ELSEVIER

Fisheries Research 60 (2003) 17–32

FISHERIES RESEARCH

www.elsevier.com/locate/fishres

Predicting freshwater fish distributions using landscape-level variables

David G. Argent [a,*], Joseph A. Bishop [b], Jay R. Stauffer, Jr. [c], Robert F. Carline [d], Wayne L. Myers [b]

[a] *Biological and Environmental Sciences Department, California University of Pennsylvania, 250 University Avenue, California, PA 15419, USA*

[b] *The School of Forest Resources, The Pennsylvania State University, Land and Water Building, University Park, PA 16802, USA*

[c] *The School of Forest Resources, The Pennsylvania State University, 2C Ferguson Building, University Park, PA 16802, USA*

[d] *The Pennsylvania Cooperative Fish and Wildlife Research Unit, 113 Merkle Laboratory, University Park, PA 16802, USA*

Received 21 August 2001; received in revised form 13 March 2002; accepted 27 March 2002

Abstract

Management and conservation of aquatic systems requires the ability to identify species' historical, current, and potential distributions. We explored how a geographic information system can be used in conjunction with a few broad landscape variables to provide watershed-scale information useful for identifying diverse aquatic areas and predicting potential fish habitat. We developed species habitat profiles for all fish species that are known to occur in Pennsylvania. Five landscape variables were used to characterize a species' habitat profile to predict its statewide distribution: presence in a major drainage basin, presence in a physiographic region, median watershed slope, level of watershed disturbance, and watershed-stream size. Each of these variables was referenced to a small watershed boundary. Using these variables, we predicted a species potential habitat range. Distribution maps that we generated were then compared to known distributions with an average accuracy of 73%. While many collections have been made in Pennsylvania over the last 50 years, we determined that many areas still remain unexplored as potential sampling locations. Among those fishes whose predicted distribution was less than the actual sampled distribution, four receive special protection in Pennsylvania and one is federally endangered. Moreover, we determined that small watersheds (1:24,000 scale) in the Allegheny River drainage, in the Pittsburgh Low Plateau Section, of small size (3–4 order), with moderate slope (2–4%), and moderate watershed disturbance (25–75%) have the highest fish species richness. Our results should facilitate the conservation of fish species and our technique should be easily repeatable in other geographic areas.

Keywords: GIS; Landscape; Fish; Habitat; Conservation

1. Introduction

The continued decline of fish biodiversity is a major concern of state and federal fish and wildlife management agencies. In the United States, fish species are among the most imperiled of any biotic group (TNC, 1996). Over the last 100 years, three genera, 27 species, and 13 sub-species of fish have been extirpated from North America (Miller et al., 1989). Recently, 364 North American fish taxa were classified as endangered, threatened, or of special concern (Williams et al., 1989; Nehlsen et al., 1991; Mayden

---

[*] Corresponding author. Tel.: +1-724-938-4200; fax: +1-724-938-1514.
E-mail address: argent@cup.edu (D.G. Argent).

0165-7836/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0165-7836(02)00076-0

et al., 1992). The need for information to aid in the conservation of the nation's aquatic resources has never been greater than it is today.

Over the past decade, advancements in geographic information systems (GIS) and landscape ecology theory have increased our understanding of aquatic systems (Wiens, 1989; Schlosser, 1991; Meaden, 1996; Allan and Johnson, 1997; Johnson and Gage, 1997). The ability to objectively characterize aquatic biodiversity, particularly at the watershed or landscape level, has only recently been explored (e.g., Angermeier and Bailey, 1992). Watersheds are particularly useful for aquatic biodiversity analyses, because they span large-land areas, encompass a connected range of stream sizes, integrate natural and altered properties of a drainage area (Imhof et al., 1996), represent subdivisions of a species' overall distribution, and incorporate differences in regional distribution that other large-scale units do not (e.g., political boundaries). Watersheds provide a link between aquatic and terrestrial environments and can form an important management and conservation unit.

Management and conservation of aquatic systems requires the ability to identify species' historical, current, and potential distributions. This idea is one of the central tenets of the gap analysis program (GAP; Scott et al., 1993). The gap analysis program has evolved from the growing consensus that efforts designed to prevent the extinction of endangered and threatened species do not address the larger problem of fragmentation, habitat loss, and disruption of ecological processes, all of which contribute to a number of once common species being placed on the endangered species list (Noss et al., 1995). The GAP authors suggest that piecemeal approaches to species-by-species conservation are not as effective at maintaining biological diversity, as are community and ecosystem wide approaches (Scott et al., 1993). To date, GAP has focused primarily on the conservation needs of terrestrial organisms with only a few studies targeted at aquatic systems (e.g., those in New York, Mississippi, Missouri, and Oregon). Ricciardi and Rasmussen (1999) demonstrated that the projected future extinction rate for freshwater fauna is about five times greater than the rate for terrestrial fauna and three times the rate for coastal marine mammals. Conservation efforts therefore, need to shift from the terrestrial to the aquatic environment and methods need to be developed that identify the potential locations of diverse fish assemblages and the locations of imperiled fish assemblages.

We explored how a GIS can be used in conjunction with a few broad landscape variables to develop a fish distribution classification that could determine the availability of potential fish sampling locations, provide watershed-scale information useful for targeting conservation measures, and identify diverse aquatic areas. The ability to assess fish assemblages and their general habitat requirements should be of great value to resource managers, because it will facilitate targeted restoration and conservation efforts, identify fish habitat, identify potential areas for recolonization, and direct sampling efforts to localities that meet specific fish habitat requirements. Our objectives were to assess the potential of our GIS-based fish distribution classification to correctly identify sampled fish habitat, to identify potential fish habitat using derived fish distribution classification variables, and to characterize the diversity of fishes within a series of small watersheds (about 0.025–266 km^2 in size).

2. Methods

2.1. Study region

Our study was conducted using geographic data derived for Pennsylvania. We selected Pennsylvania because there exists a variety of digital-geographic data that can facilitate aquatic analyses at the state level. Secondly, a watershed data layer representing an appropriate spatial scale for analysis was available. This data layer, edited by the Environmental Resources Research Institute (ERRI, 1997) from a data set produced by the Water Resources Division of the US Geological Survey using 7.5″, 1:24,000 scale quadrangle maps, contained delineated watersheds for all named streams in the state. Lastly, a large fish data set (see Argent et al., 1997) was available to evaluate the accuracy and prediction potential of our fish classification output (Table 1). For many species included in this analysis, Pennsylvania represents the eastern (e.g., Johnny darter, *Etheostoma nigrum*) and northern (e.g., Potomac sculpin, *Cottus girardi*) portion of the range, but for others Pennsylvania is near the center (e.g., Tesselated darter, *Etheostoma*

Table 1
Pennsylvania fishes databases[a]

Database or contact person	Years covered	No. collections used
Pennsylvania Fish and Boat Commission (PFBC), William Frazier	1975–1995 (stream)	10780
	1975–1997 (lake)	10019
Edwin L. Cooper (ELC)	1932–1983	1500
The Academy of Natural Sciences in Philadelphia (ANSP), William Saul and Jon Gelhaus	1900–1989	530
The Pennsylvania State University (PSU), Jay R. Stauffer Jr.	1974–1999	408
Cornell University (CU), Charles M. Dardia	1904–1989	404
The University of Michigan's Museum of Zoology (UMMZ), William L. Fink	1903–1974	165
The Smithsonian Institution (SMITH)	1900–1984	126
Environmental Protection Agency (EPA)	1993–1995	88
Robin Heard (RH)	1994–1995	70
USEPA—National Exposure Research Laboratory (MAHA), Frank H. McCormick	1994	58

[a] Only data from 1950 to the present were included among these databases for comparison between actual fish habitat and potential fish habitat, because they were used for comparable sampling methods.

olmstedi) of a species' geographic range (Lee et al., 1980). Pennsylvania's fish fauna is therefore quite diverse and includes many fishes common to other states in the eastern United States.

2.2. Study design and general procedure

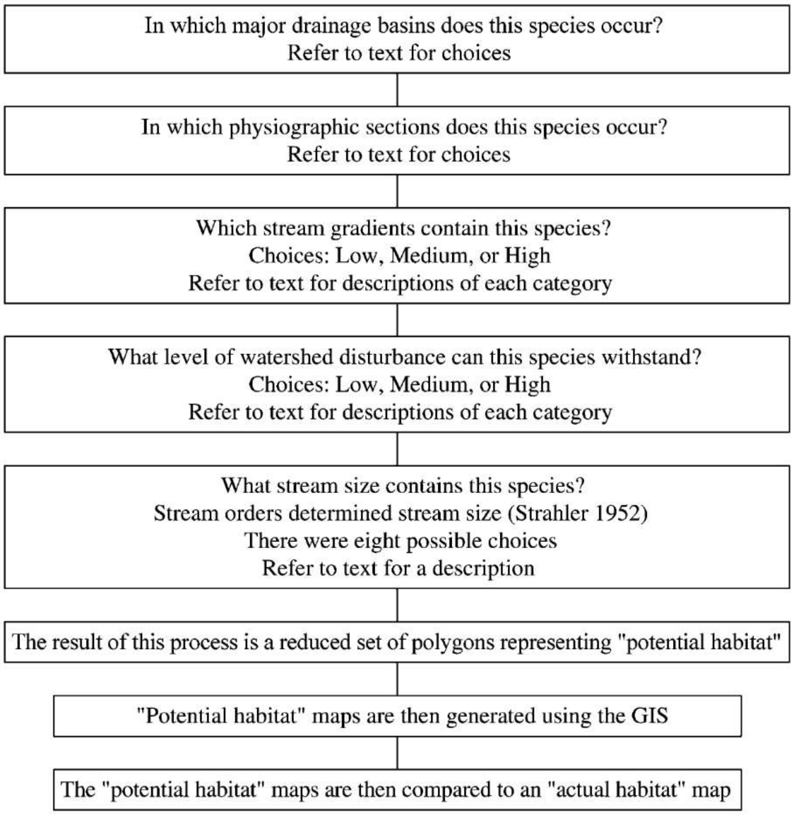
To develop a predictive fish habitat-distribution classification, we first had to select a suite of landscape-level characteristics that identify a species' habitat profile and then assemble the variety of spatial data layers that encompass those variables. We referenced several fisheries textbooks that describe regional faunas (Lee et al., 1980; Trautman, 1981; Cooper, 1983, 1985; Smith, 1985; Jenkins and Burkhead, 1994; Stauffer et al., 1995), referenced recent summaries of fish distribution patterns in the literature, conferred with local and regional experts and ground-truthed species accounts to develop our fish habitat-distribution classification variables. A habitat profile for each species was developed using these resources among five landscape-level variables (Table 2). These five variables were selected because they represented unique descriptors of the landscape relating fish distribution patterns to historical geology and land use. Each of the five variables was available for use with the Arc/Info® GIS software (ESRI, 1998). Among these five variables, we derived 34 character states (see Table 7 for summary) to develop our classification. We identified habitat variables for all fish species that are known to have had populations in Pennsylvania over the last 50 years. We selected this 50-year period because the large majority of data that we had available to test our classification came from this time period. We recognize that changes in habitat have occurred over this time period, which would have influenced the distribution of many fishes, but data covering a significant portion of Pennsylvania in the recent past is unavailable, so we have used what we consider to be the best available data for comparison. We included anadromous and catadromous species that currently occur in Pennsylvania because they use a portion of the habitat available in the state to complete their life cycle.

Using 34 character states (see Table 7 for summary), we developed a fish distribution classification that selected species-specific habitat from our GIS database. We developed distribution profiles, similar to those in Table 2 for all species using data that we derived from the literature, expert opinion, and the "fishes of books". Once these habitat profiles had been established, we queried the Arc/Info® GIS software (ESRI, 1998) data layers to create a reduced map of potential habitat for each species that met our geographic criteria (Figs. 1 and 2). The output from

Table 2
Fish distribution classification character states and predicted presence for five selected fishes[a]

Species	Variables						
	Drainages present	Phys. section present	Slope occurrence	Stream size occurrence	Disturbance occurrence	Actual collected	No. correctly predicted
Ohio lamprey, *I. bdellium*	2	3	1, 2	3	1, 2	31	17
Brook trout, *Salvelinus fontinalis*	8	16	2, 3	1, 2	1	1410	1189
Redfin pickerel *Esox a. americanus*	2	9	2	1, 2	2	88	45
Cutlips minnow, *Exoglossum maxillingua*	5	14	2	1, 2, 3	2	818	796
Gilt darter, *Percina evides*	1	3	3	1, 2, 3	1, 2	15	14

[a] Drainage column denotes species' presence out of eight possible drainage basins, Phys. Section denotes species' presence out of 16 known physiographic sections; slope, stream size, and disturbance reflects classifications as described in Section 2. Table 7 contains the complete list of characters used in this analysis, but a complete list by species is available from the authors. Numbers for slope, stream size, and disturbance represent divisions for each variable listed in Table 7.



In which major drainage basins does this species occur?
Refer to text for choices

In which physiographic sections does this species occur?
Refer to text for choices

Which stream gradients contain this species?
Choices: Low, Medium, or High
Refer to text for descriptions of each category

What level of watershed disturbance can this species withstand?
Choices: Low, Medium, or High
Refer to text for descriptions of each category

What stream size contains this species?
Stream orders determined stream size (Strahler 1952)
There were eight possible choices
Refer to text for a description

The result of this process is a reduced set of polygons representing "potential habitat"

"Potential habitat" maps are then generated using the GIS

The "potential habitat" maps are then compared to an "actual habitat" map

Fig. 1. Flowchart diagram describing how we selected fish habitat from a base map containing 9854 small watersheds, each described by the various character states found in Table 7. At each question various polygons were removed from the original set representing specific habitat for each species.

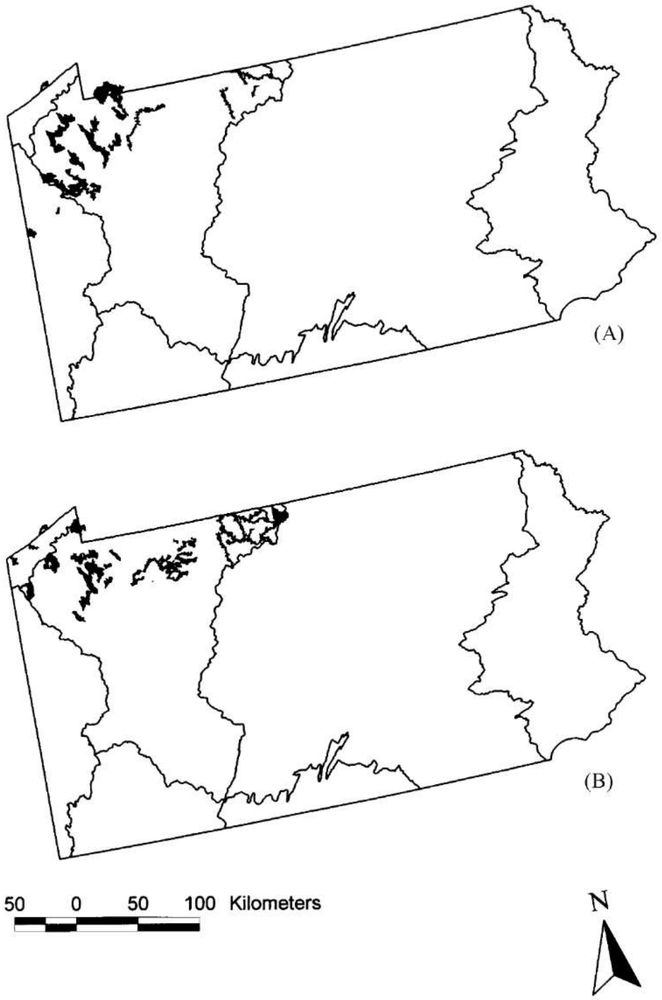


Fig. 2. An example map containing sampled (actual) habitat locations (A) and predicted (potential) habitat locations (B) for the Ohio lamprey (*Ichthyomyzon bdellium*) by small watershed. The predicted habitat locations are those that result from the procedure diagrammed in Fig. 1. Map is shown in Lambert projection. Map background denotes the eight major drainage basin divisions.

this procedure yielded 153-distribution maps specific to each species, identifying potential habitat among 9854 small-delineated watersheds. Several historically known species were omitted because we could not make accurate assessments as to their distribution among our five habitat variables (e.g., Paddlefish, *Polyodon spathula*). The potential-habitat maps were then compared to sampled-habitat maps that we generated using the fish database described below (Fig. 3).

2.3. Fish distribution classification variables

Our base geographic data layer was a small watersheds coverage that we edited to include other watershed data (ERRI, 1997). This data layer served as our smallest spatial extent for analysis. The base coverage contained 9854 separate polygons representing small watershed areas with named streams officially recognized by the Board on Geographic Names and other unofficially named

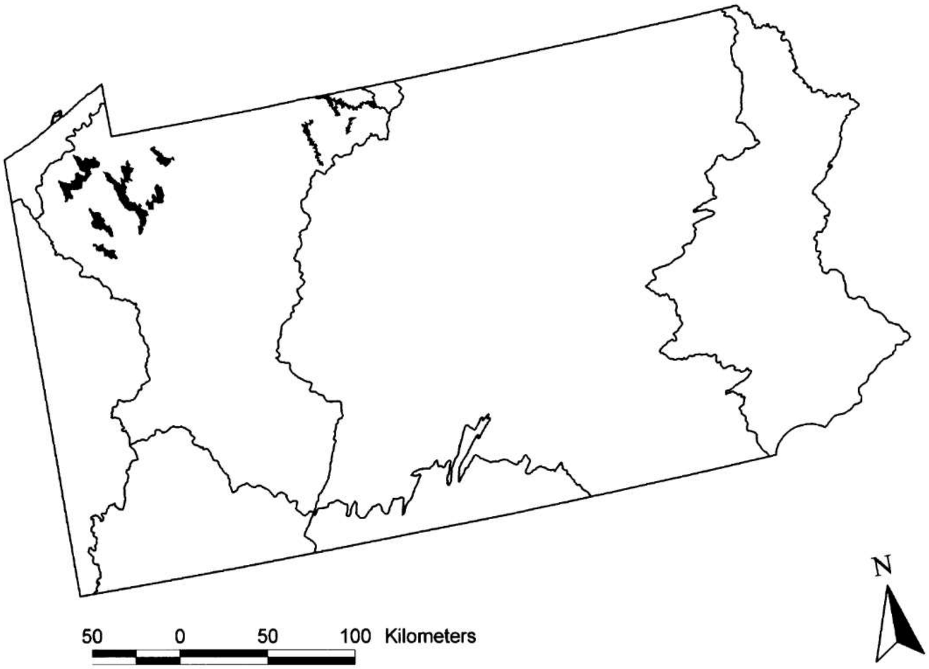


Fig. 3. Map containing the overlap of sampled (actual) habitat locations and predicted (potential) habitat locations for the Ohio lamprey (*I. bdellium*) depicted in Fig. 2. Map is shown in Lambert projection. Map background denotes the eight major drainage basin divisions.

streams that flow through named hollows (ERRI, 1997).

We added attributes to the watershed data layer that reflect the fish distribution classification variables described below, so that each small watershed contained a unique value and descriptor for each character state. Character states were derived among the five variables: presence in a major drainage basin, presence in a physiographic region, median-watershed slope, level of disturbance within each watershed, and watershed-stream size. Lastly, an attribute was added to indicate if a watershed had previously been sampled.

2.3.1. Major drainage basin coverage

Major drainage basin boundaries were delineated using 1:100,000 scale USGS quadrangle maps and 500 m digital elevation models. Drainage basin boundaries often reflect the largest spatial extent for fish species' distribution and often reflect unique fish associations (Hocutt and Wiley, 1986). Within Pennsylvania, there are six major basins, but we decided that further divisions of the Ohio River drainage were necessary to accurately predict fish faunas of the Monongahela and Allegheny river drainages. Divisions within other drainages like the Susquehanna into west and north branches were not made because we believed that the fauna of these areas was accurately represented at the scale selected. Our resulting coverage included eight drainage divisions or character states: Allegheny, Delaware, Erie, Genesee, Monongahela, Ohio, Potomac, and Susquehanna. Significant portions of the Allegheny and Monongahela rivers are contained in Pennsylvania, which form the Ohio River; over two-thirds of the Susquehanna River drainage is contained in Pennsylvania; and the headwaters of the Genesee, Potomac, and Erie drainages are contained within Pennsylvania's political border. In terms of zoogeography, Pennsylvania's drainages encompass or contribute to the large majority of streams and rivers in the eastern contiguous United States. Each species was classified as present or absent within each major drainage basin.

2.3.2. Physiographic section coverage

In 1994, the Pennsylvania Bureau of Topographic and Geologic Survey created a GIS layer at a 1:100,000 scale representing the physiographic provinces and sections of Pennsylvania. We used this data layer to represent a large spatial extent for fish distribution that reflects land physiography. Boundaries for this data layer were based primarily on geology using published geological maps. It has been hypothesized that the underlying distribution of many fishes is related to the geology of the landscape (Hocutt and Wiley, 1986); thus the importance of this category in our classification. Each species was classified as present or absent among 16 recognized physiographic sections (Eastern Lake, Mountainous high plateau, Pittsburgh, Low plateau, Glaciated Pittsburgh plateau, High plateau, Glaciated low plateau, Glaciated Pocono plateau, Allegheny mountain, Appalachian mountain, Great valley, Reading prong, South mountain, Piedmont upland, Piedmont lowland, Gettysburg-Newark lowland, and Lowland and intermediate). Sections were used over Provinces because of the increased resolution obtained from a smaller scale unit. Sections also represent units that are intermediate in scale between drainage basins and small watersheds.

2.3.3. Stream gradient

Median stream slope was determined for each small watershed using the Spatial Analyst extension in the ArcView® GIS software (ESRI, 1997). This category was selected to separate fish habitat along the longitudinal axis of a stream because some fishes occupy streams of low gradient, while others prefer higher gradient areas (Sheldon, 1968). Low gradient streams in Pennsylvania typically have sand, silt, and clay substrates, while high gradient streams typically have cobble, boulder, and rock substrates. Medium gradient streams occur between the two extremes and often have a heterogeneous mix of substrate types. Upon generating summary statistics for a 30 m × 30 m digital elevation fish distribution classification, derived by USGS (1999), we assigned three character states for watershed slope: low ($\leq 2\%$), medium ($2\% < x \leq 4\%$), and high ($>4\%$). Classes were used instead of continuous variables to simplify our overall classification procedure. Slope ranges were determined by reviewing fish life histories and by expert opinion. Each species was assigned a slope classification using these three ranges. Some species were classified in more than one of the three slope variables because of their occurrence over a broad range of gradients.

2.3.4. Watershed disturbance

Watershed disturbance was defined as any pixel within our land use coverage that contained agricultural or developed land, i.e., non-forested land (Table 3). Using a 30 m × 30 m grid data layer derived from LandsatThematic-Mapper imagery (Bishop,

Table 3
Land use data, derived from Landsat Thematic Mapper imagery (Bishop, 1998), used to quantify terrestrial disturbance in each small watershed[a]

Land use class	Description	Proportion
Water (unclassified)		
Water	Open water	1.14
Forested landscapes		
Conifer forest	Coniferous forests	1.40
Mixed forest	Mixed coniferous and deciduous stands	8.14
Broad-leaf forest	Deciduous forests	49.76
Transitional vegetation	Re-growth in clear-cut areas	9.79
Disturbed landscapes		
Perennial herbaceous	Predominantly pasture and old fields	14.77
Annual herbaceous	Predominantly agricultural land	11.32
Terrestrial unvegetated	Developed, roadways, parking lots	3.67

[a] Perennial herbaceous, annual herbaceous and terrestrial unvegetated classes were collapsed to denote disturbed land. Proportion denotes the percentage of each land use type found across Pennsylvania.

1998), we classified each small watershed by its disturbance percentage. Across Pennsylvania, 29.76% of the overall landscape (not the percent of watersheds) was classified as disturbed and 69.09% was classified as forested (Table 3). The remaining 1.14% was classified as undetermined using our procedure (the majority of which is open water). Disturbance within each small watershed was defined among three character states as low ($\leq$25% land area disturbed), medium ($25\% < x \leq 75\%$ land area disturbed), or high (>75% land area disturbed). This category was selected to represent a measure of anthropogenic stress and a measure of tolerance to human induced landscape influences. Such disturbances at the local scale may affect stream habitats (Richards et al., 1996), consequently influencing the presence or absence of fishes. Among the species classified within these three disturbance variables, some species were assigned to more than one category because we determined that some fishes could inhabit streams that span a range of disturbance levels.

2.3.5. Stream size

Stream order is a proximal measure of stream size and is often correlated with watershed area. Stream order often reflects the stream macrohabitat of a fish species (Sheldon, 1968). We determined stream order using Strahler's (1952) method. For analysis, we collapsed individual stream orders identified into one of four stream classes of similar size (Table 4).

2.3.6. Fish data layers and distribution classification

To test the predictability of our fish distribution classification, a large fish data set was assembled (Argent et al., 1997). This data set was used solely to validate our fish distribution classification variables, not to generate them. Digital representations of contemporary databases (PFBC, RH, and EPA; see Table 1 for acronyms) were created using latitude and longitude information about each collection site. By manually locating sites, using site descriptions, it was possible to generate digital representations of more historic databases (UMMZ, PSU, ELC, CU, SMITH, and ANSP; see Table 1 for acronyms) using the Arc/Info® GIS software (ESRI, 1998).

Table 4
Stream size classifications

Stream order	Classification
1–2	Headwater
3–4	Small
5–6	Medium
7–8	Large

Data were scrutinized for accuracy of species identification and site locality. Distribution maps were reviewed to identify questionable records and spurious accounts. Questionable records were removed or corrected after careful review by one of three different methods, to ensure that distribution accounts accurately represented the occurrence of Pennsylvania's fishes. First, voucher specimens were available for most collections except the PFBC and RH, and were reviewed as needed. Second, experts reviewed questionable specimens and third, the Pennsylvania State University staff made additional collections to confirm other questionable accounts. The final data set used for fish distribution validation contained over 20,000 collections from 1950 to 1999. The intersection of the fish collection points with their respective watershed indicated that the approximate watershed area sampled with these collections accounted for 52% of the total available land area within Pennsylvania or 2880 of the 9854 delineated watersheds. While prior sampling most likely does not reflect the diversity of habitats present in all watersheds, it is a good starting point to develop a classification such as ours. Extrapolation beyond the reach actually sampled may be a shortcoming of our method because data used for this classification were not expressly collected for this project. However, the inherent value of our approach is that it uses one of the most complete data sets available for Pennsylvania in a first attempt at a statewide classification. The time period selected encompassed important historic collections made by Edwin L. Cooper, recent collections made by Jay R. Stauffer Jr., and annual collections made by the PFBC. Using these data, 151 distribution maps were generated and compared with predicted species distribution maps. Two species, the silver lamprey, *Ichthyomyzon unicuspis*, and hickory shad, *Alosa mediocris* were omitted, because collections of these fishes were beyond the terrestrial borders of Pennsylvania, occurring in Lake Erie and the Delaware River, respectively.

2.4. Analysis

We assessed the ability of our habitat fish distribution classification to predict potential fish habitat in sampled watersheds and assessed the fish distribution classifications' over and under prediction of habitat relative to actual habitat areas we determined. For validation purposes only those watersheds that had been sampled were compared with watersheds correctly identified as potential habitat. The two-sample rank test, Mann–Whitney, was used to determine if the population distribution of actual habitat and area was identical to that of potential habitat and area for each fish species at the $\alpha = 0.05$ significance level. Statistical analyses were performed using JMP'IN (SAS, 1998).

3. Results

Our fish distribution classification predicted potential habitat for 153 fishes from 23 Families (e.g., Table 2). Among the 9854 watersheds, an average of 1424 (S.E. = 48) watersheds were determined to contain potential habitat for an individual species. When compared with known fish distributions, our fish distribution classification identified sampled locations with an average accuracy of 73%. Conversely, this indicates that 27% of the time our fish distribution classification did not predict a species' presence in a watershed where it was known to occur. However, predicted habitat locations were significantly correlated with actual habitat locations ($r = 0.99$, $P < 0.0001$), suggesting a tight association between actual and correctly identified potential habitat.

Of the 23 fish families represented in our data, habitat for 20 of them was predicted with at least 55% accuracy (Fig. 4). A level of 55% encompassed the majority of fish families within one standard deviation of the 73% average accuracy prediction. The three families for which our fish distribution classification failed to predict suitable habitat included: Amiidae, Acipenseridae, and Lepisosteidae. Removing these three fish families from the analysis increased our prediction accuracy to 77%. There were five species for which the fish distribution classification did not correctly predict at least one sampled watershed (Table 5). With the exception of sockeye salmon, *Oncorhynchus nerka*, an introduced species, these species all receive special protection in

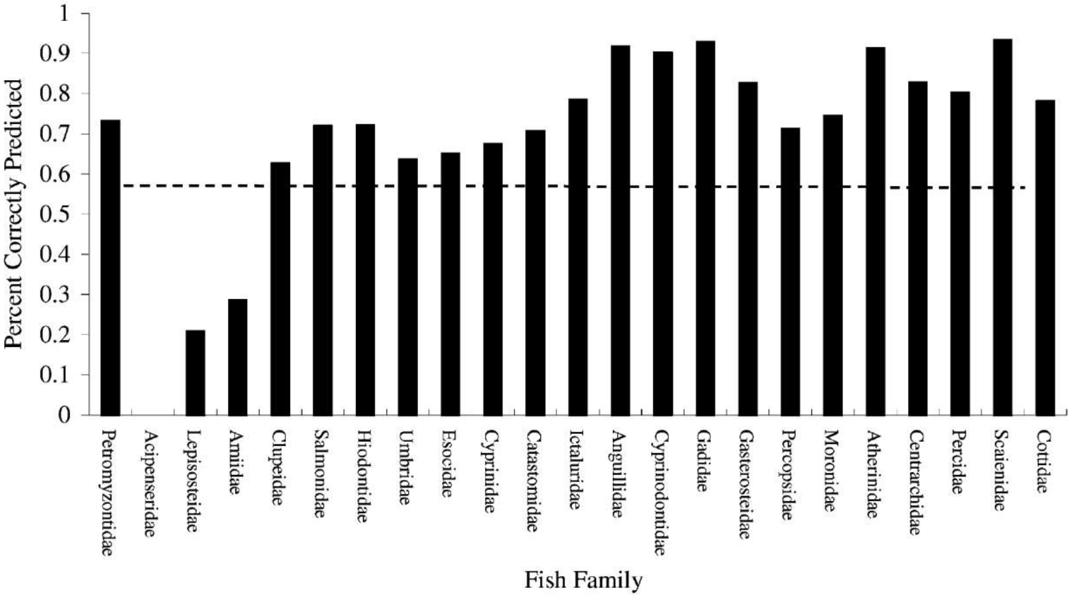


Fig. 4. Comparison of fish distribution classification accuracy among fish Families. The percent correctly predicted denotes the proportion of small watersheds correctly predicted using the fish distribution classification compared to small watersheds actually sampled.

Table 5
Species for which the fish distribution classification did not predict at least one watershed correctly

Common and scientific name	No. sampled watersheds	No. predicted watersheds
Shortnose sturgeon, *Acipenser brevirostrum*	1	1
Atlantic sturgeon, *Acipenser oxyrhynchus*	1	4
Spotted gar, *Lepisosteus oculatus*	1	114
Sockeye salmon, *O. nerka*	1	10
Ironcolor shiner, *N. chaleybaeus*	0	4

Pennsylvania. Of sites sampled containing introduced species, our classification predicted their presence correctly 74% of the time.

We determined that there are many more potential sampling locations for fish in Pennsylvania that meet suitable landscape-habitat criterion ($P < 0.001$). There was approximately three times as much potential habitat identified for a given species than had been sampled. There were several fishes whose predicted habitat was less than the actual sampled habitat, but these differences were not statistically significant ($P = 0.125$, Table 6). Among these seven fishes, four receive special protection in Pennsylvania and one is federally endangered. More potential habitat was identified using the fish distribution classification among Pennsylvania's remaining rare and endangered fishes than has previously been sampled ($P < 0.001$).

We identified potential habitat for freshwater fishes that range from relatively undisturbed areas (e.g., those in which humans have minimal access) to degraded areas (e.g., those in which have large amounts of agricultural or developed land) (Table 7). We concluded that watersheds in the Pittsburgh Low Plateau Section of the Allegheny River drainage, of small size (3–4 order), with moderate slope ($2\% < x \leq 4\%$), and moderate watershed disturbance ($25\% < x \leq 75\%$) have the highest fish species richness, while watersheds of low ($\leq 2\%$) and high slope ($>4\%$), of high watershed modification ($>75\%$), and headwater stream size (1–2 order) have the lowest fish species richness. Assemblages of lower fish species richness are also predicted to occur in the South Mountain Section and Glaciated Pocono Plateau Section as well as the Genesee and Potomac river drainages. Lower diversity in these areas occurs because they are smaller in total land area within Pennsylvania than larger major drainages and physiographic sections, and because these areas are comprised of many small headwater stream reaches within Pennsylvania.

Analysis of our actual habitat database revealed that even with extensive collections including over 20,000 sampling events spanning 50 years, only 2880 of 9854 delineated watersheds were sampled or 52% of the watersheds delineated for the state. This comparison also indicates that of larger streams (>3rd order) proportionately more have been sampled than smaller streams of first and second order, of these only 6 and 18% have been sampled at least one time, respectively (Fig. 5).

Table 6
Species whose predicted habitat area was less than the actual habitat area (areas given in the table are in km^2)

Common and scientific name	Predicted area	Actual area	Status[a]
Shortnose sturgeon, *A. brevirostrum*	2.56	8.37	FE, SE
Steelhead, *Oncorhynchus mykiss*	85.32	240.79	NS
Mooneye, *Hiodon tergisus*	80.52	161.79	ST
Eastern silvery minnow, *Hybognathus regius*	16.45	89.87	NS
Emerald shiner, *Notropis atherinoides*	624.96	1380.27	NS
Spotted darter, *Etheostoma maculatum*	116.51	146.92	ST
Channel darter, *Percina copelandi*	58.29	137.49	ST

[a] FE: Federally endangered; SE: state endangered; ST: state threatened; SC: state candidate; NS: no special status.

Table 7
Predicted species richness among five landscape habitat variables[a]

Major drainage basin	Predicted species richness
Genesee	21
Potomac	51
Erie	74
Monongahela	74
Susquehanna	78
Delaware	79
Ohio	81
Allegheny	111
Physiographic province—physiographic section	
Appalachian plateaus—Glaciated Pocono plateau	26
Blue Ridge—South mountain	32
New England—Reading prong	47
Piedmont province—Piedmont lowland	47
Appalachian plateaus—Mountainous high plateau	53
Atlantic Coastal Plain—Lowland and intermediate	53
Appalachian plateaus—Allegheny mountain	59
Appalachian plateaus—Glaciated low plateau	62
Piedmont province—Piedmont upland	65
Piedmont province—Gettysburg-Newark lowland	65
Ridge and valley—Great valley	68
Central lowland—Eastern Lake	71
Ridge and valley—Appalachian mountain	72
Appalachian plateaus—High plateau	83
Appalachian plateaus—Glaciated Pittsburgh plateau	92
Appalachian plateaus—Pittsburgh low plateau	103
Watershed-stream size—categorized from 1 to 4	
(1) Headwater	7
(2) Small	99
(3) Medium	52
(4) Large	28
Median watershed slope—categorized from 1 to 3	
(1) Low (≤2%)	68
(2) Medium (2% < x ≤ 4%)	117
(3) High (<4%)	67
Watershed disturbance—categorized from 1 to 3	
(1) Low (≤25%)	55
(2) Moderate (25% < x ≤ 75%)	126
(3) High (<75%)	26

[a] Some fish are predicted to occur in more than one category within each variable.

4. Discussion

Our results indicate that identifying potential fish habitat over large areas can be successfully accomplished using a few broad habitat variables. Spatially, those areas in northwestern Pennsylvania appear to harbor the most diverse fish assemblages, probably because of the glacial history of the region (Argent et al., 1998; Cooper, 1983) while those areas adjacent to developed lands appear to harbor modestly diverse fish assemblages, probably because of the associated disturbance of developed lands (Argent, 2000).

Other studies have noted the role that large-scale factors play in determining fish distribution patterns. Modde et al. (1991) described the role of geologic deposits in determining fish distribution and abundance

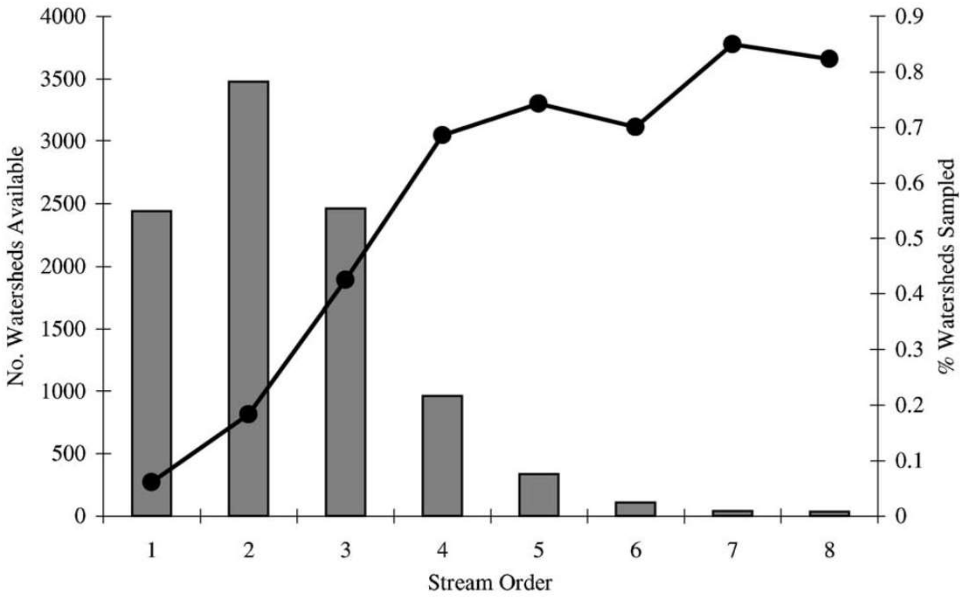


Fig. 5. Plot of number of watersheds ($N = 9854$) and percentage of sampled watersheds versus stream order. Bar graph denotes number of watersheds available at each stream order and point/line graph denotes percent of watersheds sampled at each stream order.

patterns in South Dakota. Using multivariate statistical approaches, Matthews and Robison (1988, 1998) demonstrated the strong association among Arkansas fish assemblages, environmental factors, drainage connectivity, and historical zoogeography. Others have evaluated the correspondence between ecoregions, areas of similar land-surface form, soils, and vegetation; and spatial patterns in stream ecosystems (Whittier et al., 1988; Lyons, 1989; Oberdorff et al., 1995; Wiley et al., 1997). Although not predictive in nature, these studies underscore the importance of the macro-ecological approach in explaining local fish assemblage structure in relation to larger geographic variables.

Our fish distribution classification identified watersheds that had previously been inventoried for fishes with an average accuracy of 73%. Few studies are available in the literature for comparison to assess if 73% is a reasonable level of accuracy to accept. Our correlation analysis suggests that actual habitat can be predicted using our fish distribution classification variables with a high level of confidence for most species. One recent study used simulation approaches to validate a fish habitat classification for Minnesota lakes (Stefan et al., 1995). Here, water temperature and dissolved oxygen concentrations were used to develop coldwater, coolwater, and warmwater simulation habitat models. Stefan et al. (1995) confirmed the presence of coldwater fishes with 91% accuracy, of coolwater fishes with 85% accuracy, and of warmwater fishes with 59% accuracy, or an average of 78%. While their approach relied on simulations to validate the outcome and a different approach to assessing habitat, overall accuracy is similar to the ability of our fish distribution classification to predict suitable fish habitat.

Analysis of our fish database revealed a sampling bias towards larger stream orders than smaller stream order (Fig. 5). We conclude from these results that future efforts should consider increasing the number of smaller streams sampled, while maintaining a suitable sampling level among other stream sizes. Moreover, our classification identified that streams between third and fourth order had the highest species richness. This result was counter to our a priori assumption that larger systems would contain higher fish species richness (sensu Vannotte et al., 1980). We attribute this result to the fact that numerically there

are fewer higher order streams than middle order streams, which contribute a greater diversity of fishes at the landscape scale than do larger streams.

Among the fishes not accurately represented with this classification were members of the families Amiidae, and Acipenseridae; and spotted gar, *Lepisosteus osseus*; chinook salmon *Oncorhynchus tshawytscha*; and ironcolor shiner, *Notropis chaleybaeus*. Our fish distribution classification detected only two of the five sampling locations for Amiidae probably because this species has a disjunct distribution in Pennsylvania (see Argent et al., 1997), the fish distribution classification did not appropriately account for its habitat, and we did not include distribution information for lentic systems. Collections for both sturgeon species were beyond the terrestrial borders of Pennsylvania; therefore, actual habitat maps as they relate to our watershed analysis could not be derived. This does not preclude the importance of such habitat as it may have a direct influence on the downstream areas of the Delaware Estuary. We identified spotted gar, as occupying areas of medium slope, but the one collection for this species was taken in a watershed identified with a high median slope. The sockeye salmon, a non-native species, collection representing actual habitat came from landlocked Upper Woods Pond, Wayne County in the Delaware drainage that had been stocked to support recreational angling, but the fish distribution classification predicted the presence of this species in the Erie drainage coincident with chinook salmon *O. tshawytscha* habitat. Lastly, collections for the ironcolor shiner, *N. chaleybaeus* have only recently been made (Criswell, 1998) and at the time of our analysis were not included in the databases described in Table 1. Of 22 historic sites where this species was known to occur, only one contained this species in recent collections (Criswell, 1998).

The ability to predict potential fish habitat should provide resource managers with new opportunities for management and conservation. Potential sampling locations can be identified, providing resource agencies with targeted locations to sample-specific species or specific fish assemblages. Because our fish distribution classification output generates individual maps for each species, one can easily locate the habitat range for a particular fish or for an entire assemblage, simply by overlaying several maps with one another. This could provide a valuable information source that not only saves time, but also saves money when deciding where to sample for particular species and determining which areas might receive the greatest benefit from restoration efforts. Identification of potential habitat also provides a basis for proactive habitat management that may be important for the continued persistence of resident fishes.

Two questions should be further addressed, “can the fish distribution classification be improved?” and “can the data set used to test the predictability of this classification be improved?” There are several options to address theses questions: increase the number of character states among the five variables we initially selected or increase the number of habitat variables. The first option requires the division of the slope class, disturbance level, and stream size variables further than we described. Drainage basin and physiographic province variables were incorporated at a division within our classification reflective of large landscape influences and while further division may be possible it would add little to this classification because of the relationship between fish distribution and these two variables, but the other three variables could be further sub-divided. Among these variables, slope class appears to have the greatest variation within a watershed, ranging in some cases from 0 to 20%. Division of this category into four character states may increase separation along the vertical watershed axis and better separate those fishes occupying coldwater, coolwater, or warmwater habitats.

Further division among the disturbance levels may also increase our ability to predict potential habitat using the fish distribution classification. Because we grouped a range of 50% of disturbed habitat into one variable (medium or $25\% < x \leq 75\%$ land area disturbed), we emphasized the extremes. Our values were derived from the work of Meixler et al. (1996). Perhaps an arbitrary break at 25% intervals that equally weights disturbance types would improve our fish distribution classification. Additionally, dividing disturbance types into agricultural and urban impacts or splitting these variables into their component parts, e.g., pastureland or row crop agriculture may provide a more accurate fish distribution classification.

A restructuring of the stream size category may yield a better habitat fish distribution classification along the horizontal watershed axis. This may also include the use of watershed area as a potential

classification variable. We elected not to use watershed area because there is generally a high degree of correlation between area and stream order, and because we wanted to keep or classification as simple as possible. We arbitrarily assigned ordered watersheds to one of our four stream classifications, but differences between streams within each category may be masked by our generalization of stream size. Many third and fourth order streams, for example, are quite different from one another with respect to habitat type, hydrologic characteristics, and species assemblages. Each of these variables could be treated as continuous size measures rather than as categorically as we elected to do. A final option would be the comparison of this classification with alternate methods to refine the character states and to further test the utility of GIS to predict fish habitat.

With regard to increasing the number of variables, our choices are limited due to the geographic nature of these analyses. Additional variables must describe the entire landscape under investigation (i.e., Pennsylvania), have relevance to the aquatic system or serve as surrogates for instream measures, and be derived at comparable spatial scales. Traditional fisheries data, including stream width, flow, stream temperature, dissolved oxygen, and substrate characteristics (Stefan et al., 1995) are not available for all 9854 watersheds under investigation. Moreover, the seasonal and annual variation in these variables may not improve the predictive power of the fish distribution classification. Rathert et al. (1999) used artificial constructs to describe the relationship between fish species richness and landscapes. While statistically attractive, these classifications do not reflect the lay of the land—an element we were specifically trying to incorporate.

For our analysis, perhaps the best new variables to explore would be underlying bedrock geology or soil profiles. These two variables have been correlated with fish distribution patterns, do not possess the seasonal variance typically associated with other instream measures, and are generally available for the study area in question. But, increasing the number of variables or number of divisions within a category may lead to redundancy and increased correlation among variable already in the analysis, not to mention increasing the complexity of a data matrix that is already 153 rows (species) by 34 columns (variables) in size.

“Can the data set used be improved to test the predictability of this classification?” Most certainly it can. One of the pitfalls that are now being recognized by many who undertake GIS analyses is that adequate data are lacking. Many databases (ours included) used to test analyses use data not specifically collected for a particular investigation. Other authors (Angermeier and Smogor, 1995; Lyons, 1992; Paller, 1995) address issues of sampling inadequacy; a design consideration that we believe should be incorporated into future efforts to develop landscape-level fish classifications. As McLaughlin et al. (2001) indicate, because of the “variation in the quality of the data and the evenness of sampling, analysis and interpretation of the historical data” is challenging. McLaughlin et al. (2001) list five limitations to such data sets including: lack of time series data, data gaps, seasonal variation, efficiency of sampling, and comparability of surveys. These five limitations are present in the databases that we used to compare our classification, but for now they are the best data that we have to work with and until specific collection efforts are designed with such landscape analyses in mind we must work with them—recognizing the shortcomings.

To conclude, we identified large geographic areas that may be important for fish conservation and management as well as those that may be in need of restoration. Future research should focus on the development of better data sets to test such classification schemes addressing issues of sampling adequacy. An additional study could also be undertaken with our data set, checking our classification against a random classification, permitting the use of two techniques to assess the utility of our approach. These considerations will allow us to better use GIS in our fisheries planning, allow further understanding of how fish assemblages are structured at large scales, and consider sampling issues when working with large scale analyses.

Acknowledgements

We wish to thank the USGS Biological Resources Division for providing financial support for this study. The Environmental Resources Research Institute and those individuals and agencies listed in Table 1 are thanked for providing computer and data resources.

John Arway and Paola Ferreri are acknowledged for their input to refine the variables selected for this study. Ellen van Snik Gray and five anonymous reviewers are acknowledged for their thoughtful review of earlier drafts.

References

Allan, J.D., Johnson, L.B., 1997. Catchment-scale analysis of aquatic ecosystems. Freshw. Biol. 37, 107–111.

Angermeier, P.L., Bailey, A., 1992. Use of a geographic information system in the conservation of rivers in Virginia, USA. In: Boon, P.J., Calow, P., Petts, G.E. (Eds.), River Conservation and Management. Wiley, New York, pp. 151–159.

Angermeier, P.L., Smogor, R.A., 1995. Estimating number of species and relative abundances in stream-fish communities: effects of sampling effort and discontinuous spatial distributions. Can. J. Fish. Aquat. Sci. 52, 936–949.

Argent, D.G., 2000. Fish distribution patterns in Pennsylvania: anthropogenic influences and the status of and rare fishes. Ph.D. Thesis. The Pennsylvania State University, University Park, PA.

Argent, D.G., Carline, R.F., Stauffer Jr., J.R., 1997. Historical and contemporary distribution of fishes in Pennsylvania. Final Report. Research Work Order No. 47. USGS. Biological Resources Division, The Pennsylvania Cooperative Fish and Wildlife Research Unit, The School of Forest Resources, The Pennsylvania State University, Pennsylvania.

Argent, D.G., Carline, R.F., Stauffer Jr., J.R., 1998. Changes in the distribution of Pennsylvania fishes: the last 100 years. J. Penn. Acad. Sci. 72, 32–37.

Bishop, J.A., 1998. Vegetative land cover of Pennsylvania. Pennsylvania Gap Analysis Program. Environmental Resources Research Institute, The Pennsylvania State University, University Park, PA.

Cooper, E.L., 1983. Fishes of Pennsylvania and the Northeastern United States. The Pennsylvania University Press, University Park, PA.

Cooper, E.L., 1985. In: Genoways, H.H., Brenner, F.J. (Eds.), Fishes. Species of Special Concern in Pennsylvania. Special Publication of the Carnegie Museum of Natural History, No. 11. Carnegie Museum of Natural History, Pittsburgh, PA, pp. 169–256 (Chapter 3).

Criswell, R.W., 1998. Report: status survey of fishes of special concern in the Delaware River drainage, Pennsylvania. Final Report. Wild Resources Conservation Fund.

Environmental Resources Research Institute (ERRI), 1997. Pennsylvania—small watersheds. Pennsylvania Department of Environmental Protection, Harrisburg, PA.

Environmental Systems Research Institute, Inc. (ESRI), 1997. ArcView, Version 3.0b. New York Street, Redlands, CA.

Environmental Systems Research Institute, Inc. (ESRI), 1998. Arc/Info, Version 7.2.1. New York Street, Redlands, CA.

Hocutt, C.H., Wiley, E.O., 1986. The Zoogeography of North American Freshwater Fishes. Wiley, New York.

Imhof, J.G., Fitzgibbon, J., Annable, W.K., 1996. A hierarchical evaluation system for characterizing watershed ecosystems for fish habitat. Can. J. Fish. Aquat. Sci. 53 (Suppl. 1), 312–326.

Jenkins, R., Burkhead, N., 1994. Fishes of Virginia. The American Fisheries Society Press, Bethesda, MD.

Johnson, L.B., Gage, S.H., 1997. Landscape approaches to the analysis of aquatic ecosystems. Freshw. Biol. 37, 113–132.

Lee, D.S., Gilbert, C.R., Hocutt, C.H., Jenkins, R.E., McAllister, D.E., Stauffer Jr., J.R., 1980. Atlas of North American Freshwater Fishes. North Carolina State Museum of Natural History, Raleigh, NC.

Lyons, J., 1989. Correspondence between the distribution of fish assemblages in Wisconsin streams and Omernik's ecoregions. Am. Mid. Nat. 122, 163–182.

Lyons, J., 1992. The length of stream to sample with a towed electrofishing unit when fish species richness is estimated. N. Am. J. Fish. Manage. 12, 198–203.

Matthews, W.J., Robison, H.W., 1988. The distribution of the fishes of Arkansas: a multivariate analysis. Copeia 1988, 358–374.

Matthews, W.J., Robison, H.W., 1998. Influence of drainage connectivity, drainage area and regional species richness on fishes of the interior highlands in Arkansas. Am. Mid. Nat. 139, 1–19.

Mayden, R.L., Burr, B.M., Page, L.M., Miller, R.R., 1992. The native freshwater fishes of North America. In: Mayden, R.L. (Ed.), Systematics, Historical Ecology, and North American Freshwater Fishes. Stanford University Press, Stanford, CA, pp. 827–863.

McLaughlin, R.L., Carl, L.M., Middel, T., Ross, M., Noakes, D.L., Hayes, D.B., Baylis, J.R., 2001. Potentials and pitfalls of integrating data from diverse sources: lessons from a historical database for Great Lakes stream fishes. Fisheries 26, 14–23.

Meaden, G.J., 1996. Potential for geographical information systems (GIS) in fisheries management. In: Megrey, B.A., Moksness, E. (Eds.), Computers in Fisheries Research. Chapman & Hall, London, pp. 41–79.

Meixler, M.S., Bain, M.B., Galbreath, G.H., 1996. Aquatic GAP analysis: tool for watershed scale assessment of fluvial habitat and biodiversity. In: Proceedings of the Second International Symposium on Hydraulics and Habitats, International Association for Hydraulics Research. Proceedings of the Ecohydraulics 2000 Symposium, Quebec City, Canada.

Miller, R.R., Williams, J.D., Williams, J.E., 1989. Extinctions of North American fishes during the past century. Fisheries 14, 22–38.

Modde, T., Ford, R.C., Parsons, M.G., 1991. Use of habitat-based stream classification system for categorizing trout biomass. N. Am. J. Fish. Manage. 11, 305–311.

Nehlsen, W., Williams, J.E., Lichatowich, J.A., 1991. Pacific salmon at the crossroads: stocks at risk from California, Oregon, Idaho, and Washington. Fisheries 16, 4–21.

Noss, R.F., LaRoe III, E.T., Michael Scott, J., 1995. Endangered ecosystems of the United States: a preliminary assessment of loss and degradation. Biological Report 28. US Department of the Interior, National Biological Service, Washington, DC.

Oberdorff, T., Francois, G.J., Bernard, H., 1995. Global scale patterns of fish species richness in rivers. Ecography 18, 345–352.

Paller, M.H., 1995. Relationships among number of fish species sampled, reach length surveyed, and sampling effort in South Carolina coastal plain streams. N. Am. J. Fish. Manage. 12, 110–120.

Rathert, D., White, D., Sifneos, J.C., Hughes, R.M., 1999. Environmental correlates of species richness for native freshwater fish in Oregon, USA. J. Biogeo. 26, 257–273.

Ricciardi, A., Rasmussen, J.B., 1999. Extinction rates of North American freshwater fauna. Conserv. Biol. 13, 1220–1222.

Richards, C., Johnson, L.B., Host, G.E., 1996. Landscape-scale influences on stream habitats and biota. Can. J. Fish. Aquat. Sci. 53 (Suppl. 1), 295–311.

Statistical Analysis Software Institute (SAS), 1998. JMP'IN Statistical Discovery Software, Version 3.2. Duxbury Press, Cary, NC.

Schlosser, I.J., 1991. Stream fish ecology: a landscape perspective. BioScience 41, 704–712.

Scott, J.M., Davis, F., Csuti, B., Noss, R., Butterfield, B., Groves, C., Anderson, H., Caicco, S., D'erchia, F., Edwards Jr., T.C., Ulliman, J., Wright, R.G., 1993. GAP analysis: a geographic approach to protection of biological diversity. Wild. Mon. 123, 1–41.

Sheldon, A.L., 1968. Species diversity and longitudinal succession in stream fishes. Ecology 49, 193–198.

Smith, C.L., 1985. The inland fishes of New York State. New York State Department of Environmental Conservation, Albany, NY.

Stauffer Jr., J.R., Boltz, J.M., White, L.R., 1995. The Fishes of West Virginia. The Academy of Natural Sciences, Philadelphia, PA.

Stefan, H.G., Hondzo, M., Eaton, J.G., McCormick, J.H., 1995. Validation of a fish habitat model for lakes. Ecol. Mod. 82, 211–224.

Strahler, A.N., 1952. Hypsometric (area-altitude) analysis of erosional topography. Bull. Geol. Soc. Am. 63, 1117–1142.

The Nature Conservancy (TNC), 1996. Priorities for conservation: 1996 annual report card for US plant and animal species. NatureServe.

Trautman, M.B., 1981. The Fishes of Ohio with Illustrated Keys, Revised Edition. Ohio State University Press, Columbus, OH.

US Geological Survey (USGS), 1999. National elevation dataset for Pennsylvania. EROS Data Center, Sioux Falls, SD.

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

Whittier, T.R., Hughes, R.M., Larsen, D.P., 1988. Correspondence between ecoregions and spatial patterns in stream ecosystems in Oregon. Can. J. Fish. Aquat. Sci. 45, 1264–1278.

Wiens, J.A., 1989. Spatial scaling in ecology. Funct. Ecol. 3, 385–397.

Wiley, M.J., Kohler, S.L., Seelbach, P.W., 1997. Reconciling landscape and local views of aquatic communities: lessons from Michigan trout streams. Freshw. Biol. 37, 133–148.

Williams, J.E., Johnson, J.E., Hendrickson, D.A., Contretas-Balderas, S., Williams, J.D., Nawarro-Mendoza, M., McAllister, D.E., Deacon, J.E., 1989. Fishes of North America endangered, threatened, or of special concern. Fisheries 14, 2–20.