increasingly common to use model-predicted habitat data as inputs for other modeling purposes (Baker et al. 2005; Creque et al. 2005; Wall et al. 2004; McKenna et al. 2006; Zorn and Wiley 2006). Based on the results from the present study, we would encourage researchers to exercise caution when using model-predicted data to study fish assemblage-habitat relationships. Although we found that model-predicted data in some cases can yield results that are similar to those obtained with field-measured data, in other cases the results can be quite different from those obtained with field-measured data. Such differences can occur even if local habitat features are predicted with apparently high accuracy. If model-predicted local habitat data are included in analyses meant to elucidate fish assemblage-habitat relationships, then the results from our research on Michigan and Wisconsin streams suggest that conclusions regarding observed environmental gradients or directional relationships between local habitat features and individual species or assemblage metrics should only be considered as tentative.

Acknowledgments

Funding for this research was provided through U.S. Environmental Protection Agency STAR Grant R-83059601. The authors thank Steve Aichele, Ed Bissell, Arthur Cooper, and Jana Stewart for assistance with the GIS attribution of habitat features for the sampling sites. Kirk Krueger, Tammy Newcomb, and several anonymous reviewers provided helpful critiques on earlier drafts of this manuscript.

References

- Allan, J. D., D. L. Erickson, and J. Fay. 1997. The influence of catchment land use on stream integrity across multiple spatial scales. *Freshwater Biology* 37:149–161.
- Angermeier, P. L., and M. R. Winston. 1999. Characterizing fish community diversity across Virginia landscapes. *Ecological Applications* 9:335–349.
- Baker, E. A., K. E. Wehrly, P. W. Seelbach, L. Wang, M. J. Wiley, and T. Simon. 2005. A multimetric assessment of stream condition in the Northern Lakes and Forests Ecoregion using spatially explicit statistical modeling and regional normalization. *Transactions of the American Fisheries Society* 134:697–710.
- Baker, M. E., M. J. Wiley, and P. W. Seelbach. 2001. GIS-based hydrologic modeling of riparian areas: implications for stream water quality. *Journal of the American Water Resources Association* 37:1615–1628.
- Borcard, D., P. Legendre, and P. Drapeau. 1992. Partialling out the spatial component of ecological variation. *Ecology* 74:1045–1055.
- Brenden, T. O., R. D. Clark, Jr., A. R. Cooper, P. W. Seelbach, L. Wang, S. S. Aichele, E. G. Bissell, and J. S. Stewart. 2006. A GIS framework for collecting, managing, and analyzing multiscale landscape variables across large regions for river conservation and management. Pages 49–74 in R. M. Hughes, L. Wang, and P. W. Seelbach, editors. *Landscape influences on stream habitats and biological assemblages*. American Fisheries Society, Symposium 48, Bethesda, Maryland.
- Creque, S. M., E. S. Rutherford, and T. G. Zorn. 2005. Use of GIS-derived landscape-scale habitat features to explain spatial patterns of fish density in Michigan rivers. *North American Journal of Fisheries Management* 25:1411–1425.
- Daly, K., P. Mills, B. Coulter, and M. McGarrigle. 2002. Modeling phosphorus concentrations in Irish rivers using land use, soil type, and soil phosphorus data. *Journal of Environmental Quality* 31:590–599.
- Fisher, W. L., and F. J. Rahel. 2004. Geographic information systems applications in streams and river fisheries. Pages 49–84 in W. L. Fisher and F. J. Rahel, editors. *Geographic information systems in fisheries*. American Fisheries Society, Bethesda, Maryland.
- Frissell, C. A., W. J. Liss, C. E. Warren, and M. D. Hurley. 1986. A hierarchical framework for stream habitat classification: viewing streams in a watershed context. *Environmental Management* 10:199–214.
- Hastenrath, S. 1990. Diagnostics and prediction of anomalous river discharge in northern South America. *Journal of Climate* 3:1080–1096.
- Heitke, J. D., C. L. Pierce, G. T. Gelwicks, G. A. Simmons, and G. L. Siegwarth. 2006. Habitat, land use, and fish assemblage relationships in Iowa streams: preliminary assessment in an agricultural landscape. Pages 287–303 in R. M. Hughes, L. Wang, and P. W. Seelbach, editors. *Landscape influences on stream habitats and biological assemblages*. American Fisheries Society, Symposium 48, Bethesda, Maryland.
- Hsieh, W. W., Yuval, J. Li, A. Shabbar, and S. Smith. 2003. Seasonal prediction with error estimation of Columbia River streamflow in British Columbia. *Journal of Water Resources Planning and Management* 129:146–149.
- Hunsaker, C. T., and D. A. Levine. 1995. Hierarchical

- approaches to the study of water quality in rivers. *BioScience* 45:193–203.
- Kilgour, B. W., and L. W. Stanfield. 2006. Hindcasting reference conditions in streams. Pages 623–639 in R. M. Hughes, L. Wang, and P. W. Seelbach, editors. *Landscape influences on stream habitats and biological assemblages*. American Fisheries Society, Symposium 48, Bethesda, Maryland.
- Lyons, J., L. Wang, and T. D. Simonson. 1996. Development and validation of an index of biotic integrity for coldwater streams in Wisconsin. *North American Journal of Fisheries Management* 16:241–265.
- McKenna, J. E., Jr., R. P. McDonald, C. Castiglione, S. S. Morrison, K. P. Kowalski, and D. R. Passino-Reader. 2006. A broad-scale fish–habitat model development process: Genesee Basin, New York. Pages 533–554 in R. M. Hughes, L. Wang, and P. W. Seelbach, editors. *Landscape influences on stream habitats and biological assemblages*. American Fisheries Society, Symposium 48, Bethesda, Maryland.
- Nelson, B. D. 2001. Variable reduction for modeling using PROC VARCLUS. Proceedings of the Twenty-sixth Annual SAS User's Group International Conference. SAS Institute Inc., Cary, North Carolina. Available: www2.sas.com/proceedings/sugi26/p261–26.pdf. (November 2006).
- Økland, R. H. 1999. On the variation explained by ordination and constrained ordination axes. *Journal of Vegetation Science* 10:131–136.
- Olden, J. D., D. A. Jackson, and P. R. Peres-Neto. 2002. Predictive models of fish species distributions: a note on proper validation and chance prediction. *Transactions of the American Fisheries Society* 131:329–336.
- Poff, N. L. 1997. Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. *Journal of the North American Benthological Society* 16:391–409.
- R Development Core Team. 2006. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available: www.r-project.org. (November 2006).
- Richards, C., L. B. Johnson, and G. E. Host. 1996. Landscape-scale influences on stream habitats and biota. *Canadian Journal of Fisheries and Aquatic Sciences* 53(Supplement 1):295–311.
- Sandin, L., and R. K. Johnson. 2004. Local, landscape, and regional factors structuring benthic macroinvertebrate assemblages in Swedish streams. *Landscape Ecology* 19:501–514.
- SAS Institute. 2004. SAS/STAT 9.1 user's guide. SAS Institute, Cary, North Carolina.
- Seelbach, P. W., M. J. Wiley, P. A. Soranno, and M. T. Bremigan. 2002. Aquatic conservation planning: using landscape maps to predict ecological reference conditions for specific waters. Pages 454–478 in K. Gutzwiller, editor. *Conceptual applications of landscape ecology in biological conservation*. Springer-Verlag, New York.
- Simonson, T. D., J. Lyons, and P. Kanehl. 1994. Guidelines for evaluating fish habitat in Wisconsin streams. U.S. Forest Service General Technical Report NC-164. Available: http://ncrs.fs.fed.us/pubs/gtr/gtr_nc164.pdf. (November 2006).
- Steen, P. J., D. R. Passino-Reader, and M. J. Wiley. 2006. Modeling brook trout presence and absence from landscape variables using four different analytical methods. Pages 513–532 in R. M. Hughes, L. Wang, and P. W. Seelbach, editors. *Landscape influences on stream habitats and biological assemblages*. American Fisheries Society, Symposium 48, Bethesda, Maryland.
- ter Braak, C. J. F., and I. C. Prentice. 1998. A theory of gradient analysis. *Advances in Ecological Research* 18:271–317.
- ter Braak, C. J. F., and P. Smilauer. 1998. CANOCO reference manual and user's guide to CANOCO for Windows: software for canonical guild ordination, version 4. Microcomputer Power, Ithaca, New York.
- Tonn, W. M. 1990. Climate change and fish communities: a conceptual framework. *Transactions of the American Fisheries Society* 119:337–352.
- Tutz, G., and H. Binder. 2006. Generalized additive modeling with implicit variable selection by likelihood-based boosting. *Biometrics*. 62:961–971.
- Venables, W. N., and C. M. Dichmont. 2004. GLMs, GAMs, and GLMMs: an overview of theory for applications in fisheries research. *Fisheries Research* 70:319–337.
- Wall, S. S., C. R. Berry, Jr., C. M. Blausey, J. A. Jenks, and C. J. Kopplin. 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., J. Lyons, P. Rasmussen, P. Seelbach, T. Simon, M. Wiley, P. Kanehl, E. Baker, S. Niemela, and P. Stewart. 2003. Watershed, reach, and riparian influences on stream fish assemblages in the Northern Lakes and Forest Ecoregion, U.S.A. *Canadian Journal of Fisheries and Aquatic Sciences* 60:491–505.
- Wang, L., P. W. Seelbach, and R. M. Hughes. 2006a. Introduction to landscape influences on stream habitats and biological assemblages. Pages 1–23 in R. M. Hughes, L. Wang, and P. W. Seelbach, editors. *Landscape influences on stream habitats and biological assemblages*. American Fisheries Society, Symposium 48, Bethesda, Maryland.
- Wang, L., P. W. Seelbach, and J. Lyons. 2006b. Effects of levels of human disturbance on the influence of catchment, riparian, and reach-scale factors on fish assemblages. Pages 199–219 in R. M. Hughes, L. Wang, and P. W. Seelbach, editors. *Landscape influences on stream habitats and biological assemblages*. American Fisheries Society, Symposium 48, Bethesda, Maryland.
- Wehrly, K. E., M. J. Wiley, and P. W. Seelbach. 2006. Influence of landscape features on summer water temperatures in lower Michigan. Pages 113–128 in R. M. Hughes, L. Wang, and P. W. Seelbach, editors. *Landscape influences on stream habitats and biological assemblages*. American Fisheries Society, Symposium 48, Bethesda, Maryland.
- Weigel, B. M., J. Lyons, P. W. Rasmussen, and L. Wang. 2006. Relative influences of environmental variables at multiple spatial scales on fishes in Wisconsin's warm-water nonwadeable rivers. Pages 493–511 in R. M. Hughes, L. Wang, and P. W. Seelbach, editors. *Landscape influences on stream habitats and biological assemblages*. American Fisheries Society, Symposium 48, Bethesda, Maryland.
- Wiley, M. J., S. L. Kohler, and P. W. Seelbach. 1997. Reconciling landscape and local views of aquatic

- communities: lessons from Michigan trout streams. *Freshwater Biology* 37:133–148.
- Zorn, T. G., P. W. Seelbach, and M. J. Wiley. 2004. Utility of species-specific, multiple linear regression models for prediction of fish assemblages in rivers of Michigan's lower peninsula. Michigan Department of Natural Resources, Fisheries Research Report 2072, Ann Arbor. Available: www.michigandnr.com/PUBLICATIONS/PDFS/ifr/ifrlibra/research/reports/2072rr.pdf. (November 2006).
- Zorn, T. G., and M. J. Wiley. 2006. Influence of landscape characteristics on local habitat and fish biomass in streams of Michigan's lower peninsula. Pages 375–394 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Landscape influences on stream habitats and biological assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.