Classifying Lotic Systems for Conservation: Methods and Results of the Pennsylvania Aquatic Community Classification

Mary C Walsh, Jeremy Deeds, and Betsy Nightingale

June 2007

Pennsylvania Natural Heritage Program
Western Pennsylvania Conservancy

Pennsylvania Natural Heritage Program is a partnership of:

Western Pennsylvania Conservancy, Pennsylvania Department of Conservation and Natural Resources, Pennsylvania Fish and Boat Commission, and Pennsylvania Game Commission.

The project was funded by:

Beneficia Foundation
Heinz Endowments
Pennsylvania Department of Conservation and Natural Resources
Pennsylvania Department of Environmental Protection
Pennsylvania Department of Transportation
William Penn Foundation

Suggested report citation:

Walsh, M.C., J. Deeds, and B. Nightingale. 2007. Classifying Lotic Systems for Conservation: Methods and Results of the Pennsylvania Aquatic Community Classification. Pennsylvania Natural Heritage Program, Western Pennsylvania Conservancy, Middletown, PA, and Pittsburgh, PA.

Cover photos by:

PA Natural Heritage Program and M.C. Barnhart

Acknowledgements

Our sincerest thanks are given to the agency and foundation funding sources of the Pennsylvania Aquatic Community Classification Project.

Numerous individuals have contributed data, time, effort, and support to ensure the success of the project. They include:
Dave Argent, Mike Bilger, Charles Bier, Clare Billet, Greg Czarnecki, Emily Bond, Jacqui Bonomo, Bill Botts, Mike Boyer, Mark Bryer, Robin Brightbill, Frank Borsuk, Susan Buda, Meghan Carfioli, Ann Cavanaugh, Brian Chalfant, Pete Cinotto, Rob Criswell, Tony Davis, Dave Day, Charles Decurtis, Alice Doolittle, Marsha Dulaney, Jane Earle, Ryan Evans, William Fairchild, Su Fanok, Jean Fike, Jon Gelhaus, Chuck Hawkins, Alan Herlihy, Jonathan Higgins, Jen Hoffman, Dave Homans, Greg Hoover, Terry Hough, John Jackson, Nels Johnson, Sally Just, Kevin Kelley, Rod Kime, Andy Klinger, Susan Klugman, Michelle Laudenslager, Bill Lellis, Betsy Leppo, Owen McDonough, Charlie McGarrell, Kathy McKenna, Andrew Nevin, Ted Nutall, Arlene Olivero, Maggie Passmore, Greg Pond, Greg Podniesinski, Derek Price, Sally Ray, Dave Rebuck, Lou Reynolds, Bob Ross, Erika Schoen, Tamara Smith, Tammy Smith, Tony Shaw, Andy Shiels, Bob Schott, Judy Soule, Scott Sowa, Rick Spear, Ron Stanley, David Strayer, Sue Thompson, Chris Urban, John Van Sickle, Joel Van Noord, Rita Villela, Jeff Wagner, Cindy Walters, Carrie Wengert, Kellie Westervelt, Jeremy Yoder, and Chen Young.

Many institutions generously made data available to us. We would like to acknowledge the following:
Academy of Natural Sciences, Aquatic Systems, Alliance for Aquatic Resource Monitoring, Brodhead Watershed Association, Chester County Water Authority, ClearWater Conservancy, Civil and Environmental Consultants, Cornell University Museum, Delaware Museum of Natural History, Delaware River Basin Commission, Delaware Riverkeeper, Dinkins Biological Consulting, Ecological Specialists, Enviroscience, Inc., Florida Museum of Natural History, Harvard University Museum of Comparative Zoology, Illinois Natural History Survey, Michigan State University, Monroe County, National Park Service, New Jersey Heritage Program, New York Department of Conservation, New York State Museum, Ohio State University, Ohio River Sanitation Commission, Pennsylvania Department of Environmental Protection, Pennsylvania Department of Transportation, Pennsylvania Fish and Boat Commission, Pennsylvania Natural Heritage Program, Pennsylvania State University, Philadelphia Water Department, Pike County Conservation District, PPL Generation, Smithsonian Institute, Susquehanna River Basin Commission, The Nature Conservancy, US Environmental Protection Agency, US Forest Service – Allegheny National Forest, US Geological Survey, University of Michigan, and Western Pennsylvania Conservancy.

We regret any inadvertent omissions of project contributors from this list.

Executive Summary

In the course of the Pennsylvania Aquatic Community Classification Project, extensive effort was spent to determine the best approach to classifying the flowing waters of Pennsylvania. Varied types of classification systems and methods for developing them have been applied in other regions. We ultimately classified streams based on community assemblages of macroinvertebrates, mussels, and fish and based on physical stream types, with the intention of describing biodiversity patterns and habitat gradients.

Because we developed the classifications with datasets acquired from state and regional monitoring projects, the project resources were focused on analysis and applications, instead of data collection. Development of a project database with comprehensive aquatic datasets enabled a large, regional analysis of existing community survey data. Multivariate ordination and cluster analysis were used to determine initial community groups. Indicator Species Analysis, classification strength and review by taxa experts helped to refine community types. Lastly, community groupings were evaluated with a validation analysis of a secondary dataset. We compared taxonomic level for grouping macroinvertebrate communities with genus- and family-level datasets. Final community groupings include 13 mussel communities, 11 fish communities, 12 communities of genus-taxonomy macroinvertebrate communities, and 8 family-taxonomy macroinvertebrate communities. Seasonal influences on macroinvertebrate abundance and basin specificity of fish and mussels were used to modify classifications. Datasets within a spring index period were used to classify macroinvertebrates. Three separate basin classifications were necessary to describe mussel communities, while two separate basin classifications were applied to fish communities.

Water chemistry, stream channel and watershed data were attributed to stream reaches, reach watersheds, and catchments and were used to describe communities. We combined classes of bedrock geology, stream gradient, and watershed size in into physical stream types for each reach in the study area. Models were developed to predict community presence based on channel and watershed attributes for all mussel, fish, and macroinvertebrate communities.

We analyzed the condition of streams and watersheds to better understand relationships between communities and stream quality and to prioritize areas for restoration and conservation. Least Disturbed Streams were designated as those having little human disturbance; we used watershed and riparian landcover, mines and points sources, road – stream crossings, and dams as disturbance indicators. Watershed conservation and restoration priorities met criteria for the density of Least Disturbed Streams, community habitats, and community quality metrics.

By systematically evaluating communities, habitats, and conditions across the waterways of Pennsylvania, we have gained better understanding of the aquatic natural diversity and its threats. Many ACC project applications are currently underway, including conservation planning and watershed management.

Table of Contents

List of Tables and Figures

1. Project Introduction

To create a systematic categorization of flowing water ecosystems in Pennsylvania and its watersheds, the Pennsylvania Aquatic Community Classification (ACC) was developed by the Pennsylvania Natural Heritage Program. The ACC defines types of stream and river reaches based on aquatic communities, their habitats, and watershed properties. The project products were designed for natural resource applications including assessment, monitoring, resource planning, and conservation.

In this project, aquatic assemblage types and habitat types, their distribution, and relationship to water quality were described. Relative water quality and habitat conditions were evaluated for aquatic assemblages. Potential habitat types for communities were modeled and gave further insights to the relative importance of environmental characteristics in defining community habitats. High quality or rare communities and habitats were used in a watershed conservation prioritization analysis. Similarly, communities and habitats in poor condition were described for prioritizing watershed restoration.

Since its initiation in 2001, the ACC project was guided by three major objectives: 1) to develop a region-wide classification of riverine systems as a basis for conserving aquatic biodiversity, 2) to determine aquatic environments and assemblages in the greatest need of conservation and protection, and 3) to apply the classification system to natural resource management and conservation planning. The steps to develop the ACC outlined in this report include data mining, managing the project database, evaluating data types, developing methods to classify aquatic assemblages and habitats, and analyzing the condition of stream reaches. We also analyzed watersheds based on conservation value and restoration need. In this document the institutional knowledge gained from the project is shared with other natural resource agencies and organizations; we discuss the project approach, methods, analyses, lessons learned, and information gained about aquatic resources.

A system for managing aquatic communities and their habitats

Using ecological community units as the basis for conservation and management is not a new concept for resource managers. Mapping vegetation communities came into popularity in the last half of the 20th century and has been largely embraced as a tool for land management by agencies and conservation organizations like the National Park Service, the US Forest Service, the US Department of Defense, and The Nature Conservancy (TNC).

What classifications exist for managing plant communities?

As a primer to national vegetation types, the *U.S. National Vegetation Classification* (Grossman et al. 1998) developed a standard for classifying vegetation stands and has been used in a hierarchy for further delineation of vegetation types at regional and sub-regional scales (e.g. Faber-Langendoen 2001). Classes of vegetation in Pennsylvania were described in *Terrestrial and Palustrine Communities of Pennsylvania* (Fike 1999). Surveys of rare vegetation community types are currently documented in the Pennsylvania Natural Diversity Inventory Database, which records locations of rare organisms and communities for the Commonwealth.

Conservation across aquatic ecosystems would be best guided by a uniform classification system. Without such a system, protection and management decisions are made without reference to their ecological context; furthermore, information sharing across agencies without a common classification is difficult because of a lack of common ecological units (McMahon et al. 2001). Information from conservation programs, monitoring, and inventories cannot be easily compared across jurisdictional units, such as national parks, state land holdings, and other agency units without a common classification (Bryer et al. 2000). Using a classification system, similar ecological units can be assessed within and across political or agency jurisdictional boundaries.

State, regional, and national conservation initiatives recognize the need for comprehensive aquatic habitat information. Objectives in the Pennsylvania Comprehensive Wildlife Conservation Strategy (2005) acknowledge the gap in systematic habitat protection and recommend development of standardized habitat classification under Operational Objective 2.2.1.: "*Develop a standardized community/habitat classification system that works at both vertebrate and invertebrate scales.*" The Pennsylvania Department of Conservation and Natural Resource's Biodiversity Workgroup Report (2001) and State Forest Resource Management Plan (2005) identified classification of aquatic communities as a priority for conservation of biodiversity and natural resources for the agency. In recognition of habitat conservation needs across the entire United States, the National Fish Habitat Action Plan (2006) has begun developing hierarchical aquatic habitat classes at a national scale.

We anticipate incorporating the results of the ACC into hierarchical aquatic habitat classes from the forthcoming regional and national habitat classifications. To our knowledge, the ACC is the first aquatic classification effort of this magnitude for Pennsylvania.

Why do scientists recommend that we classify our ecosystems?

Researchers and conservationists, identifying flaws in current aquatic management methods, have looked to classifications to aid resource protection and management, create common ecological units, develop standard terms for communication, and allow collaboration across the scientific and conservation communities (Davis and Henderson 1978; Platts 1980; Lotspeich and Platts 1982; Higgins et al. 2005). A framework for aquatic resource planning and management ought to be based on a system that includes multiple species, that is focused on habitats, and that is linked to ecological and watershed processes (Maybury 1999; Higgins et al. 2005).

Comparing streams and rivers to ecologically similar waters has particular relevance when developing standards for water quality regulations and conducting aquatic research studies. Assessing waterways with biological surveys, as popularized in state and federal agency biomonitoring protocols (e.g., Barbour et al. 1999), necessitates having reasonably-correct expectations about the condition of unimpaired rivers relative to those most disturbed by human alteration. Minimizing natural variation within river – to – river comparisons would facilitate assessments and strengthen protection measures based on biological assessments (Herlihy et. al 2006). Other disciplines (e.g., fisheries science, benthic ecology, water quality, groundwater management, watershed management, conservation planning, and restoration ecology) would also benefit from knowledge of the diversity and the comparability of aquatic riverine systems.

Project collaboration

Significant benefits to this project were gained from the expertise and data contributed by the many project collaborators. A committee of project advisors consisted of biologists, agency representatives, natural resource planners, and conservationists from universities, state and federal agencies, inter-government river basin commissions, natural heritage, conservation, and watershed organizations. Five major advisory meetings held since 2001 brought together approximately 25 project advisors who provided scientific guidance on the project methods and feedback on preliminary results. A pilot study completed in 2004 (see Nightingale et al. 2004) was extensively reviewed by project advisory group members; lessons learned from the pilot study were applied to analyses reported here.

Review of specific project results was sought from state and regional aquatic ecologists and taxonomic specialists. Researchers from the Utah State University, Pennsylvania State University, Oregon State University, Smithsonian Institution, and others provided expertise on aquatic classification topics. Additionally, collaboration with natural resource and Pennsylvania regulatory agencies like the Department of Environmental Protection (DEP), the Department of Conservation and Natural Resources (DCNR), and the Fish and Boat Commission (FBC) facilitated discussion on project methods and applications of ACC products.

Two land conservancies in Pennsylvania, The Nature Conservancy and the Western Pennsylvania Conservancy, maintained funding and staff positions within the Pennsylvania Natural Heritage Program for this project. Conservation planning efforts at both organizations have already incorporated ACC project products.

Project methodology

To complete the Pennsylvania Aquatic Community Classification Project, a series of major steps was undertaken:

- Developing a study approach;
- Mining and managing data;
- Creating biological classifications;
- Associating environmental data with communities and developing a physical stream classification;
- Evaluating and refining biological classifications;
- Modeling community habitats;
- Identifying high quality streams and watersheds;
- Selecting poor quality watersheds for restoration prioritization.

2. Classification Approach

The desired applications of a classification system should dictate its approach. For this reason, we reviewed a number of classification systems in light of the goals of the Pennsylvania Aquatic Community Classification.

The Pennsylvania Aquatic Community Classification was intended to create a classification system founded in ecological patterns. It stratifies stream reaches based on aquatic animal communities. Communities are defined as recurring assemblages of organisms found together and that respond to similar environmental factors. A physical stream classification, using watershed and reach attributes, was also developed; this classification used a similar approach to "macrohabitat" classifications developed by The Nature Conservancy (Higgins et al. 2005). The physical classification defines ecological gradients in geology, stream slope, and stream size that related to biodiversity patterns.

Review of classification types

Considering the diverse types of classifications, our primary questions were:

- *What are the benefits and drawbacks of each type of aquatic classifications?*
- *What type is most appropriate for our goals?*

A review of those topics is briefly summarized here.

Current scientific knowledge about flowing waters recognizes that spatial patterns structure aquatic ecosystems at different scales and that processes at multiple spatial scales may have an influence on aquatic biodiversity. Regional patterns of biogeography, climate, drainage patterns, and landforms influence physical aquatic systems and biological patterns (Frisell et al. 1986; Maxwell et al. 1995; Omernik 1995; Oswood et al. 2000; and others). Other factors that occur on a stream segment scale, such as hydrology, temperature, channel morphology, and water chemistry, are related to communities and species distributions and have be used in stream classifications (Reash and Berra 1987; Poff and Alan 1995; Richards et al. 1997; USGS 1998; Wehrly et al. 1998; and others). Reach scale- and micro-habitats also explain distributions of aquatic species and assemblages (Blanck et al. 2007, Usio 2007, Haag and Warren 2007) and are likely related to life history traits and habitat preferences. Because of the diversity of factors identified as stratifying aquatic environments, a number of variables have been applied in aquatic classification strategies.

Ecoregional classifications, such as ecoregions and physiographic provinces, are commonly used to stratify habitats in aquatic classifications (Lotspeich and Platts 1982; Hudson et al. 1992; Hughes 1995; Maxwell et al. 1995). Flowing waters within ecoregion types may have similar climate, vegetation, geology, and soils, resulting in comparable aquatic habitat characteristics like water quality, stream substrates, and channel characteristics (Omernik 1995; Omernik and Bailey 1997). Ecoregion classifications are commonly used to group aquatic environments into similar types (Griffith et al. 1999) for applications like biomonitoring.

Studies that examine coarse-scale landforms find that they perform poorly as the sole classifiers of aquatic assemblages. Ecoregions and physiographic boundaries alone do not classify aquatic habitats as well as biological data do (McCormick et al. 2000; Waite et al. 2000; Sandin and Johnson 2000; Hawkins and Vinson 2000; Herlihy et al. 2006, and others). Identifying the ecoregion that best represents a flowing water body may be difficult because watersheds often cross more than one landform type and the relative influence of each type is unknown. The Susquehanna River, for instance, crosses seven Level-4 Omernik Ecoregions from its confluence at the North and West Branches to the mouth at the Chesapeake Bay. Nevertheless, ecoregions may be useful to standardize aquatic units at a large scale in a hierarchical classification with nested sub-units (Omernik and Bailey 1997; Griffith et al. 1999). A fine-scale classification that sub-divides ecoregion types into smaller units would be more useful for delineating aquatic habitats than landform classes alone (Lyons 1989; Heino et al. 2002).

Classifications that incorporate hydrological processes, geomorphology, and physical habitats have also been proposed (Kellerhalls and Church 1989; Rosgen 1994). These types of classifications have not been widely applied by biologists, perhaps because of the detailed information gathering and mapping necessary for regional application. Physical descriptors of channel segments have been the basis for some classification studies. Aquatic ecological systems (grouping of watersheds), and sub-classes of "macrohabitat" stream reaches are used by The Nature Conservancy for classifying stream types and guiding conservation (Higgins et al. 2005). Physically similar systems are grouped by a multivariate analysis of gradient, elevation, stream size, stream connectivity, geology and hydrologic regime (Higgins et al. 2005). The National Fish Habitat Action Plan (http://www.fishhabitat.org/) has adopted similar methods for discriminating aquatic habitats at a national scale.

Aquatic GAP analysis programs have also begun to converge on an ecological classification construct based on channel and watershed characteristics, such as bedrock and surficial geology, soils, climate, gradient, and sinuosity. Channel types are then grouped based on multivariate procedures to identify similar adjacent stream reaches, called "valley segments". Stream classifications using these methods were completed in New York, Michigan, Wisconsin, Illinois, and Missouri (USGS 2003; Sowa et al. 2005; McKenna et al. 2006; Seelbach et al. 2006).

A bottom-up riverine classification of environmental characters applied in a valley segment classification can be used to infer biological gradients from physical habitats. Valley segment classes were related to macroinvertebrate diversity, fish abundance, and fish spawning habitat in some studies (Brosse et al. 2001; Baxter and Houser 2000). A valley segment classification can provide the appropriate context for understanding species habitat characteristics. The valley segment classes defined the geomorphological and groundwater characteristics of bull trout (*Salvelinus confluentus*) spawning habitats in one study (Baxter and Houser 2000).

Some classification systems integrate biological classification with physical habitat types. Using biological data to stratify habitat classifications incorporates gradients that shape assemblages or species habitat. In aquatic GAP analysis protocols, habitat classes are refined (e.g., Sowa et al. 2005) by assemblages ranges; the assemblages assist in defining valley segments.

Aquatic assemblages have inherent properties that make them useful for classification and conservation. Assemblages respond to ecological changes in the flowing water environment related to resource availability (Vannote et al. 1980); thus, they are sentinels for ecological gradients. The value of ecological services provided by communities (Costanza et al. 1997) makes understanding communities particularly relevant to natural resource management. Biological classification in aquatic habitats has been suggested as a method for fine-tuning water quality assessment criteria, benchmarks for rare species listing, and for species habitat assessment (Herlihy et al. 2006). Additionally, communities are valuable targets for conservation since they encompass biodiversity and habitat more broadly than single species (Stein and Davis 2000). Many science and conservation organizations recognize the conservation need for communities because of inherent natural, economic, or societal value. For instance, conservation planning implemented by The Nature Conservancy incorporates terrestrial communities and their supporting habitats as conservation priorities (Groves et al. 2000).

Desired outcomes

Ultimately, we chose a community classification and a "macrohabitat" type approach to describe physical and biological diversity in Pennsylvania.

We anticipate that project outcomes will lead to more effective conservation of aquatic natural resources.

Project outcomes

- Develop a classification system based on patterns in communities;
- Understand aquatic animal biodiversity and its relationship to landscape and local habitat factors;
- Develop a physical stream classification and relate it to communities;
- Centralize aquatic data for Pennsylvania in a public database;
- Identify potentially high quality stream reaches and watersheds as conservation priorities;
- Identify poor quality watersheds as restoration priorities.

3. Data Management

The community types were developed from existing datasets for reasons discussed in this chapter.

Study area

The study area was chosen to encompass Pennsylvania and its contributing watersheds. The study area includes the entire Delaware, Susquehanna, Allegheny, and Monongahela River basins, and parts of the Erie, Genessee, Potomac, and Ohio River Basins (Figure 3-1).

Prior studies had not examined community types on a similar scale in the study region. The quantity of biological data and the geographic scale of the Pennsylvania Aquatic Community Classification differentiate this study from others. Datasets for the project analysis had a broad geographic scope (e.g., a river basin) and together contained thousands of biological community records of mussel, fish, and macroinvertebrate surveys.

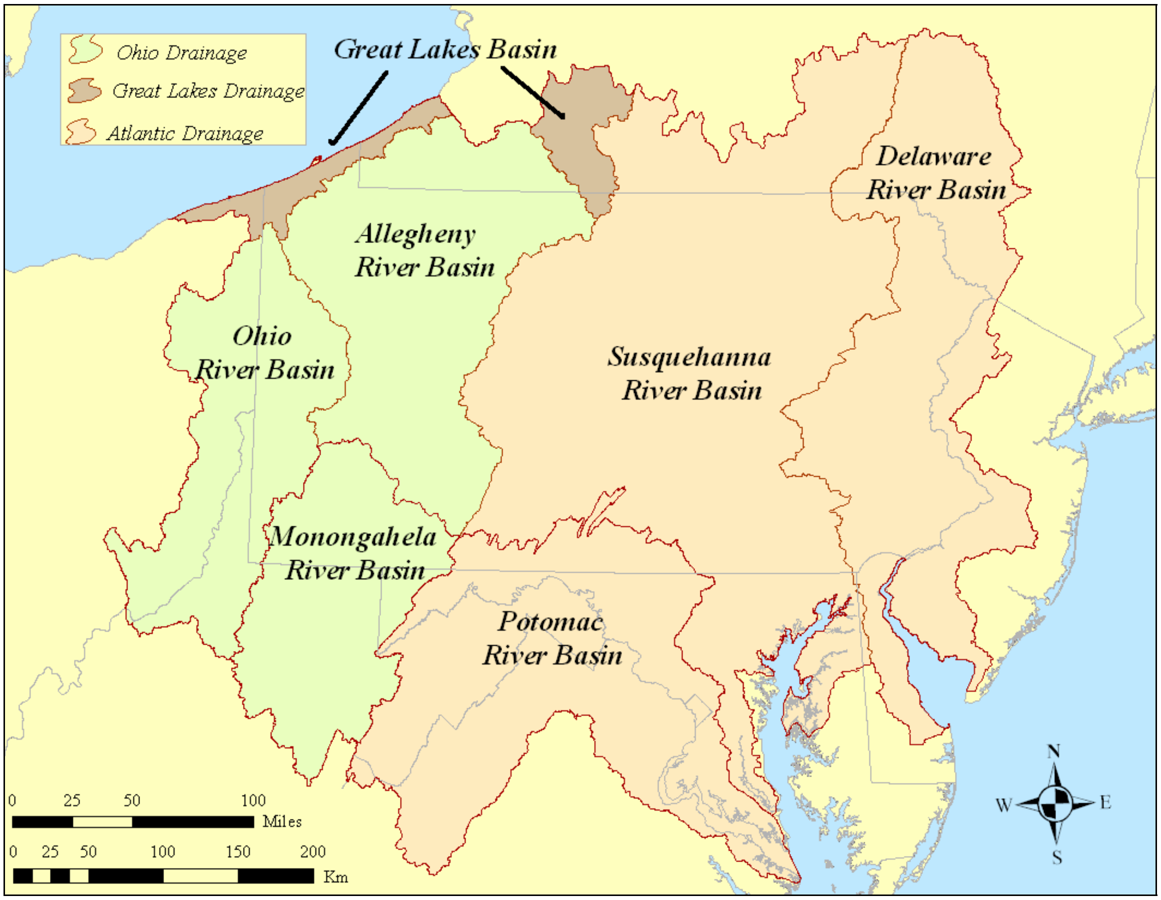


Figure 3-1. The ACC study area included the major drainage basins in Pennsylvania flowing to the Atlantic Ocean, the Ohio River, and the Great Lakes. The Ohio River drainage encompasses its tributaries – Allegheny and Monongahela Rivers. The Atlantic drainage contains the Potomac, Susquehanna, and Delaware Rivers.

Data gathering

The benefits and drawbacks of collecting a dataset of biological, water quality, watershed, and habitat information over a large region were weighed. Such a dataset collected with methods tailored to the project and with a randomized design clearly has advantages like standardized methods and field/lab data collection. However, to complete an extensive field collection involved more resources than available to project staff. Using existing field-collected and regional landscape datasets (e.g., GIS data) focused project effort on data analysis.

Collecting and formatting data

A number of organizations and institutions were surveyed for available and applicable datasets. Organizational programs with long-term datasets or those spanning large geographic areas were targeted for data requests, such as the US EPA Environmental Monitoring and Protection Program and PA DEP Water Quality Network datasets. Additional requests for electronic data were made to other state and federal agencies, river basin commissions, academic researchers, watershed groups, museums, water authorities, and county conservation districts.

Data in electronic format largely fulfilled the need for data geographically representing the study area. In some parts of the study area, electronic datasets were lacking and aquatic studies in hard-copy reports were obtained. Data reports in print from DEP and the DCNR Wild Resource Conservation Program were transcribed into electronic format and study locations were mapped in a GIS (ESRI ArcMap 9.1®). In total, 94 paper and electronic datasets from 44 organizations and agencies were obtained.

We initially invested much time in developing a centralized database to organize the project data. The resulting project database, Pennsylvania Aquatic Database (PAD), contains data for public distribution and includes most datasets used in the project analysis. The PAD runs in a Microsoft Access® platform (Microsoft Office 2000). The model for the database was the Ecological Data Application System® (v.3) developed by TetraTech, Inc. (http://www.ttwater.com/Ecological_details.htm). The database was originally designed to store fish, benthic macroinvertebrate, algae, chemistry, physical character, and habitat data and we modified it to accommodate mussel survey data.

Pennsylvania Aquatic Database

Stores information including:

- Biological, chemical, physical habitat samples;
- Survey locations;
- Survey methods and sampling gear;
- Data source contact information;
- Taxa lists.

Standardized taxa lists were created for the Pennsylvania Aquatic Community Classification study. State lists for mussels, fishes, and some macroinvertebrate groups are maintained by the Pennsylvania Natural Heritage Program and the Pennsylvania Biological Survey. Some invertebrates had inconsistent nomenclature and less documented ranges. Where possible, recommendations from taxonomic experts on range and nomenclature were used to establish taxa lists for the database. Experts provided information on the following invertebrate taxa: Ephemeroptera, Trichoptera, Plecoptera, Odonata, Tipulidae, Simuliidae, Culicidae, Syrphidae, Amphipoda, Isopoda, and Decapoda (Nightingale et al. 2004). Datasets with outdated taxa names or taxa that have distribution known to be outside the study were updated in the database.

4. Data Screening

To characterize aquatic assemblages, the availability of appropriate data and their geographic ranges were considered. We chose three types of taxa to develop separate biological classifications: macroinvertebrates, fishes, and mussels. Each type of taxon occupies niches different from the others and responds to environmental gradients uniquely. Datasets with geographic coverage across the study area in varied habitats and watersheds, those that were community surveys, and those with a high degree of taxonomic certainty were selected (Table 4-1).

Macroinvertebrates
- Are sensitive to water quality;
- Are sampled in wadeable streams and selectively in non-wadeable habitats;
- Occur in stream benthic habitats;
- Have limited habitat range.

Fishes
- Are sampled in wadeable and non-wadeable habitats;
- Are thermally sensitive;
- Occupy a range of food niches, but are top predators in many aquatic systems;
- Are relatively mobile.

Mussels
- Are sampled in wadeable and non-wadeable habitats;
- Occur in 3rd order and larger streams;
- Are sensitive to toxins and habitat alteration;
- Are filter feeders and live in benthic habitats;

We extensively screened datasets to evaluate their appropriateness for analysis. The effect of data collection time, data collector, and collection method on taxa composition and classification was evaluated. Additionally, the influence of rare taxa and exotic taxa on classification statistical results was also examined. Taxa richness and taxa occurrence by Julian day, month, season, sampling method, and data collector were evaluated. Macroinvertebrate, fish, and mussel samples were analyzed with multivariate ordination (non-metric multi-dimensional scaling) for patterns related to the same variables.

Seasonal and temporal patterns

Analysis results revealed that macroinvertebrate occurrence is strongly related to seasons. The relative abundance of the winter stoneflies (Capniidae) demonstrates the pattern of taxonomic shift in macroinvertebrate samples by season (Figure 4-1). The Capniidae stoneflies are abundant in winter samples, coincidental with their maturation into the final larval stage before hatching in cold months. Because of seasonal taxa shifts in larvae, many stream surveys are conducted within an index period.

Figure 4-1. Monthly mean index of relative abundance and standard deviation of Capniidae stoneflies. Relative abundance is significantly different across months (ANOVA, F = 228.71, p < 0.001). Error bars represent standard deviation.

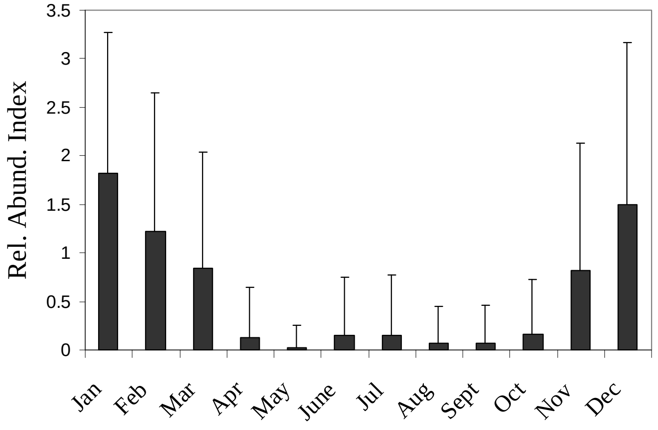


Table 4-1 (a-d). Datasets used to develop community classifications were compiled from a number of data sources for a) fish, b) macroinvertebrates identified to family taxonomy (MI – Fam), c) macroinvertebrates identified to genus taxonomy (MI – Genus), d) mussels.

Dataset	Data sources
a) Fish	Ohio River Sanitation Commission US Geological Survey PA Natural Heritage Program, Western Pennsylvania Conservancy US Environmental Protection Agency PA Fish and Boat Commission Pennsylvania State University PA Department of Environmental Protection Philadelphia Water Department US Forest Service – Allegheny National Forest NY Department of Environmental Conservation US Army Corps of Engineers
b) MI – Family	PA DEP In-stream Comprehensive Evaluation Program
c) MI – Genus	Ohio River Sanitation Commission Susquehanna River Basin Commission US Environmental Protection Agency PA Natural Heritage Program, Western Pennsylvania Conservancy PA Department of Environmental Protection US Forest Service – Allegheny National Forest
d) Mussels	Aquatic Systems, Inc. Civil & Environmental Consultants, Inc. Dinkins Biological Consulting Enviroscience, Inc. Western Pennsylvania Conservancy US Geological Survey Wildlife Resource Conservation Fund, PA Department of Conservation and Natural Resources Ecological Specialists New York State Museum

We examined classifications for macroinvertebrate fauna within index periods. Indices of how well classifications portioned data were applied to evaluate the most appropriate index period. Classification strength[1], indicator values (IV), and mean p-values derived from Monte-Carlo simulations in Indicator Species Analysis (ISA)[1] were compared for classifications developed for 1) all seasons, 2) spring (April − June), 3) summer − fall (July − October), and 4) winter (November − March).

As an index of the ability to parcel data, classification strength compares within-group variability to between-group variability. Higher classification strength indicates a classification is better at portioning data than a classification with low classification strength. Classification strengths for summer − fall sampling and all-season sampling were the weakest, but were the strongest for spring and winter (Table 4-2).

Indicator values were slightly higher for the winter index period than for the spring index period (Table 4-2). Relatively high indicator values and low mean Indicator Species Analysis p-values[1] suggest that a classification's indicator taxa are strongly associated with the community groups Because spring sample index periods are commonly used by Pennsylvania agencies, we concluded that a classification of communities from that time period would be most appropriate. While the number of sampling events for fish and mussels was concentrated in warmer months, strong seasonal patterns in those taxa abundances were not found. Classifications of seasonal index periods produced similar results. We included data from all seasons in the final classifications of fish and mussels.

Table 4-2. Classification strength (CS), indicator values (IV), and Indicator Species Analysis (ISA) Monte-Carlo simulation p-values for macroinvertebrate classifications of all sampling periods, for spring (April − June), summer − fall (July − October), and winter (November − March).

	All Seasons	Spring	Summer – Fall	Winter
CS	0.14	0.18	0.14	0.18
IV	12.20	12.84	10.05	14.15
ISA p-value	0.05	0.18	0.14	0.22

---

[1] See Appendix 1 for details on classification strength and Indicator Species Analysis.

Influence of data collector and sampling methods

Among the datasets chosen for community analysis, sampling method and data collector did not greatly influence the grouping of sites with multivariate ordination and clustering analysis. For instance, among 6,698 fish samples with 34 variations of sampling gear and collection methods, no discernable patterns were identified in the site-taxa ordination.

Analysis of other taxa groups yielded similar results. Although some influence of sample methods and effort on taxa composition was expected, the number of samples may overwhelm the appearance of such patterns. However, we feel that community composition was adequately represented in the datasets chosen for analysis and that shifts in taxa due to sample efforts and methods were relatively minor.

Datasets and collection methods

Fish
Collecting gear – boat electrofishing, backpack electrofishing, seine, trapnet, gillnet, rotenone;

Macroinvertebrates
Collecting gear – D-frame nets, kicknets, Surber samplers, artificial substrate samplers;
Laboratory and field identification;
Family and genus identification level;
100-300 count sub-samples;

Mussels
Collecting gear – mussel buckets, snorkel, SCUBA;
Qualitative surveys, timed search surveys;
Quantitative surveys within fixed areas;
Semi-quantitative transect surveys.

Taxonomic resolution of macroinvertebrate datasets

Datasets with varied levels of taxonomic identification were evaluated for this project. Widely accepted macroinvertebrate protocols (e.g., Barbour et al. 1999) recommend that macroinvertebrates data for stream health assessment use genus- or species-level data. However, the most comprehensive macroinvertebrate dataset obtained by the ACC had data with family-level identifications.

Because the debate among scientists about the appropriate level of taxonomy of macroinvertebrates for classifying patterns in flowing waters has yet to be resolved, two levels of taxonomy were compared in this study. One dataset with family-level data collected by the Pennsylvania Department of Environmental Protection for the In-stream Comprehensive Evaluation program was compared to other datasets with genus-level data. Comparisons of macroinvertebrate taxonomic level in community assemblages are made in Chapter 5.

Dataset refinement

Rare taxa may unduly influence multivariate analysis by adding excess variation (McCune and Grace 2002). Other similar studies have removed rare species from community analysis for this reason (e.g., Herlihy et al. 2006). For the purposes of this study, rare species were defined as those present at <1% of study locations.

The influence of rare species data was evaluated for each of the macroinvertebrate, fish, and mussel datasets, and rare species were removed on a case-by-case basis. For the mussel dataset, which had relatively few taxa, removal of rare species was not a viable option for the analyses. Without rare taxa, no successful NMS ordination solution could be created (See Chapter 5); thus, rare species were not removed.

Generally, samples in the fish, mussel, and macroinvertebrate datasets were not collected in a uniform manner. Density or relative abundance was either not available or not comparable between datasets (Table 4-3). The determination to use presence-absence for most data analysis was based on the aforementioned reasons. The exception was the family macroinvertebrate dataset because data collection, sub-sampling, and identification were uniform. Because patterns in biological communities are detectable for presence-absence information in large-scale studies of diverse communities (Gauch 1982), we felt it was appropriate to use presence-absence data.

Exotic and stocked taxa

Although native aquatic communities would be ideal baselines for assessing communities, several issues prevented analysis of native-species-only assemblages for this project. The transplantation of aquatic species from other continents and basins has been a common practice for several centuries in Pennsylvania basins, making assessments of native communities difficult. For some organisms, like some macroinvertebrate and fish taxa, species native ranges have not been thoroughly documented in the study area. Current assemblages are likely influenced by non-native taxa; in most instances non-native species were included in the community analyses.

Where taxa surveyed in community datasets are not permanent community members, we attempted to remove them from the analysis. In many cases, non-native species (e.g., brown trout (*Salmo trutta*) from Europe and rainbow trout (*Oncorhynchus mykiss*) from western North America) have become naturalized and are captured in fish community surveys. We identified temporary community members in the case of stocked rainbow and brown trout in a put-and-take fishery, where the stocked fish do not become permanently established in the assemblage. In stocked streams that were not designated as cold-water fisheries (defined as having wild-trout reproduction by the Pennsylvania Fish and Boat. Commission) brown trout and rainbow trout were removed from the dataset.

Cool- and warm-water game fish like the muskellunge (*Esox masquinongy*), walleye (*Stizostedion vitreus*), channel catfish (*Ictalurus punctatus*), largemouth bass (*Micropterus salmoides*), smallmouth bass (*Micropterus dolomieu*), yellow perch (*Perca flavescens*), and bluegill (*Lepomis macrochirus*) are also stocked regularly across the state and were included in the community datasets. However, stocked cool- and warm-water stocked species are thought to establish natural reproduction in many locations after stocking (according the PA Fish and Boat Commission, see http://sites.state.pa.us/PAExec/Fish_Boat/stockwarmc_prior.htm). For this reason, cool- and warm-water stocked fish species remain in the analysis.

Another non-native species, the Asian clam (*Corbicula fluminea*), was included in macroinvertebrate community analyses. The Asian clam has become established as part of the community in many study area waterways and was regularly sampled in macroinvertebrate surveys.

Why is the taxonomic level of macroinvertebrates appropriate for scientific studies under debate?

In many cases, aquatic insect species are not yet fully described by entomologists and often taxonomic keys are not available for larval forms typically collected in aquatic surveys. Scientists have weighed the costs and benefits of identifying macroinvertebrates to varying levels of taxonomy. The benefits gained by species identification include more detailed information related to each species about water pollution tolerance, habitat, and evolutionary history. However, the effort and expertise necessary to identify species, and even genera, may be unattainable for some aquatic projects.

A number of studies have provided evidence that a lower level of taxonomy is more suited for classifying and assessing stream health when compared to higher taxonomy. Lower levels of taxonomic resolution (e.g., genus and species) may be more appropriate for informing classifications; higher levels of taxonomy (e.g., family, order, and phylum) dilute patterns in environmental or biological gradients (Marchant 1990; Bowman and Bailey 1997; Lenat and Resh 2001). However, other studies find that similar classifications are produced when data with family-level taxonomy are classified relative to genus or species data (Marchant et al. 1995; Hewlett 2000).

Table 4-3. The number of samples, number of rare taxa removed, final number of taxa, and taxa abundance or presence-absence in fish, macroinvertebrate, and mussel community classification datasets. MI – Family = macroinvertebrates identified to family taxonomy. MI – Genus = macroinvertebrates identified to genus taxonomy.

Dataset	Basin	Total number samples	Number Rare Taxa Removed	Final Number of Taxa	Relative Abundance or Presence-Absence
Fish	Atlantic Basin	4284	60	80	Pres-Abs
	Ohio – Great Lakes Basins	2027	65	75	Pres-Abs
MI – Family		3261	---	63	Rel Abund
MI – Genus		863	163	138	Pres-Abs
Mussels	Delaware Basin	844	---	9	Pres-Abs
	Susquehanna – Potomac River Basins	145	---	14	Pres-Abs
	Ohio – Great Lakes Basins	170	---	32	Pres-Abs

Basins and zoogeographic patterns

The most appropriate geographic extent of classifications for macroinvertebrates, fish, and mussels was examined. For macroinvertebrates, there did not appear to be many basin-specific distributions of taxa in the genera and families. To date, the geographic distribution of many macroinvertebrates has not been well studied in Pennsylvania.

Fish and mussel classifications were influenced by zoogeographic characteristics. For both taxa, watersheds draining to the Atlantic Slope and those draining to the Ohio River Basin differ greatly in the faunal characteristics. Based on knowledge of species distributions and patterns in classification analyses, we developed classifications for basins or groups of basins. For instance, fish were grouped into two basin groups for the purpose of identifying community types: 1) watersheds on the Atlantic Slope, including the Delaware River, Susquehanna River, and Potomac River Basins, (hereafter called the Atlantic Basins), and 2) Allegheny River, Monongahela River, Ohio River, and Great Lakes Basins (hereafter referred to at the Ohio – Great Lakes Basins) (Figure 3.1). Patterns for mussel communities revealed that assemblage classifications of the following three basin groups produced the strongest results: 1) Allegheny River, Monongahela River, Ohio River Basin and the Great Lakes Basin (hereafter referred to at the Ohio – Great Lakes Basins), 2) Susquehanna River and Potomac River Basins, and 3) Delaware River Basin (Figure 3.1).

5. Community Classification Analysis

To initially develop community groups, we used two multivariate grouping methods: cluster analysis and ordination. Non-metric multi-dimensional scaling (NMS) ordination was used to evaluate the relationships between cluster groups in ordination space and refine cluster groups. We used Indicator Species Analysis and classification strength to identify the appropriate number of community groups.

Environmental data were associated with cluster groups. We described community habitats by mean water chemistry, in-stream habitat variables, and other stream reach and watershed variables.

Lastly, we used a variant of Classification and Regression Tree Analysis, called Random Forest Analysis, to predict community membership of stream reaches based on environmental data. Community groups were predicted for streams without biological samples. The analysis procedure was performed for each taxa group dataset (fish, mussels, and macroinvertebrates) and for the genus and family macroinvertebrate datasets. The community groups were developed by applying the data procedure to 70% of the datasets. Then, the remaining 30% was analyzed in a separate validation of the community groups.

Environmental data

We developed a number of datasets to describe community occurrences.

Water quality

Water chemistry was often measured at the same sampling locations where a biological sample was collected. However, in some cases, few water chemistry data were available. Dissolved oxygen, pH, alkalinity, conductivity, and water temperature were attributed to stream reaches (based on the EPA River Reach files, Version 3.0, www.epa.gov/waters/doc/techref.html). Water chemistry and quality were evaluated for a subset of the communities based on data availability.

Habitat

In-stream habitat assessment surveys were completed with many of the fish, mussel, and macroinvertebrate surveys in the analysis datasets. Variations on the EPA Rapid Bioassessment Habitat Protocols (RBP) (Plafkin et al. 1989; Barbour et al. 1999) were completed in many state and federal agency and river basin commission surveys, obtained for this study. Similar to water quality information, there were some locations in the analysis dataset without habitat assessments. To standardize variations in habitat assessment protocols, we calculated the percent total RBP habitat score for each stream reach. We assessed habitat scores for community types.

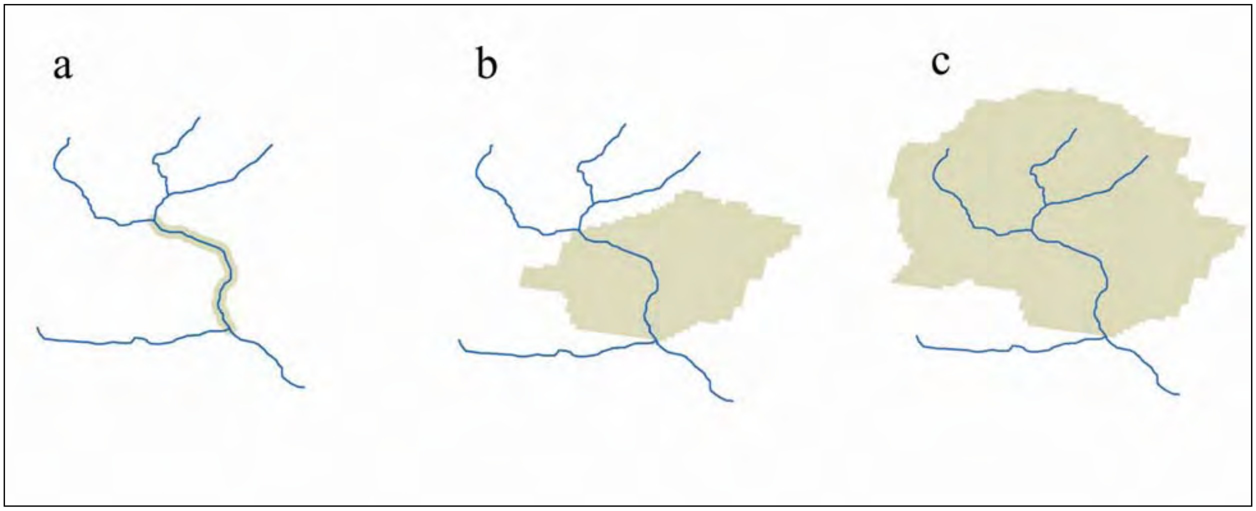
Longitudinal and watershed variables

Environmental data were attributed for each river reach in the study area. Stream variables and watershed landscape characteristics were calculated and attributed to stream reaches and to their associated reach riparian buffers, reach watersheds, and catchments (Table 5-1 and Figure 5-1). The EPA River Reach files dataset (v 3.0) (http://www.epa.gov/waters/doc/techref.htm) delineated stream reaches,

Table 5-1. Attributes summarized for reaches, riparian buffers, reach watersheds, and catchments.

Reach	Riparian buffer	Reach watershed	Catchment
Arbolate sum	Land cover	Dams	Dams
Elevation		Geology	Geology
Gradient		Landcover	Landcover
Link		Point sources	Point sources
Strahler order		Road – stream crossings	Road – stream crossings
Water chemistry			Catchment area
RBP habitat			

Figure 5-1 (a-c). Spatial boundaries of a riparian buffer surrounding a stream reach, a reach watershed, and a catchment. Areas are shaded for a) a riparian buffer, b) a reach watershed, and c) a catchment (Adapted from Brenden et al. 2006).



as the units bounded by upstream and downstream confluences. Stream reaches are hydrologically ordered and are attributed with flow direction. Riparian buffers, extending 100 m laterally from stream reaches, were created (Figure 5-1). Reach watersheds, small watersheds consisting of the land area draining directly to the reach, were developed by Anderson and Olivero (2003) (Figure 5-1). Reach watersheds were nested within catchments, which drain the entire land area upstream of each reach (Figure 5-1).

To calculate environmental variables, stream reach, reach riparian buffer, reach watershed, and catchment variables were analyzed with ArcView® (ESRI 1982-2000), Visual Basic®, and Arc/INFO® (ESRI 1982-2000) watershed tools created by The Nature Conservancy (Fitzhugh 2000). Environmental variables in Table 5-2 were summarized and calculated by Anderson and Olivero (2003) and by the report authors. Reach position in the watershed (e.g., arbolate sum, link, and Strahler order) and channel characteristics (e.g., elevation, gradient, water chemistry, and RBP habitat) were summarized for stream reaches (Table 5-1; Table 5-2).

Watershed land cover types were summarized as indicators of riparian and watershed conditions. The area and proportions of land cover classes (from the 1992 National Landcover Dataset, www.landcover.usgs.gov/uslandcover.php) were calculated within riparian buffers, reach watersheds, and catchments (Table 5-2). We summarized landcover for catchments with watershed tools (Fitzhugh 2000). Some landcover classes were aggregated (e.g., total catchment agriculture, total catchment forest, total catchment wetlands, and total catchment urban land cover types (Table 5-2)).

Geologic bedrock formations for New York, Pennsylvania, New Jersey, West Virginia, Virginia, and Ohio were evaluated for their hydrologic and chemical properties and were assigned to 6 geologic classes: sandstone, shale, calcareous, crystalline silicic, crystalline mafic, and unconsolidated formations (Table 5-2). Proportions of geologic classes were calculated for each reach watershed; watershed tools summarized the area and proportion of geology classes within catchments (Fitzhugh 2000).

In addition, information about point sources, roads, and dams was summarized for reach watersheds and catchments (Table 5-2). Industrial point sources, permitted discharges, and mines datasets were combined into a point source dataset; the density and number of point sources were calculated for reach watersheds. The number of locations where streams are crossed by roads and the density of point sources were calculated. Similarly, we attributed reach watersheds with data about hydrologic alteration from dams, including the number, density, and storage capacity of dams. We used watershed tools to summarize catchment point sources, road – stream crossings, and dams (Fitzhugh 2000).

Stream reach, riparian buffer, reach watershed, and catchment attributes were related to each EPA river reach in the study area (where data were available) in a GIS. Attributes described aquatic community occurrences and were used in community predictive models. Environmental variables were also applied in a physical stream classification (See Chapter 6) and in

Table 5-2 Environmental variables and variable codes for data attributed to stream reaches, reach riparian buffers, reach watersheds, and catchments in the study area.

Variable	Code	Definition	Data source
Physical stream class	ABIOCLASS	A stream class category combining geology class, gradient class, and watershed area class	See Chapter 6 for details.
Gradient class	GRAD_CLASS	Class of gradient (low, med, high)	See Chapter 6 for details.
Geology class	DOMUPSGEO; DOMLOCGEO	Class of dominant catchment geology; class of dominant reach watershed geology (sandstone, shale, calcareous, crystalline silicic, crystalline mafic, unconsolidated materials)	See Chapter 6 for details.
Catchment area class	WSHEDCLASS	Watershed size class	See Chapter 6 for details.
Least Disturbed Stream Class	REFSEG2	Class of stream quality	See Chapter 6 for details.
Arbolate sum	ARBOLATE_2	Total upstream stream miles	Access Visual Basic tool. TNC Stream Macrohabitats. Anderson, M.A. and A.P. Olivero. 2003. Lower New England Ecoregional Plan. The Nature Conservancy.
Elevation	AVGELV	Average reach elevation	Stream gradient and elevation AML - TNC Stream Macrohabitats. Anderson, M.A. and A.P. Olivero. 2003. Lower New England Ecoregional Plan. The Nature Conservancy.
Link	D_LINK; LINK	Number of downstream links (first order streams) in the catchment; number of upstream links in the catchment	Access Visual Basic tool. TNC Stream Macrohabitats. Anderson, M.A. and A.P. Olivero. 2003. Lower New England Ecoregional Plan. The Nature Conservancy.
Gradient	GRADIENT	Average reach slope ((elevation at ‘from node’ – elevation at ‘to node’)/ reach length)	Stream gradient and elevation AML -TNC Stream Macrohabitats. Anderson, M.A. and A.P. Olivero. 2003. Lower New England Ecoregional Plan. The Nature Conservancy.

Table 5-2 (cont'd).

Variable	Code	Definition	Data source
Catchment area	SQMI	Area (mi.^{2})	Access Visual Basic tool. TNC Stream Macrohabitats. Anderson, M.A. and A.P. Olivero. 2003. Lower New England Ecoregional Plan. The Nature Conservancy.
Stream order	STRORDER	Strahler stream order of reach	Access Visual Basic tool. TNC Stream Macrohabitats. Anderson, M.A. and A.P. Olivero. 2003. Lower New England Ecoregional Plan. The Nature Conservancy.
Catchment hydrologic impairment	DAMACCUM; DAMDENS; DAMSTACCU; DAMSTDENS	Number of upstream dams in catchment; density of dams in catchment; accumulated dam storage in catchment; density of dams * storage capacity in catchment	National Inventory of Dams in BASINS (http://www.epa.gov/OST/BASINS/)
Reach hydrologic impairment	DAMS_12; DAMSTORA_2	Number of reach dams; reach dam storage capacity	
Catchment point source pollution	PS_ACCUM	Number of point sources in catchment	Point sources were identified as mines, industrial point sources, and permitted discharges from several national datasets: USBM Mineral Availability System (http://minerals.er.usgs.gov/minerals/pubs); Superfund/CERCLIS (EPA Comprehensive Environmental Response, Compensation, and Liability Information System (http://www.epa.gov/superfund); IFD (Industrial Facilities Discharge) (http://www.epa.gov/ost/basins); TRI (Toxic Release Inventory Facilities (http://www.epa.gov/enviro/html/tris/tris_overview.html)
Reach watershed point source pollution	PTSOURCE_2; PSDENSITY	Total number of point sources in reach watershed; density of point sources in reach	

Table 5-2 (cont'd)

Variable	Code	Definition	Data source
Catchment roads	RSC_ACCUM; RSC_DENSIT	Number of catchment road – stream crossings; density of road – stream crossings in catchment	Census 2000 Tiger line files (http://www.census.gov/geo/www/tiger/)
Reach watershed roads	RDSTRXINGS; RDSTR_DENS	Number of road – stream crossings; density of reach road – stream crossings in reach watershed	
Percent of reach watershed bedrock geology	LOCALGEO1; LOCALGEO2; LOCALGEO3; LOCALGEO4; LOCALGEO5; LOCALGEO6;	% sandstone geology class in reach watershed; % shale geology class in reach watershed; % calcareous geology class in reach watershed; % crystalline silicic geology class in reach watershed; % crystalline mafic geology class in reach watershed; % unconsolidated materials geology class in reach watershed	Bedrock geology data sources: Pennsylvania - http://www.dcnr.state.pa.us/topogeo/map1/bedmap.aspx New Jersey - http://www.state.nj.us/dep/njgs/, New York - http://www.nysm.nysed.gov/gis/, Delaware - http://www.udel.edu/dgs/dgsdata/GeoGIS.html, Virginia -http://www.mme.state.va.us/dmr/DOCS/MapPub/map_pub.html, West Virginia - http://wvgis.wvu.edu/data/data.php
Percent of catchment bedrock geology	UPSTRGEO1; UPSTRGEO2; UPSTRGEO3; UPSTRGEO4; UPSTRGEO5; UPSTRGEO6	% sandstone geology class in catchment; % shale geology class in catchment; % calcareous geology class in catchment; % crystalline silicic geology class in catchment; % crystalline mafic geology class in catchment; % unconsolidated materials geology class in catchment	

Table 5-2 (cont'd)

Variable	Code	Definition	Data source
Percent of catchment landcover	PC_COMMIND;	% commercial/industrial/transportation in catchment;	National Land Cover Dataset, 1992 (http://landcover.usgs.gov/usgslandcover.php)
	PC_DECFOR;	% deciduous forest in catchment;	
	PC_EMRWET;	% emergent wetland in catchment;	
	PC_EVEFOR;	% evergreen forest in catchment;	
	PC_GRASS;	% grassland in catchment;	
	PC_HIGHURB;	% high intensity residential in catchment;	
	PC_LOWURB;	% low intensity residential in catchment;	
	PC_MIXFOR;	% mixed forest in catchment;	
	PC_NONRCAG;	% non-row crop agriculture in catchment;	
	PC_OPNWATR;	% open water in catchment;	
	PC_ORCH;	% orchard in catchment;	
	PC_PASTURE;	% pasture/hay in catchment;	
	PC_QUARMN;	% quarries / strip-mines / gravel pits in catchment;	
	PC_ROCK;	% bare rock/sand/clay in catchment;	
	PC_ROWCROP;	% agriculture in row crops in catchment;	
	PC_SCRUB;	% scrubland in catchment;	
	PC_SMGRAIN;	% small grains in catchment	
	PC_TOTAG2;	% agriculture in catchment;	
	PC_TOTFOR2;	% forest in catchment;	
	PC_TOTURB2;	% urban in catchment;	
	PC_TRANS;	% transitional in catchment;	
	PC_URBREC;	% urban/recreational grasses in catchment;	
	PC_WDYWET;	% woody wetland in catchment;	
	PCTOTWETL2	% wetland in catchment	

Table 5-2 (cont'd)

Variable	Code	Definition	Data source
Catchment landcover area	TOT_BARERO;	Area bare rock/sand/clay in catchment;	National Land Cover Dataset, 1992 (http://landcover.usgs.gov/usgslandcover.php)
	TOT_COMM_I;	Area commercial/industrial/ transportation in catchment;	
	TOT_DECFOR;	Area deciduous forest in catchment;	
	TOT_EMERWE;	Area emergent wetland in catchment;	
	TOT_EVEFOR;	Area evergreen forest in catchment;	
	TOT_GRASS;	Area grassland in catchment;	
	TOT_HIGHIN;	Area high intensity residential in catchment;	
	TOT_LOWINT;	Area low intensity residential in catchment;	
	TOT_MIXFOR;	Area mixed forest in catchment;	
	TOT_OPENWA;	Area open water in catchment;	
	TOT_ORCH;	Area orchard in catchment;	
	TOT_PASTUR;	Area pasture/hay in catchment;	
	TOT_QUARMI;	Area quarries/stripmines/gravel pits in catchment;	
	TOT_ROWCRO;	Area agriculture in row crops in catchment;	
	TOT_SCRUB;	Area scrubland in catchment;	
	TOT_SMGRAI;	Area small grains in catchment;	
	TOT_TRANS;	Area transitional in catchment;	
	TOT_URBREC;	Area urban/recreational grasses in catchment;	
	TOT_WOODYW	Area woody wetland in catchment	

Table 5-2 (cont'd)

Variable	Code	Definition	Data source
Reach watershed landcover area	IN_BAREROC;	Area bare rock/sand/clay in reach watershed;	National Land Cover Dataset, 1992 (http://landcover.usgs.gov/usgslandcover.php)
	IN_COMM_IN;	Area commercial/industrial/transportation in reach watershed;	
	IN_DECFOR;	Area deciduous forest in reach watershed;	
	IN_EMERWET;	Area emergent wetland in reach watershed;	
	IN_EVEFOR;	Area evergreen forest in reach watershed;	
	IN_GRASS;	Area grassland in reach watershed;	
	IN_HIGHINT;	Area high intensity residential in reach watershed;	
	IN_LOWINTR;	Area low intensity residential in reach watershed;	
	IN_MIXFOR;	Area mixed forest in reach watershed;	
	IN_OPENWAT;	Area open water in reach watershed;	
	IN_ORCH;	Area orchard in reach watershed;	
	IN_PASTURE;	Area pasture/hay in reach watershed;	
	IN_QUARMIN;	Area quarries/stripmines/gravel pits in reach watershed;	
	IN_ROWCROP;	Area agriculture in row crops in reach watershed;	
	IN_SCRUB;	Area scrubland in reach watershed;	
	IN_SMGRAIN;	Area small grains in reach watershed;	
	IN_TRANS;	Area transitional in reach watershed;	

Table 5-2 (cont'd)

Variable	Code	Definition	Data source
Reach watershed landcover area	IN_URBRECG; IN_WOODWET	Area urban/recreational grasses in reach watershed; Area woody wetland in reach watershed	National Land Cover Dataset, 1992 (http://landcover.usgs.gov/usgslandcover.php)
Percent riparian landcover	RIP_AG; RIP_BARREN; RIP_DEVEL; RIP_FOREST; RIP_WATER; RIP_WETL	% agriculture in reach riparian zone; % barren in reach riparian zone; % developed in reach riparian zone; % forest in reach riparian zone; % open water in reach riparian zone; % wetland in reach riparian zone	National Land Cover Dataset, 1992 (http://landcover.usgs.gov/usgslandcover.php)

assessments of potential watershed quality (See Chapter 7).

Classification methods

Grouping and refinement of stream assemblages

Outlying sites that were greater than 2.3 standard deviations from the mean were removed from analysis (McCune and Grace 2002). Cluster analysis with the Sorensen distance measure and flexible beta linkage ($\beta = -0.1$) was performed in PC-ORD (version 4.26, MjM Software Design) to group sites based on similarities in taxa composition.

Non-metric multidimensional scaling (NMS) was used as a secondary classification technique. NMS has been shown to be one of the most effective methods of ordination for ecological community data (McCune and Grace 2002). The NMS was conducted using Sorensen's distance, an appropriate distance measure for ordination of presence – absence data (McCune and Grace 2002). The number of ordination dimensions was determined by evaluating the NMS stress (McCune and Grace 2002). There is not a statistical criterion developed for selecting the appropriate number of dimensions (Kruskal and Wish 1978), but a stress of 20 or below indicates a stable solution (McCune and Grace 2002). Percent variance explained by each axis of the NMS ordination was calculated for each NMS analysis to measure the effectiveness of the ordination, how well the ordination results represent the variance in the original data, and whether the ordination axes are independent.

Other analyses refined community classes. Indicator Species Analysis (ISA) (Dufrêne and Legendre 1997) was used to determine the percent affinity of taxa in each cluster group. Mean indicator values resulting from the ISA were used as an index to evaluate patterns found in cluster groups and ordination. The classification strength was also used to prune the cluster dendrograms.

Final communities were selected representing the best grouping of sample locations with the strongest ISA values and lowest Monte-Carlo simulation p-values, and highest classification strength. Community geographic distribution and species composition were evaluated. Best professional judgment ultimately determined the most appropriate grouping of sample locations and community types. We described communities by the strongest significant taxa indicators, their distribution, reach water quality and habitat conditions, and reach and catchment environmental variables.

Predictive community modeling methods

Community presence for stream reaches was predicted by Random Forest models. Random Forest analysis[1] is a modification of Classification and Regression Tree Analysis that aggregates data into increasingly similar groups based on recursively partitioning the dataset. The analysis ultimately results in a decision tree model in which classifying characters split the dataset (McCune and Grace 2002).

Community stream reaches with stream and watershed physical attributes were used to train the community prediction models. Variables included catchment and reach bedrock geology, catchment and riparian landcover, hydrologic alteration, point source pollution, road – stream crossings, and a number of reach and watershed

---

[1] Random Forest analysis is described in Appendix 2.

attributes (Table 5-2). Five variables at each node were randomly selected to develop the models based on 1,000 trees. Predictive models were created for each classification of fish, mussel, and macroinvertebrate communities at two levels of taxonomy (genus and family).

Validating community assemblage classifications

We evaluated the community assemblages with an independent dataset. Analysis of the 30% data that remained, after the initial 70% was used to develop the final community assemblages, provided insights into the community distribution and repeatability of assemblage groupings and habitat associations. The same methods of grouping sample locations, choosing number of groups, associating habitat and environmental variables, and predicting community locations were applied to the validation dataset. If the validation analysis produced similar groupings of species and geographic distribution, and were associated with similar habitats, original community assemblages were affirmed. In other cases, if new assemblage groupings occurred, or if community assemblages were not repeated in the validation analysis, then adjustments to finalized community assemblages were made to represent the new findings.

Expert review of the project findings and field visits to community locations also informed the classification results. Peer review of community assemblages by experienced aquatic biologists at the project advisory meetings confirmed the best grouping of community assemblages. In addition, we visited approximately 100 community locations and measured water chemistry and in-stream habitat conditions, and surveyed for mussels and macroinvertebrates. No samples of fish were collected in the validation phase because of limited staff time and resources. We compared habitats, water quality, and taxa with expected conditions at community locations.

Community classification results and discussion

Community comparisons

Community analysis revealed that eleven fish communities and thirteen mussel communities occur in the study area. Depending on the analysis, there were eight to twelve macroinvertebrate communities. Macroinvertebrates in the family-level dataset had eight communities, but twelve communities were described by the genus-level dataset. Indicator species and descriptive community names are listed in Appendices 3-9. Descriptions of communities, including indicator species and habitat descriptions, can be found in the accompanying document, *User's Manual and Data Guide to the Pennsylvania Aquatic Community Classification*. The next sections compare the taxonomic classifications and their habitat affinities.

To evaluate the relative strength of classification types, classification strength, indicator values, and Indicator Species Analysis Monte-Carlo simulation p-values were compared among community classifications of fish, mussels, macroinvertebrates (genus and family analyses). The three taxa groups had differing abilities to classify streams by community types. Mussels were the strongest classifiers of flowing waters, having mean indicator values two to four times greater than macroinvertebrate classifications (Table 5-3). Mean indicator values were the highest for Susquehanna – Potomac River, Delaware River, and Ohio – Great Lakes Basins mussel classifications,

Table 5-3. Classification strength, Indicator Species Analysis, and non-metric multidimensional scaling (NMS) ordination results for mussel, fish, and macroinvertebrate community classifications. Values included are mean indicator values, Indicator Species Analysis (ISA) randomized Monte-Carlo p-values, classification strength, mean NMS stress, NMS total variance explained, and final number of NMS dimensions for the final solution for all community classifications.

Community type	Mean indicator value	Mean ISA p-value	Class strength	Mean stress	Iterations to obtain NMS solution	Total variance explained in NMS ordination	# NMS ordination dimensions
Ohio – Great Lakes Basins Mussels	23.30	0.06	0.13	19.50	22	0.68	3
Susquehanna – Potomac River Basins Mussels	47.05	0.29	0.52	12.49	20	0.90	3
Delaware River Basin Mussels	42.31	0.36	0.85	10.60	12	---	1
Ohio – Great Lakes Basins Fish	17.39	0.01	0.22	18.71	187	0.73	3
Atlantic Basin Fish	19.40	0.01	0.25	18.18	90	0.81	3
Macroinvertebrate – Genus	13.59	0.01	0.16	20.86	200	0.71	3
Macroinvertebrate – Genus (Genera Grouped to Family)	12.84	0.18	0.18	20.48	84	0.81	3
Macroinvertebrate – Family	11.58	0.05	0.20	33.71	50	0.70	3

followed by Atlantic and Ohio – Great Lakes Basins fish classifications (Table 5-3). Macroinvertebrates were the weakest classified assemblages. Indicator values for the family and generic macroinvertebrate classifications spanned from 11.58 to 13.59 and were less than a third of the strongest mussel classification mean indicator value. Genus-level macroinvertebrate dataset had indicator values that were marginally higher than the family-level macroinvertebrate dataset (Table 5-3; also see Appendices 3-9 for complete Indicator Species Analysis results).

Results from the Monte-Carlo simulations in Indicator Species Analysis (ISA) were compared among classification types. The p-values generated from Monte-Carlo simulations in ISA tell us whether the indicator taxa are statistically significant and are a metric of how well the dataset is classified. Indicator taxa were on average statistically significant ($p<0.05$) for most classifications (Table 5-3). However, some indicator species for mussel classifications in the Delaware River Basin and Ohio – Great Lakes Basins were not statistically significant, resulting in mean ISA p-values for each classification ranging from 0.22 to 0.36 (Table 5-3; Appendix 3 and Appendix 5).

Although the mussels in the Delaware River and Ohio – Great Lakes Basins had strong assemblage affinities overall, some mussel species were not strong indicators of any community type. In the Ohio – Great Lakes Basins, 16 out of 25 species were strongly associated (ISA p-value < 0.05) with an assemblage (Appendix 3). Classification of the Delaware River Basin mussel assemblages revealed that four of six mussel species have a strong community affinity (ISA p-value < 0.05) (Appendix 5). However, species like yellow lampmussel (*Lampsilis cariosa)* and triangle floater (*Alasmidonta undulata)* did not associate strongly with any community group (Appendix 5) in the Delaware River Basin classification.

Classification strengths reveal patterns about taxa and assemblage classifications similar to those suggested by the ISA. The mussel communities had some of the highest classification strengths in the analysis, ranging from 0.52 to 0.85 for the Susquehanna – Potomac River Basins and Delaware River Basin classifications, respectively (Table 5-3). Classification strengths were intermediate for fish assemblages at values ranging from 0.22 to 0.25. The lowest classification strengths spanned 0.16 to 0.20 for macroinvertebrate communities. The Ohio – Great Lake Basins mussel classification had low classification strength relative to other communities. However, the Ohio – Great Lake Basins mussel classification was determined to be a reasonable classification of mussel communities because it had relatively high indicator values and a mean NMS stress within an acceptable range. Additionally, many community indicators were strongly associated with community types.

Taxonomic type and level of taxonomy influenced classification patterns. Datasets with species-level taxonomy such as the fish and mussel datasets produced the best classifications. However, life history and basin affinity are other factors that could influence the classifications.

Taxa types differ in their mobility, zoogeographic limitations, and habitats. Mussels are limited in their dispersal by the mobility of their fish hosts during the glocidial stage, since they move relatively little as adults (Villella et al. 2004). Fish distribution is limited by zoogeography determined by current and past drainage patterns (Unmack 2001; Oberdorff et al. 1999).

Mussel communities may be the result of individuals that can survive in spatially overlapping habitats with amenable conditions. Because of their mobility, fish have more habitat choices within the available environments and may actively seek preferred habitats. Nevertheless, assemblages may be formed by species associating in preferred and overlapping habitats (Clements 1916; Clements 1920).

Macroinvertebrate communities may be less strongly defined than fish and mussel communities for several reasons. The communities shift with the season. Additionally, macroinvertebrate larvae may drift downstream during unfavorable conditions and have the opportunity to move over land as terrestrial adult forms. However, the relationship between macroinvertebrate dispersal ability and their distribution and ecology has not been addressed in many studies and is largely unknown (Bohonak and Jenkins 2003).

The taxonomic level of classification datasets confounds the comparison of classification strengths. While fish and mussel species were used to characterize assemblages, no species-level datasets were available for macroinvertebrates. Genera and families in the macroinvertebrate datasets usually encompass one or more species and their collective ranges. For example, a macroinvertebrate family with numerous species, each species having distinct habitat preferences, may be described as a generalist taxa. Since each species has a diverse niche, the family taxon collectively occurs in wide range of habitats. For a multi-family macroinvertebrate assemblage, the diversity of niches occupied by all the assemblage taxa may be even greater. A single species or species assemblage may have a more sharply defined habitat than do higher level macroinvertebrate taxa assemblages.

Comparisons of family and genus macroinvertebrate data as classifiers of aquatic communities suggest that genus taxonomy is most appropriate for classification, but it does not give a strong advantage over family taxonomy. The mean indicator values for the family dataset were slightly lower than for the genus macroinvertebrate dataset (Table 5-3). Confounding the comparison is the fact that the datasets differ in the number of samples, data source, and sampling methods.

We standardized the comparison of genus- and family-data in an additional analysis that grouped the genus dataset taxa into their respective family taxa. The grouped (family) dataset was classified with the same number of community groups as the genus classification. The classification using family taxonomy of the same dataset had a 12.84 mean indicator value, compared to the 13.59 indicator value of the genus classification. Results from a pilot study of the Pennsylvania Aquatic Community Classification also demonstrate the benefit of genus-taxonomy in classification; the study found macroinvertebrate classifications with genus data were three to six times stronger than classifications with family information (Nightingale et al. 2004).

There is no consensus among aquatic ecologists about the best taxonomic resolution for stream bioassessment and classification, but the use of genus and species taxonomy for those purposes is supported by a number of studies. The tradeoff between the effort and costs for high taxa resolution and gains in additional information should be weighed. Some researchers have found that for bioassessment, family taxa resolution is adequate (Bailey et al. 2001; Waite et al. 2004), particularly for studies where a very large number of sites must be sampled and the study goal is to detect coarse differences between sites (Lenat and Resh 2001).

Others have found that genus or species taxonomic units are better able to distinguish gradients of impairments and classify ecological units of flowing waters in subtle or sometimes dramatic ways. Ecological stream types in a large study of European streams had the least statistical overlap for species data, but were less distinct when family macroinvertebrate data was analyzed (Verdonschot 2006); ordinations distinguishing stream types were able to separate mountain streams from lowland streams and from Mediterranean streams with lower taxonomy datasets, but family taxonomy produced greater overlap in ordinations between types. Other studies similarly found that ordination stress, amount of variance explained, and correlations with environmental variables improved in ordinations of genus taxa compared to ordinations of family taxa (Arscott et al. 2006; Metzeling et al. 2006). The ability to distinguish between impaired and non-impaired streams also improves with some macroinvertebrate metrics (e.g., total taxa richness, EPT taxa richness, and % EPT taxa) when genus macroinvertebrate data are compared to family-level data (Waite et al. 2004; Metzeling et al. 2006).

Our study sought to identify ecological patterns in streams based on macroinvertebrate biological composition across varied aquatic systems. Methods for classifying biological stream types were improved with genus taxonomy of macroinvertebrates. Ordinations of genus macroinvertebrate datasets produced solutions with less stress for datasets and explained more total variance than those of family macroinvertebrate datasets (Table 5-3). Genus and species taxonomies add information about the evolutionary history and biodiversity of aquatic systems relative to higher taxonomic levels, and are recommended for selection of locations for conservation (Lenat and Resh 2001). Based on our findings, we agree with recommendations to use generic or lower taxonomy for stream classification and for setting conservation priorities.

Among fish, mussel, and macroinvertebrate taxa, our study identified mussels as the strongest classifiers of aquatic systems. Fish datasets performed intermediately at stratifying aquatic systems, while macroinvertebrates had the least robust classification. Other comparisons among other taxa as ecological classifiers in aquatic systems are few, but Paavola et al. (2003) found that fish were better classifiers of headwater streams than macroinvertebrates based on classification strength. Other papers demonstrate that biological classifications of fish typically have relatively high classification strengths, ranging from 0.35 to 0.53 (McCormick et al. 2000; Van Sickle and Hughes 2000), compared to macroinvertebrate classifications, in which class strengths varied from 0.06 to 0.15 (Gerritsen et al. 2000; Hawkins and Vinson 2000; Sandin and Johnson 2000; Waite et al. 2000).

The type of taxa used in an ecological classification should be decided based on the project goals. Freshwater mussels are not found in all flowing water systems; headwater and medium-sized streams are not usually populated by mussels. Strayer (1993) found that many Atlantic Basin mussels occurred infrequently in flowing waters with having a watershed area less than 75 mi^2. Fish and macroinvertebrate communities are found in flowing waters habitats varying from small streams to large rivers and may be more appropriate for stratifying biological stream types across a wider range of habitats. Each taxa type highlights different environmental gradients that may be important depending on study priorities.

Community predicted habitats

Community habitat associations and prediction abilities were varied within and among the aquatic animal assemblages. Comparisons of Random Forest models reveal that types of fish, mussels and macroinvertebrate communities are related to different channel and landscape variables.

Community occurrences predicted by Random Forest models were most strongly associated with longitudinal gradients of stream and river systems, landcover in catchment and riparian zone, geology, road – stream crossings, and dams. Variables indicative of longitudinal gradients like elevation, arbolate sum, and stream link are among the variables with the highest importance values[2] (importance value ≥ 1) for community classes (Appendix 10).

Mussel communities were predicted by a variety of variables across basin analyses. The classification of mussels in the Ohio – Great Lakes Basins were most strongly predicted by the percentage of several local land cover types and percent upstream landcover types, in addition to elevation, downstream link, number of accumulated catchment dams, and total catchment area (Appendix 10). In the Susquehanna – Potomac River Basins mussel classification, factors with strong importance values were catchment area in calcareous geology, catchment landcover, area of open water in the reach watershed, and reach geology classes (Appendix 10). Mussel communities of the Delaware River Basin were related to stream longitudinal variables (such as average stream reach elevation, upstream and downstream reach elevation, and stream link), the density of road – stream crossings, catchment woody wetlands, and total catchment area in shale bedrock (Appendix 10). No variables were strong predictors of community types (importance value ≥ 1) among all three basin mussel classifications. Variables that indicated stream size (e.g., catchment area), position in the watershed (e.g., elevation, downstream link), and landcover were strong predictors among at least two basin mussel assemblage models.

Fish communities were also predicted by position in the watershed and the total upstream catchment land cover. Fish in the Atlantic Basin were most strongly associated with elevation variables, a number of catchment land cover variables, arbolate sum, catchment road stream crossings, and link number, among other variables (Appendix 10). There were few strong predictor variables of fish habitat in the Ohio – Great Lakes Basins. They consisted of total catchment area in pasture and hay landcover (along with two other land cover types), arbolate sum, link, catchment area, catchment dam storage, and average elevation (Appendix 10). There was overlap of strong predictor variables between the two basin models. Predictors that had importance values > 1 in both models were arbolate sum, catchment area, catchment deciduous forest landcover, catchment pasture landcover, catchment row crop landcover, and catchment road – stream crossings.

---

[2] See Appendix 2 for a description of random forest importance values.

Macroinvertebrate community distribution was best predicted by elevation. Both genus and family assemblages had two elevation variables as strong predictors. The family classes did not have additional strong predictors, but were also associated with landcover types, such as percent catchment agriculture and percent low- and high-intensity urban catchment landcover (Appendix 10). Models of genus macroinvertebrate classes had 24 strong predictors. The strongest importance values for the prediction model included catchment area of pasture/hay landcover, arbolate sum, catchment area of deciduous forest landcover (among other catchment landcover types), catchment area, reach elevation, and gradient (Appendix10).

Physical stream[3] types were found to be less valuable predictors of community classes than other variables in Random Forest models. Importance values ranged from 0.22 to 0.65 for stream classes in all assemblage Random Forest models (Table 5-4). Among taxa classifications, fish communities in the Atlantic Basins had the highest importance values for stream classes, followed by mussel communities in the Delaware River Basin, Susquehanna – Potomac River Basins mussel communities, and family macroinvertebrate communities (Table 5-4). Importance values of stream classes were the lowest for Ohio – Great Lakes Basins mussel communities.

Geology classes were more strongly related to macroinvertebrate communities than the stream classes combining geology, gradient, and watershed size. For instance, reach watershed geology classes had higher importance values and were better predictors of the Susquehanna - Potomac River Basins mussel communities than the stream classes (Table 5-4). Local and/or upstream geology classes had higher importance values for genus macroinvertebrate assemblages and family macroinvertebrate assemblages than did the stream classes.

Model performance

Random Forest models predicted occurrences of communities with Out-of-the-Bag (OOB)[4] error rates ranging from nearly 2.9% to 51.9% (Table 5-5), where lower error rates indicate better model prediction. Mussel communities were most variable in their predictive capability. In the Delaware River Basin, mussel communities had extremely low OOB error rates (2.9%), while Susquehanna – Potomac River Basins and Ohio – Great Lakes Basins mussel communities had much higher error rates. Since the Delaware River Basin was dominated by one community, the model predicted the dominant community with a high degree of certainty. Mussels in other basins with more variable communities had more uncertainty with community predictions, and OOB error ranged from 39.8% to 48.9%. Macroinvertebrate communities, for both the family and genera datasets, had the poorest predictions of all community types. OOB rates ranged from 49.7% to 51.9%. Fish models had intermediate predictive ability, with OOB rates spanning 28.0% to 38.1%.

Assemblage types within each classification analysis had variable prediction ability in Random Forest models. For the mussel classification, four community types of thirteen communities were consistently predicted by Random Forest models, having class error ≤ 40% (Table 5-6; Appendix 11). Eight of eleven communities for fish prediction models met that criterion (Table

---

[3] See Chapter 6 for descriptions of physical stream types.

[4] Appendix 2 describes Random Forest OOB error rates.

5-6; Appendix 11). Genus macroinvertebrate Random Forest models (four out of twelve assemblages), and models of family assemblages (one out of eight assemblages) had few communities with high predictability (40% ≤ class error) (Table 5-6; Appendix 11).

We concluded that community Random Forest models are successful at predicting some assemblages based on landscape and channel characteristics. Fish and mussel models had the lowest OOB error rates and performed the best. For macroinvertebrate communities there was less certainty for habitat prediction. The high OOB error rates and poor class error rates indicated that models were less reliable. Other environmental variables and finer scale habitat characteristics may be needed to develop macroinvertebrate community habitat models.

Table 5-4. Random Forest importance values for dominant reach watershed geology (Reach WS Geol), dominant catchment geology (Catchment Geol), watershed size, gradient, and stream classes for each community classification.

Community type	Reach WS Geol	Catchment Geol	Watershed Size	Gradient	Stream class
Ohio – Great Lakes Basins Mussels	0.26	-0.12	0.49	0.04	0.22
Susquehanna – Potomac River Basins Mussels	1.10	0.03	0.00	-0.13	0.62
Delaware River Basin Mussels	0.10	0.39	0.84	0.30	0.64
Ohio – Great Lakes Basins Fish	0.35	0.29	0.86	0.36	0.44
Atlantic Basin Fish	0.19	0.52	0.58	0.71	0.65
Genus Macroinvertebrates	0.76	0.72	0.51	0.61	0.55
Family Macroinvertebrates	0.61	0.57	0.15	0.42	0.61

Table 5-5. Out-of-the-Bag error (OOB) estimate for each community classification Random Forest model.

Community type	OOB Error
Ohio – Great Lakes Basins Mussels	38.8%
Susquehanna – Potomac River Basins Mussels	48.9%
Delaware River Basin Mussels	2.9%
Ohio – Great Lakes Basins Fish	28.0%
Atlantic Basin Fish	38.1%
Genus Macroinvertebrates	49.7%
Family Macroinvertebrates	51.9%

Table 5-6 (a-c). Percent class error for each community type from Random Forest models of a) mussel communities, b) fish communities, and c) macroinvertebrate communities. Classification error rates ≤ 40% in bold type. (See Appendices 3-9 for community descriptive names and the *User's Manual and Data Guide to the Pennsylvania Aquatic Community Classification* for more details about each community.)

a) Mussels

Ohio - Great Lakes Basins		Susquehanna - Potomac River Basins		Delaware River Basin	
Community Name	Class. Error	Community Name	Class. Error	Community Name	Class. Error
Fatmucket	**26.7%**	Eastern Elliptio	**14.9%**	Eastern Elliptio	**1.0%**
Spike	62.5%	Squawfoot	81.8%	Alewife Floater	100.0%
Fluted shell	**34.1%**	Eastern Floater	100.0%	Other	50.0%
Pink Heelsplitter	100.0%	Yellow Lamp-mussel	76.9%		
		Elktoe	100.0%		
		Lanceolate Elliptio	100.0%		

b) Fish

Atlantic Basin		Ohio - Great Lakes Basins	
Community Name	Class. Error	Community Name	Class. Error
Warmwater 1	**23.4%**	Coolwater	**17.1%**
Warmwater 2	56.0%	Warmwater	**35.7%**
Coolwater 1	66.5%	Coldwater	**37.5%**
River & Impoundment	50.0%	Large Ohio River	**38.4%**
Coolwater 2	43.6%		
Coldwater	**19.9%**		
Lower Delaware River	**24.2%**		

.

Table 5-6 (a-c). (cont'd)

c) Macroinvertebrates

Genus		Family	
Community Name	Class. Error	Community Name	Class. Error
High Quality Small	**38.7%**	Low Gradient Valley	51.2%
High Quality Headwater	78.0%	High Quality Small	**27.3%**
High Quality Large	**28.4%**	Common Headwater	70.7%
Sluggish Headwater	96.0%	Limestone / Agricultural	52.7%
Common Large	75.0%	High Quality Headwater	66.6%
Limestone / Agricultural	51.6%	Common Large	64.1%
Small Urban Stream	52.2%	High Quality Mid-Sized	53.4%
Large Stream Generalist	77.8%	AMD	100.0%
Forested Headwater	100.0%		
Common Small	**38.3%**		
Ohio River	**8.7%**		
Mixed Land Use	84.6%		

6. Physical Stream Type Classification

Recognizing that the diversity of flowing water systems includes physical attributes that support biological communities, we developed a classification designed to group physical environments based on similar ecological characters.

A physical classification of flowing waters delineating ecological gradients can be applied in conservation efforts, aquatic resource monitoring, and assessment (See Chapter 2). We used ecological classification to expand on the knowledge about stream natural diversity gained from biological assemblages. The ecological gradients identified in the biological classification can be used to describe potential "habitats" for other aquatic communities. For instance, in an initial review of variables related to community occurrences, we found that traits, related to bedrock geology, watershed size, and stream reach gradient (or slope) were distinct among biological stream community types. These characters were chosen as variables to stratify the physical types of stream reaches in the study area. We explored how well patterns in a physical classification related to biological classes.

The physical classification was designed to highlight the ecological variation in the study area, including common and rare stream classes, to preliminarily assess threats to classes, and to highlight conservation value of others.

Effective conservation of biological assemblages may be limited without considering the ecological context. Some patterns in biodiversity are difficult to detect or are muted by pollution and habitat destruction. Community types and species may be specialized for habitats that have become entirely disturbed from human influences. For example, agricultural pollution degrades many limestone streams in the study area; headwater streams flowing from slopes in watersheds with little buffering capacity are commonly acidified by polluted precipitation in Pennsylvania. Specialized or endemic communities may be threatened or extirpated in habitats that are commonly polluted.

By classifying physical stream types we strive to identify habitats potentially occupied by diverse community assemblages. As mentioned in Chapter 2, ecological classifications of aquatic systems as "macrohabitats" or "valley segments" have been created for some regions in the United States. The approach for the physical classification undertaken by this project is based on The Nature Conservancy's macrohabitat methodology (Higgins et al. 2005) and other similar approaches (e.g., Seelbach et al. 2006). Segments of river reaches are classified by abiotic variables that are thought to have relatively uniform influence on biological patterns within the reach. In the macrohabitat development, flowing waters are classified by a combination of stream gradient, elevation, stream size, connectivity, drainage network position, geology, and hydrologic regime data that can be analyzed for large regions with GIS applications (Higgins et al. 2005).

In the same vein, several abiotic variables, including bedrock geology, gradient, and watershed size, were combined to create physical stream classes in the study area (Table 6-1). The result was a reasonable number of physical classes that were simply defined and had captured many of the same elements as other similar classifications.

Table 6-1. Abiotic variables associated with stream reaches to create the physical stream type classification. Table adapted from Higgins et al. (2005).

Abiotic Attribute	Rationale
Geology	Geology classes can capture influence of geology on many ecosystem attributes: water source (ground or surface), temperature, chemistry, substrate, stream geomorphology & hydrological regime.
Stream Gradient	Correlated with flow velocity, substrate material, channel morphology and stream habitat types (pools, riffles, runs, etc.).
Stream Size	Measured as drainage area and correlated with channel morphology, habitat types, habitat stability and flow volume.

The primary subclasses of geology, gradient, and watershed size were determined based on the classes from other macrohabitat classifications created by researchers at The Nature Conservancy (Anderson and Olivero 2003) and examination of the relationship between the abiotic variables and aquatic biota. While this physical classification is based primarily on abiotic variables, it was our objective that the stream types developed be biologically meaningful.

Geology type

Geology classes were defined based on simple properties that influence water chemistry and hydrologic regime, similar to those classes used by Anderson and Olivero (2003). In order to create a similar classification based on watershed geology, we gathered information about geology types as they are classified by their primary lithology in the Geologic Map of Pennsylvania (PA Bureau of Topographic and Geologic Survey 1980; Reese, personal communication; Podniesinski, personal communication). Six geology classes reflected chemical and hydrological variables adequately for Pennsylvania: sandstone, shale, calcareous, crystalline silicic, crystalline mafic and unconsolidated materials (Table 6-2). We assigned these geology classes to the bedrock geology datasets from Pennsylvania's bordering states in the study area in order to create seamless geological classes across the study area. Unfortunately, the geological data for Maryland was not available at the time of analyses and watersheds in Maryland were excluded from the physical stream classification.

The geology type that was most dominant in the catchment was associated with each stream reach (See Figure 5-1). Dominant catchment geology accounts for the cumulative effects of watershed geology on water chemistry and substrate material at a location, rather than localized effects of underlying geology at a single stream reach.

Stream gradient

The stream gradient data in this analysis were calculated by researchers at The Nature Conservancy as the proportion of change in elevation from the start and ending nodes of an individual stream reach (Anderson and Olivero 2003). Stream segments were defined by the River Reaches files (Version 3.0) (RF3), created by the U.S. Environmental Protection Agency (Dewald and Olsen 1994; see http://www.epa.gov/waters/doc/techref.html). Three categories were used that reflect patterns in biological assemblages as well as patterns in the stream gradient dataset (Table 6-2). Three gradient categories were defined as: low, medium and high (Table 6-2). These classes were chosen because of their relationship to patterns in biological communities.

Table 6-2 (a-c). Classes of a) geology, b) gradient, and c) watershed size in the physical stream types.

Abiotic Variables & Categories		Description
a) Geology Classes		
1	Sandstone	Most common type in study area. Sedimentary rock composed of sand-sized particles.
2	Shale	Second-most common geology type in study area. Fine-grained sedimentary rock.
3	Calcareous	Limestone and dolomite rock types. Small amounts of calcareous geology can have a disproportionate effect on water chemistry and biotic assemblages.
4	Crystalline silicic	Igneous or metamorphic rock containing silica ions.
5	Crystalline mafic	Igneous or metamorphic rock containing calcium, sodium, iron and magnesium ions.
6	Unconsolidated materials	Sands & gravels (mainly along coastal zones and larger rivers).
b) Stream Gradient		
1	Low gradient	0.0 – 0.5%
2	Medium gradient	0.51 – 2.0%
3	High gradient	Over 2.0%
c) Watershed Size		
1	Headwater stream	0 – 2 mi^2 (0 – 5.2 km^2)
2	Small stream	3 – 10 mi^2 (5.2 – 25.9 km^2)
3	Mid-reach stream	11 – 100 mi^2 (25.9 – 259.0 km^2)
4	Large streams and rivers	Over 100 mi^2 (259.0 km^2)

For example, some communities were found primarily in streams with gradients less than 0.5%. Communities may have preferences for the relatively slow current, the water chemistry characteristics, food types and availability, or other environmental factors associated with low gradient streams.

Watershed Size

Researches from The Nature Conservancy calculated the catchment area for each stream reach by summing the land area that contributes to the basin of each stream reach (Anderson and Olivero 2003). We determined four categories of watershed size that were associated with gradients in community distribution.

Size 1 watersheds represent the smallest headwater streams (0-2 mi^2 watershed area; 19,000 stream reaches). These streams support mainly the headwater macroinvertebrate communities. Size 2 watersheds (3-10 mi^2; 13,000 reaches) are still small in size, but support a greater diversity of macroinvertebrate and small-stream fish communities. Watersheds in the Size 3 category (11-100 mi^2; 12,000 reaches) represent mid-reach streams and are habitat for larger-stream macroinvertebrate communities and many types of fish communities. Size 4 streams (100+ mi^2; 7,000 reaches) represent the larger streams, small rivers, and large rivers of the study area. They hold nearly all mussel communities and the large-river fish communities. Size 4 streams are broad categories, encompassing large ranges in potential habitats.

Because zoogeographic factors, hydrology, and human influences may differ greatly among the large rivers, we considered each river with > 2,000 mi^2 watershed area to be an environment unique to each basin. For example, mussel fauna differs between the Susquehanna River and the Delaware River, as does the degree of connectivity. The lack of dams on the main river channel of the Delaware River may distinguish it hydrologically from the Susquehanna River, where there many impoundments. When ACC data users apply the physical stream classes, they should consider the differences between large river systems that cannot be captured with our classification.

Data Processing

The geological, gradient, and watershed size data were combined to create a physical type for each stream reach (in the EPA River Reach file dataset) in the study area. We assigned a three-digit code to the stream classes developed using the code accompanying each variable category from Table 6-2. In the code, the geology category was the first digit, the gradient category was the second digit, and the watershed area category was the third. For example, a sandstone-dominated (‘1’), medium gradient (‘2’), small stream (‘2’) would receive stream class code of ‘122’. Once the physical stream classes were defined, the stream types were assigned to reaches with biological community groups.

Stream type classification results and discussion

Stream classes and communities

Stream types were affiliated with biological communities, but the strength of the relationship between stream types and communities varied greatly across community types and taxa classifications (Table 6-3). Macroinvertebrate communities were found across stream classes to be associated with patterns of geology, gradients, and watershed size classes. For classifications of macroinvertebrates with genus- and family-level data, we found that communities were affiliated with one physical stream class at 7.9% to 95.6% of sample locations (Table 6-3). Nearly 96% of the community habitat for the macroinvertebrate (genus-level) Ohio River Community was characterized as the ‘114’ stream class, a sandstone geology, low-gradient, large-watershed class (Please see the *User’s Manual and Data Guide to the Pennsylvania Aquatic Community Classification* for details on the community types). However, most macroinvertebrate communities were associated with one single physical class in less than 30% of locations (Table 6-3).

Although a single physical class did not always delimit the habitat of each macroinvertebrate community, communities were found in multiple classes with a similar class feature, such as geology. Most notable are the genus – and family – level macroinvertebrate communities that were commonly found in calcareous geology: these communities were consistently found in physical stream classes with this geology type, denoted by stream classes where the first digit is ‘3.’ In the genus-level macroinvertebrate classification, the Limestone / Agricultural Stream Community was commonly associated with the ‘313’, ‘312’, ‘322’, and ‘323’ stream classes (Table 6-3).

Headwater macroinvertebrate communities were often associated with shale and sandstone geologies (stream classes beginning with ‘1’ and ‘2’, respectively) and in medium to high gradients (the middle digit of the stream class is ‘2’ and ‘3’ respectively). The ‘132’ stream class was associated with the macroinvertebrate (genus-level) High Quality Small Stream Community at 25.8% of community habitat. Crystalline mafic and crystalline silicic geology physical stream classes (first digit of the stream class code begins with ‘5’ and ‘4’, respectively) were associated with some communities, like the macroinvertebrate (genus-level) Small Urban Stream Community and the Common Small Stream Community (genus-level).

The mussel communities were almost exclusively associated with physical stream classes indicating large streams, lower gradients, and sandstone or shale geologies. Some communities, like the Alewife Floater Community (Delaware River Basin), the Other Community (Delaware River Basin), the Pink Heelsplitter Community (Ohio – Great Lakes Basins), and the Lanceolate Elliptio Community (Susquehanna – Potomac River Basins) occurred in only one physical stream type (e.g., the ‘114’ stream class – sandstone, low gradient, large watershed, or the ‘113’ stream class – sandstone, low gradient, medium-sized watershed). We did not identify any mussel communities that were associated with calcareous, crystalline mafic, or crystalline silicic geologies. However, there were few mussel occurrences in reaches having those dominant geology types. Mussel communities occurred in watersheds dominated by sandstone or shale geology, which are found in large proportions of most large watersheds.

The limited variation of aquatic environments for the largest streams and rivers characterized by the physical stream classification may explain why most mussel communities were found in only a few stream classes. Many mussels in the Atlantic slope drainages are found in flowing waters with watershed areas greater than 75 sq. mi (Strayer 1993). Communities were associated mainly with stream classes of the largest watershed size class (watershed area over 100 mi^2). Large streams and rivers generally had low gradients.

Table 6-3 (a-c). The physical stream classes associated with a) macroinvertebrates, b) mussels, and c) fish communities and percent community occurrence for the most strongly associated stream class. Among the associated stream types, the class most associated with community is in bold. (See the *User's Manual and Data Guide to the Pennsylvania Aquatic Community Classification* for details on the community types).

Community Name	Common Stream Classes	% Community Occurrence
a) Macroinvertebrates		
Genus-level		
High Quality Small Stream	**132**, 131, 122, 123	25.8%
High Quality Headwater Stream	**131**, 132, 122, 231	26.9%
High Quality Large Stream	**123**, 122, 113, 223, 132	23.7%
Sluggish Headwater Stream	**121**, 221, 122, 313, 322	14.3%
Common Large Stream	**122,** 113, 123, 213, 222	12.5%
Limestone / Agricultural Stream	**313,** 231, 312, 322, 323	14.6%
Small Urban Stream	**421**, 413, 113, 131	20.0%
Large Stream Generalist	**213,** 221, 113, 222, 114	15.8%
Forested Headwater Stream	**131**, 122, 132	36.8%
Common Small Stream	**522**, 213, 221, 113, 123	7.9%
Ohio River	**114**	95.6%
Mixed Land Use Stream	**123,** 331, 221, 232	28.0%
Family-level		
Low Gradient Valley Stream	**222,** 122, 113, 221, 213	9.2%
High Quality Mid-Sized Stream	**131**, 132, 122, 123, 231	21.9%
Common Headwater Stream	**131,** 132, 231, 122	25.0%
Limestone / Agricultural Stream	**313,** 322, 122, 114, 331	10.1%
High Quality Small Stream	**122,** 123, 222, 131, 113	14.1%
Common Large Stream	**131,** 231, 122, 221, 222	18.7%
High Quality Headwater Stream	**131,** 122, 132, 231, 222	18.0%
AMD Stream	**122,** 132, 231	10.7%

Table 6-3 (cont'd)

b) Mussels

Delaware River Basin		
Eastern Elliptio	**114,** 214	61.9%
Alewife Floater	**114**	100.0%
Other	**222**	100.0%
Ohio - Great Lakes Basins		
Pink Heelsplitter	**114**	100.0%
Fluted Shell	**114**, 113	57.1%
Fatmucket	**114,** 113	47.3%
Spike	**114**, 113	82.9%
Susquehanna - Potomac River Basins		
Lanceolate Elliptio	**113**	100.0%
Squawfoot	**114,** 213, 113, 313	23.8%
Yellow Lampmussel	**214**, 114, 223, 113	23.5%
Eastern Elliptio	**214**, 114, 213, 113	24.2%
Eastern Floater	**114**, 124, 214, 322	40.0%
Elktoe	**214**, 213, 231	50.0%

c) Fish

Atlantic Basin		
Warmwater 1	**113**, 123, 114, 213	18.2%
Warmwater 2	**214**, 213, 113, 114	15.2%
Coolwater 1	**132**, 123, 122	21.7%
Coolwater 2	**123**, 122, 313	17.3%
Coldwater	**132**, 123, 131, 122	29.4%
River and Impoundment	**114,** 214, 113	36.4%
Lower Delaware River	**114**, 214, 213	94.0%
Ohio - Great Lakes Basins		
Warmwater	**114**, 113, 123, 213	32.7%
Coldwater	**122**, 132, 123, 131, 113	31.3%
Coolwater	**113**, 122, 123, 213, 223	15.4%
Ohio Large River	**114**, 214	89.7%

Similar to mussel communities, fish communities were related to a subset of physical stream classes. Shale and sandstone geologies (physical class codes beginning with ‘1’ or ‘2’) were most commonly associated with fish communities. The Coldwater Communities were found in higher gradient classes and exclusively in sandstone-dominated geology streams (stream classes ‘132’ or ‘122’) (Table 6-3). No fish communities were strongly associated with calcareous, crystalline silicic, crystalline mafic, and unconsolidated material geologies. A review of fish communities in watersheds dominated by those geologies found that there were diverse assemblages present in each with no single community solely present in any one of those types. For example, in the Atlantic Basin, the Coldwater Community, the Coolwater Community 1, the Coolwater Community 2, the Warmwater Community 1, and the Warmwater Community 2 occurred in watersheds dominated by calcareous bedrock.

As expected, fish communities occurring in the valley streams were found in the largest watershed size class. The Atlantic Basin and Ohio – Great Lakes Basins Warmwater Communities occurred in larger streams (e.g., stream classes ‘113’, ‘114’, and ‘214’) with sandstone geologies and lower gradients (Table 6-3). Lastly, the largest river communities, like the River and Impoundment Community (Atlantic Basin), the Lower Delaware River (Atlantic Basin), and the Ohio Large River Community (Ohio – Great Lakes Basins), were associated with lower gradients and large stream classes (e.g., stream class ‘114’) (Table 6-3).

Physical stream types were found to be less valuable predictors of community classes than other variables in our predictive random forest analysis models (described in the Chapter 5). Fish communities in the Atlantic Basin had the highest importance values for stream classes followed by mussel communities in the Delaware River Basin, Susquehanna – Potomac River Basin mussel communities, and family macroinvertebrate communities. Importance values for stream classes were the lowest for Ohio – Great Lakes Basins mussel communities (See Chapter 5). In some cases, the components of the physical stream classification were more strongly related to biological assemblages than the overall stream class types. For instance, local geology classes were better predictors of the Susquehanna – Potomac River Basin mussel communities than were the stream classes. Reach watershed geology and catchment geology classes had higher importance values for genus and family macroinvertebrate groups than did the stream classes.

We further compared physical and biological stream classification with classification strength analysis. Stream classes had much lower classification strength values than biological community classes for most classifications (Table 6-4). Macroinvertebrate and mussel classes had classification strengths up to 40-fold greater than the physical stream classification. The classification strength value for the stream classification was similar to the classification strength of fish communities in the Ohio – Great Lakes Basins (Table 6-4). However, streams were not well classified by the physical stream classes in the Atlantic Basin compared to fish community classes.

Despite the fact that the physical stream classification generally performed poorly at classifying stream types relative to communities and weakly predicted most community occurrences, there were some relationships between biological assemblages and stream types. We learned that stream sizes, gradient, and bedrock geology in flowing waters were related to community types; macroinvertebrate communities varied with geology types and stream sizes. Watershed size and gradient classes differed among fish communities, while mussels were mostly found in the largest watershed size classes.

Table 6-4. Classification strength of physical stream classes and community classes.

Classification	Physical stream class	Community class
Ohio – Great Lakes Basins Mussels	0.07	0.13
Susquehanna – Potomac River Basins Mussels	0.05	0.52
Delaware River Basin Mussels	0.02	0.85
Ohio – Great Lakes Basins Fish	0.18	0.22
Atlantic Basin Fish	0.04	0.25
Macroinvertebrate – Genus	0.06	0.16
Macroinvertebrate – Family	0.05	0.20

7. Conservation Applications

To relate communities and streams types to quality habitats and aquatic conservation, we developed several relative measures of quality in a condition analysis. Stream reaches and small watersheds were evaluated using indicators of biological and watershed condition. The results were used to develop a ranking of potentially high quality stream reaches and watersheds meeting criteria indicating the best conditions. Conversely, the watersheds were also ranked for having intermediate and poor conditions.

The least amounts of human disturbance from non-point source pollution, point sources, mines, roads, and hydrologic alteration were assessed for stream reaches. We used watershed landcover, riparian landcover, and road – stream crossings, as indicators of non-point sources. The amount of natural and altered landcover types in the watershed was used as an index of disturbance. Disruption in riverine connectivity and hydrologic regime was relatively measured by the presence of dams. Potential water quality degradation was indicated by permitted point sources and mines. Stream reaches having the least amounts of non-point source pollution, point sources, mines, and hydrologic alteration were categorized as "Least Disturbed Streams."

To identify watershed units having relatively good and poor conditions across the study area, we also developed tiers of quality for 12 – digit hydrologic units (HUC 12), akin to small watersheds. HUC 12 units were determined to be the highest quality and second highest quality watersheds in Pennsylvania, if they met criteria for high quality community habitat, for fish and mussel indicator metrics values[1] (representing a diverse and healthy community), and for the number of stream reaches designated as "Least Disturbed Streams." The highest quality Conservation Priority Watersheds were given a "Tier 1" rank; a "Tier 2" designated the second highest quality watersheds (Table 7-1). Conservation Priority Watersheds contrasted with Restoration Priority Watersheds, which were prioritized for remediation actions. Restoration Priority Watersheds met poor quality criteria and had no Least Disturbed Streams, had low community metric scores, and had many occurrences of communities that were indicators of poor quality. Restoration Priority Watersheds that indicated the poorest conditions were given a "Tier 1" rank and those with secondarily poor conditions had a "Tier 2" rank (Table 7-1).

Table 7-1. Conservation designations for HUC 12 watersheds and tiers of watershed quality.

Conservation designation	Tier	Relative quality
Conservation priorities	Tier 1	Watershed quality ↑
	Tier 2	
Enhancement areas	Tier 1	
	Tier 2	
Restoration priorities	Tier 2	
	Tier 1	

Intermediate quality HUC 12 watersheds were also designated based on the same quality indicator variables. The Watershed

---

[1] Metrics, like taxa richness, are often used in water quality studies to relative measure community health. Macroinvertebrate and fish metrics were applied in the restoration and conservation prioritization of HUC 12s. See the *User's Manual and Data Guide to the Pennsylvania Aquatic Community Classification* for more details on community metrics.

Enhancement Areas support some high quality communities and habitats, but have more indicators of disturbance than the Conservation Priority Watersheds. “Tier 1” watersheds of the Watershed Enhancement Areas have quality indicator values below the Conservation Priority Watersheds, while “Tier 2” Watershed Enhancement Areas indicate more relative disturbance than “Tier 1” (Table 7-1).

Results of the watershed prioritization analyses reflected human landuse and pollution patterns in Pennsylvania. Conservation Priority Watersheds coincided with the greatest amounts of public lands in the northern tier of Pennsylvania. Watersheds along the ridges of the Ridge and Valley Province in the center of Pennsylvania were also designated Conservation Priority Watersheds. Urban and suburban watersheds in southeast and southwest Pennsylvania, those in agricultural valleys throughout the Commonwealth, and those in watersheds with many coal mines were classified as Restoration Priority Watersheds. Most watersheds in Pennsylvania were designated as Watershed Enhancement Areas, demonstrating that there is a large number of streams and rivers with moderate pollution and disturbance.

Details of analyses and maps of Least Disturbed Streams, Conservation Priority Watersheds, Watershed Enhancement Areas, and Restoration Priority Watersheds can be found in the *User’s Manual and Data Guide to the Pennsylvania Aquatic Community Classification.*

The high quality areas on the Allegheny River were highlighted as “Least Disturbed Stream” reaches and conservation priority watersheds.

8. Conclusions

Diverse mussel, fish, and macroinvertebrate communities occur in Pennsylvania's watersheds. Thirteen mussel communities from three basin analyses were found. Assemblages of fish grouped into eleven communities across two basins. Macroinvertebrate community types differed between datasets with family taxonomy and genus taxonomy. Eight community types were found for the family-level macroinvertebrate dataset, but genus-level macroinvertebrates were best described by twelve communities. Although there were benefits to using communities based on family taxonomy, finer taxonomic information in the genus-level dataset was most useful for macroinvertebrate community classification. We ultimately endorse genus-level in place of family-level taxonomy for defining aquatic macroinvertebrate communities.

Macroinvertebrate taxa demonstrate seasonal patterns in abundance, necessitating index periods for community analysis. Some differences in classification strength were found among index periods. Spring and winter index periods produced the best classifications. Ultimately, a community classification of macroinvertebrates for spring was developed for this project.

Sixty-four physical stream types were developed for the project study area based on geology, stream gradient, and watershed size. Nineteen stream types were most common and occurred in > 1,000 stream reaches. Headwater streams on steep slopes in shale and sandstone bedrock geology occurred most frequently in the study area. Large rivers were considered unique systems and not otherwise classified as a stream type. Community types had varied affiliations with physical stream types. Macroinvertebrate communities differed between types of geology and gradient. Some fish communities were strongly associated with habitats based on watershed size. Mussel communities were found in few stream types and tended to occur in large streams, with lower gradients and with either sandstone or shale geologies.

Modeling of community occurrences with random forest predictions was moderately successful. Mussel groups from the Ohio – Great Lakes Basins and fish community occurrences were well predicted based on channel and watershed characteristics. However, mussel communities from Susquehanna – Potomac River Basins and macroinvertebrate communities were predicted with higher error rates. Catchment landcover (e.g., amount of catchment pasture and catchment forest landcover) and longitudinal gradient characteristics (e.g., watershed area, elevation, and link) were the strongest predictors of community habitats.

In an analysis of stream condition, we examined watershed and riparian landcover, mines and points sources, road – stream crossings, and dams as indicators of habitat quality and water quality. Community types, metrics of community quality, and Least Disturbed Stream condition were used to select Conservation Priority Watersheds, Watershed Enhancement Areas, and Restoration Priority watersheds. The majority of Pennsylvania watersheds were categorized as having intermediate quality in Watershed Enhancement Area class. We determined poor quality watersheds classified as Restoration Priority Watersheds occurred in areas with extensive coal mining, agriculture, or urban development. The highest quality watersheds in

Pennsylvania, the Conservation Priority Watersheds, were found primarily in north-central Pennsylvania, where much land is in public ownership. We recommend additional protection of the high quality aquatic resources and remediation measures for watersheds with moderate to high levels of disturbance. Watershed managers, conservation planners, granting agencies, and monitoring agencies and programs can scale their efforts and funding for conservation and restoration based on the regional prioritization.

Next steps

Future efforts of the Pennsylvania Aquatic Community Classification project will focus on integrating community classes into hierarchical aquatic classification structures evolving from national and regional habitat classifications, such as the National Fish Habitat Initiative. We also anticipate refining community prediction models and evaluating other habitat variables. Prediction of rare and high quality communities will be used to target conservation in under-sampled areas.

In the near future we plan to incorporate information on community groups, physical stream types, and stream and watershed quality into local and regional conservation planning efforts. Project staff will ensure that results from the project are integrated into watershed and land management plans by conservation organizations, watershed groups, land trusts, land management agencies, and others. To date, the project staff has worked with the Western Pennsylvania Conservancy, The Nature Conservancy, and other groups to apply results of the Pennsylvania Aquatic Community Classification into conservation plans. Continued integration of the Pennsylvania Aquatic Community Classification into ongoing planning efforts like ecoregional plans at The Nature Conservancy and conservation action plans at the Western Pennsylvania Conservancy is planned. We anticipate working with state and federal agencies to advocate that assessment programs of water quality and biological surveys integrate community and physical stream types.

References Cited

Anderson, M.A. and A.P. Olivero. 2003. TNC Stream Macrohabitats. Lower New England Ecoregional Plan. The Nature Conservancy. Boston, MA.

Arscott, D.B., J.K. Jackson, and E.B. Kratzer. 2006. Role of rarity and taxonomic resolution in a regional and spatial analysis of stream macroinvertebrates. Journal of the North American Benthological Society 25(4): 977-997.

Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: Periphyton, benthic macroinvertebrates, and fish. EPA 841-B-99-002. U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

Bailey, R.C., R.H. Norris, and T.R. Reynoldson. 2001. Taxonomic resolution of benthic macroinvertebrate communities in bioassessments. Journal of the North American Benthological Society 20(2): 280-286.

Baxter, C.V. and F.R. Houser. 2000. Geomorphology, hyporheic exchange, and selection of spawning habitat by bull trout (*Salvelinus confluentus*). Canadian Journal of Fisheries and Aquatic Sciences 57: 1470-1481.

Blanck, A., P.A. Tedesco, and N. Lamouroux. 2007. Relationship between life-history strategies of European freshwater fish species and their habitat preferences. Freshwater Biology 52: 843-859.

Bohonak, A.J. and D.G. Jenkins. 2003. Ecological and evolutionary significance of dispersal by freshwater invertebrates. Ecology Letters. 6: 783-796.

Bowman, M. F. and R. C. Bailey. 1997. Does taxonomic resolution affect the multivariate description of the structure of freshwater benthic macroinvertebrate communities? Canadian Journal of Fisheries and Aquatic Sciences 54(8): 1802-1807.

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* Hughes, R. M., L. Wang, and P. W. Seelbach, *Eds.* Landscape influences on stream habitats and biological assemblages. American Fisheries Society Symposium 48. American Fisheries Society, Bethesda, MD. 698 pp.

Brosse, S., S. Lek, and C.R. Townsend. 2001. Abundance, diversity, and structure of freshwater invertebrates and fish communities: An artificial neural network approach. New Zealand Journal of Marine and Freshwater Research 35(1): 135-145.

Bryer, M.T., K. Maybury, J.S. Adams, and D.H. Grossman. 2000. More than the sum of the parts: Diversity and status of ecological systems. Pages 201-237 *In* Stein, B.A., L.S. Kutner, and J.S. Adams, *Eds*. Precious heritage: The status of biodiversity in the United States. The Nature Conservancy and Association for Biodiversity Information. Oxford University Press, New York, NY. 399 pp.

Clements, F.E. 1916. Plant succession: An analysis of the development of vegetation. Carnegie Institute of Washington, Publication 242. Carnegie Institute of Washington, Washington, D.C.

Clements, F.E. 1920. Plant indicators: The relation of plant communities to process and practice. Carnegie Institute of Washington, Publication 290. Carnegie Institute of Washington, Washington, D.C.

Costanza, R., R. d'Arge, R. de Groot, S. Farber, M. Grasso, B. Hannon, K. Limburg, S. Naeem, R.V. O'Neill, J. Paruelo, R.G. Raskin, P. Sutton, and M. van den Belt. 1997. The value of the world's ecosystem services and natural capital. Nature 387: 253-260.

Davis, L.S. and J.A. Henderson. 1978. Many uses and many users: Some desirable characteristics of a common land and water classification system. Pages 13-34 *In* Classification, inventory and analysis of fish and wildlife habitat. U.S. Fish and Wildlife Service, Washington, D.C., FWS/OBS-78/76.

Dewald, T.G. and M. V. Olsen. 1994. The EPA Reach File: A national spatial data resource. www.epa.gov/waters/doc.

Dufrene, M. and P. Legendre. 1997. Species assemblages and indicator species: The need for a flexible asymmetrical approach. Ecological Monographs 67(3): 345-366.

Faber-Langendoen, D., Editor. 2001. Plant communities of the Midwest: Classification in an ecological context. Association for Biodiversity Information, Arlington, VA. 61 pp. + Appendix (705 pp.)

Fike, J. 1999. Terrestrial and palustrine communities of Pennsylvania. Pennsylvania Natural Diversity Inventory, Pennsylvania Department of Conservation and Natural Resources. Harrisburg, PA. 86 pp.

Fitzhugh, T. 2000. GIS tools for stream and lake classification and watershed analysis: A tool primer. The Nature Conservancy. Arlington, VA. http://conserveonline.org/coldocs/2000/11/tnctools_10_20.zip/view?searchterm=tnctools_10_20

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(2): 199-214.

Gauch, H.G., Jr. 1982. Multivariate analysis in community ecology. Cambridge University Press, New York, NY.

Gerritsen, J., M.T. Barbour, and K. King. 2000. Apples, oranges, and ecoregions and littoral macroinvertebrate assemblages. Journal of the North American Benthological Society 19(3): 487-496.

Griffith, G.E., J.M. Omernick, and A.J. Woods. 1999. Ecoregions, watersheds, basins and HUCs: How state and federal agencies frame water quality. Journal of Soil and Water Conservation 54(4): 655-677.

Grossman, D.H., D. Faber-Langendoen, A.S. Weakley, M. Anderson, P. Bourgeron, R. Crawford, K. Goodin, S. Landaal, K. Metzler, K. Patterson, M. Pyne, M. Reid, and L. Sneddon. 1998. Terrestrial vegetation of the United States, Volume I. The national vegetation classification system: Development, status, and applications. The Nature Conservancy, Arlington, VA. 126 pp.

Groves, C.R., L.S. Kutner, D.M. Stoms, M.P. Murray, J.M. Scott, M. Schafale, A.S. Weakley, and R.L. Pressey. 2000. Owning up to our responsibilities: Who owns lands important for biodiversity? Pages 275- 299 *In* Stein, B.A., Kutner, L.S., and J.S. Adams, *Eds*. Precious heritage: The status of biodiversity in the United States. The Nature Conservancy and Association for Biodiversity Information. Oxford University Press, New York, NY. 399 pp.

Haag, W.R. and M.L. Warren. 2007. Freshwater mussel assemblage structure in a regulated river in the Lower Mississippi River Alluvial Basin, USA.Aquatic Conservation: Marine and Freshwater Ecosystems 17: 25-36.

Hawkins, C.P. and M.R. Vinson. 2000. Weak correspondence between landscape classifications and stream invertebrate assemblages: Implications for bioassessment. Journal of the North American Benthological Society 19(3): 501-517.

Heino, J., T. Muotka, R. Paavola, H. Hamalainen, and E. Koskenniemi. 2002. Correspondence between regional delineations and spatial patterns in macroinvertebrate assemblages of boreal headwater streams. Journal of the North American Benthological Society 21(3): 397-413.

Herlihy, A.T., R.M. Hughes, and J.C. Sifneos. 2006. Landscape clusters based on fish assemblages in the conterminous USA and their relationship to existing landscape classifications. Pages 87-112 *In* Hughes, R. M., L. Wang, and P. W. Seelbach, *Eds.* Landscape influences on stream habitats and biological assemblages. American Fisheries Society Symposium 48. American Fisheries Society, Bethesda, MD. 698 pp.

Hewlett, R. 2000. Implications of taxonomic resolution and sample habitat for stream classification at a broad geographic scale. Journal of the North American Benthological Society 19(2): 352-361.

Higgins, J.V., M.T. Bryer, M.L. Khoury, and T.W. Fitzhugh. 2005. A freshwater classification approach for biodiversity conservation planning. Conservation Biology 19(2): 432-445.

Hudson, P.L., R.W. Griffiths, and T.J. Wheaton. 1992. Review of habitat classification schemes appropriate to streams, rivers, and connecting channels in the Great Lakes Drainage Basin. Pages 73-107 *In* Busch, W.D.N. and P.G. Sly, *Eds*. The development of and aquatic habitat classification system for lakes. CRC Press, Boca Raton, FL. 225 pp.

Hughes, R.M. 1995. Defining acceptable biological status by comparing with reference conditions. Pages 31-47 *In* Davis, W.S. and T.P. Simon, *Eds*. Biological assessment and criteria: Tools for water resource planning and decision making. CRC Press, Boca Raton, FL. 415 pp.

Kellerhalls, R. and M. Church. 1998. Morphology of large rivers: Characterization and management. Pages 31-34. *In* D.P. Dodge, *Ed*. Proceedings of the International Large River Symposium. Canadian Journal of Fisheries and Aquatic Sciences. Special Publication 106.

Kruskal, J.B. and M. Wish. 1978. Multi-dimensional scaling. Sage Publications, Beverly Hills, CA.

Lenat, D.R. and V.H. Resh. 2001. Taxonomy and stream ecology: The benefits of genus- and species-level identifications. Journal of the North American Benthological Society 20(2): 287-298.

Liaw, A. and M. Wiener. 2002. Classification and regression by randomForest. R News. 2/3: 18-22.

Lotspeich, F.B. and W.S. Platts. 1982. An integrated land – aquatic classification system. North American Journal of Fisheries Management 2: 138-149.

Lyons, J. 1989. Correspondence between the distribution of fish assemblages in Wisconsin streams and Omernik's ecoregions. American Midland Naturalist 122:163-182.

Marchant, R. 1990. Robustness of classification and ordination techniques applied to macroinvertebrate communities from the La Trobe River, Victoria. Australian Journal of Marine and Freshwater Research 45: 945-962.

Marchant, R., L.A. Barmuta, and B.C. Chessman. 1995. Influence of sample quantification and taxonomic resolution on the ordination of macroinvertebrate communities from running waters in Victoria, Australia. Marine and Freshwater Research 46(2): 501-506.

Maxwell, J.R., C.J. Edwards, M.E. Jensen, S.J. Paustian, H. Parrott and D.M. Hill. 1995. A hierarchical framework of aquatic ecological units in North America (Nearctic Zone). USDA Forest Service, North-Central Forest Experiment Station, General Technical Report NC-176, St. Paul, MN.

Maybury, K.P., Editor. 1999. Seeing the forest and the trees: Ecological classification for conservation. The Nature Conservancy, Arlington, VA. 40 pp.

McCormick, F.H., D.V. Peck, and D.P. Larsen. 2000. Comparison of geographic classification schemes for Mid-Atlantic stream fish assemblages. Journal of the North American Benthological Society19(3): 385-404.

McCune, B. and J.B. Grace. 2002. Analysis of ecological communities. MjM Software Design, Gleneden Beach, OR.

McKenna, J.E., Jr., R.P. McDonald, C. Castiglione, S.S. Morrison, K.P. Kowalski, and D. R. Passino-Reader. 2006. A broadscale fish-habitat model development process: Genesee Basin, New York. Pages 533-554 *In* Hughes, R. M., L. Wang, and P. W. Seelbach, *Eds*. Influence of landscapes on stream habitats and biological assemblages. American Fisheries Society Special Publication, Bethesda, MD. 698 pp.

McMahon, G., S.M. Gregonis, S.W. Waltman, J.M. Omernik, T.D. Thorson, J.A. Freeouf, A.H. Rorick, and J.E. Keys. 2001. Developing a spatial framework of common ecological regions for the conterminous United States. Environmental Management 28(3): 293-316.

Metzeling, L., S. Perriss, and D. Robinson. 2006. Can the detection of salinity and habitat simplification gradients using rapid bioassessment of benthic invertebrates be improved through finer taxonomic resolution or alternative indices? Hydrobiologia 572: 235-252.

National Fish Habitat Action Plan. 2006. http://www.fishhabitat.org/

Nightingale, B., M. Walsh, D.D. Homans, R. Evans, E. Bond, and J. Deeds. 2004. Pennsylvania Aquatic Community Classification Project: Phase I Final Report. The Nature Conservancy, Middletown, PA and Western Pennsylvania Conservancy, Pittsburgh, PA.

Oberdorff, T., S. Lek., and J.F. Guégan. 1999. Patterns of endemism in riverine fish of the Northern Hemisphere. Ecology Letters 2: 75-81.

Omernik, J.M. 1995. Ecoregions: A spatial framework for environmental management. Pages 49-62 *In* Davis, W.S. and T.P. Simon, *Eds.* Biological assessment and criteria: Tools for water resource planning and decision making. CRC Press, Boca Raton, FL. 415 pp.

Omernik, J.M. and R.G. Bailey. 1997. Distinguishing between watersheds and ecoregions. Journal of the American Water Resources Association 33(5): 935 949.

Oswood, M.W., J.B. Reynolds, J.G. Irons, III, and A.M. Milner. 2000. Distributions of freshwater fishes in ecoregions and hydroregions of Alaska. Journal of the North American Benthological Society 19(3): 405-418.

PA Bureau of Topographic and Geologic Survey. 1980. Geologic Map of Pennsylvania. www.dcnr.state.pa.us/topogeo/

PA Department of Conservation and Natural Resources. 2001. Biodiversity Workgroup Report. PA DCNR, Harrisburg, PA.

PA Department of Conservation and Natural Resources. 2005. State Forest Resource Management Plan. PA DCNR, Harrisburg, PA.

Paavola, R., T. Muotka, R. Vitanen, J. Heino, and P. Kreivi. 2003. Are biological classifications of headwater streams concordant across multiple taxonomic groups? Freshwater Biology 48(10): 1912-1923.

Pennsylvania Comprehensive Wildlife Conservation Strategy, Version 1.0. September 2005. PA Game Commission and PA Fish and Boat Commission. http://sites.state.pa.us/PA_Exec/Fish_Boat/promo/grants/swg/00swg.htm

Plafkin, J.L., M.T. Barbour, K.D. Porter, S.K. Gross, and R.M. Hughes. 1989. Rapid bioassessment protocols for use in streams and rivers: Benthic macroinvertebrates and fish. U.S. Environmental Protection Agency, Office of Water Regulations and Standards, Washington, D.C. EPA440-4-89-001.

Platts, W.S. 1980. A plea for fishery habitat classification. Fisheries 5(1): 2-6.

Podniesinski, G. Personal communication. PA Natural Heritage Program, Western Pennsylvania Conservancy.

Poff, N.L. and J.D. Allan. 1995. Functional organization of stream fish assemblages in relation to hydrological variability. Ecology 76(2): 606-627.

Prasad, A.M., L.R. Iverson, and A. Liaw. 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological predictions. Ecosystems 9(2): 181-199.

Reash, R.J. and T.M. Berra. 1987. Comparison of fish communities in a clean-water stream and an adjacent polluted stream. American Midland Naturalist 118: 301-322.

Reese, S.O. Personal communication. PA Bureau of Topographic and Geologic Survey, PA DCNR.

Richards, C., R.J. Haro, L.B. Johnson, and G.E. Host. 1997. Catchment and reach-scale properties as indicators of macroinvertebrate species traits. Freshwater Biology 37: 219-230.

Rosgen, D.L. 1994. A classification of natural rivers. Elsevier Science B.V. 22: 169-199.

Sandin, L. and R.K. Johnson. 2000. Ecoregions and benthic macroinvertebrate assemblages of Swedish streams. Journal of the North American Benthological Society 19(3): 462-474.

Seelbach, P.W., M.J. Wiley, M.E. Baker, and K.E. Wehrly. 2006. Initial classification of river valley segments across Michigan's lower peninsula. Pages 25-48 *In* Hughes, R. M., L. Wang, and P. W. Seelbach, *Eds.* Landscape influences on stream habitats and biological assemblages. American Fisheries Society Symposium 48. American Fisheries Society, Bethesda, MD. 698 pp.

Sowa, S.P., D.D. Diamond, R. Abbitt, G. Annis, T. Gordon, M.E. Morey, G.R. Sorensen, and D. True. 2005. A GAP analysis for riverine ecosystems of Missouri. Final Report, submitted to the USGS National GAP Analysis Program. 1675 pp.

Stein, B.A. and F.W. Davis. 2000. Discovering life in America: Tools and techniques of biodiversity inventory. Pages 19-53 *In* Stein, B.A., L.S. Kutner, and J.S. Adams, *Eds*. Precious heritage: The status of biodiversity in the United States. The Nature Conservancy and Association for Biodiversity Information. Oxford University Press, New York, NY. 399 pp.

Strayer, D.L. 1993. Macrohabitats of freshwater mussels (Bivalvia: Unioacea) in streams of the North Atlantic Slope. Journal of the North American Benthological Society 12(3): 236-246.

US Geologic Survey. 1998. Fish communities and their relation to physical and chemical characteristics of streams from selected environmental settings in the lower Susquehanna River Basin, 1993-1995. US Department of the Interior, USGS, Lemoyne, PA. 43 pp.

US Geologic Survey. 2003. Great Lakes regional aquatic GAP analysis: Preserving biodiversity in the Great Lakes Basin. USGS Great Lakes Science Center, Ann Arbor, MI., GLSC Fact Sheet 2003-1.

Unmack, P.J. 2001. Biogeography of Australian freshwater fishes. Journal of Biogeography 28: 1053-1089.

Usio, N. 2007. Endangered crayfish in northern Japan: Distribution, abundance, and microhabitat specificity in relation to stream and riparian environment. Biological Conservation 134(4): 517-526.

Vannote, R.L., G.W. Minshall, K.W. Cummins, J.R. Sedell, and C.E. Cushing. 1980. The river continuum concept. Canadian Journal of Fisheries and Aquatic Sciences 37: 130-137.

Van Sickle, J. 1997. Using mean similarity dendrograms to evaluate classifications. Journal of Agricultural, Biological and Environmental Statistics 2: 1-25.

Van Sickle, J. and R.M. Hughes. 2000. Classification strengths of ecoregions, catchments, and geographic clusters for aquatic vertebrates in Oregon. Journal of the North American Benthological Society 19(3): 370-384.

Verdonschot, P.F.M. 2006. Data composition and taxonomic resolution in macroinvertebrate stream typology. Hydrobiologia 566: 59-74.

Villella, R.F., D.R. Smith, and D.P. Lemarié. 2004. Estimating survival and recruitment in a freshwater mussel population using mark-recapture techniques. American Midland Naturalist 151: 114-133.

Waite, I.R., A.T. Herlihy, D.P. Larsen, and D.J. Klemm. 2000. Comparing strengths of geographic and nongeographic classifications of stream benthic macroinvertebrates in the Mid-Atlantic Highlands, USA. Journal of the North American Benthological Society 19(3): 429-441.

Waite, I.R., A.T. Herlihy, D.P. Larsen, N.S. Urquhart, and D.J. Klemm. 2004. The effects of macroinvertebrate resolution in large landscape bioassessments: An example from the Mid-Atlantic Highlands, U.S.A. Freshwater Biology 49: 474-489.

Wehrly, K.E., M.J. Wiley, and P.W. Seelbach. 1998. A thermal habitat classification for lower Michigan rivers. Fisheries Research Report No. 2038. Michigan Department of Natural Resources.

Appendix 1. Description of Indicator Species Analysis and classification strength methods

Indicator Species Analysis (ISA) calculates the frequency of occurrence, relative abundance, and proportion indication score of each taxon within the cluster groups. The indication score ranges from 0 to 100 percent of perfect indication for taxa. Using 10,000 Monte-Carlo simulations, ISA randomly assigns sites to cluster groups and calculates the proportion of runs that the percent indication in the random simulations is greater than the observed percent indication. The resulting p-value is calculated (McCune and Grace 2002). Following recommendations from McCune and Grace (2002), for each taxa classification we used the lowest mean Monte-Carlo simulation p-value and highest indication score among 2 to 20 cluster groups to choose the best grouping. Any significant indicators (with $p < 0.05$, unless otherwise noted) were noted from the ISA as indicators of that particular community group.

Classification strength is based on similarity between all pairs of sites. The relationship between within- and between-cluster group similarities indicates relative strength of a classification (Van Sickle 1997). In strong classifications, the similarities between classification units in the same group are much larger than similarities between sites that are in different groups. Using PC-ORD and MEANSIM6 (Van Sickle 1997) we calculated the mean of between-group similarities (Bbar), within-group similarity (W_i) for all *i* classes (where $i = 1,2,3\ldots k$), and the weighted mean of within-group similarity (Wbar).

$\text{Wbar} = \Sigma_i \, (n_i/N)W_i$

Where

n_i = number of sites in each group *i*
N = total number of sites

The class strength statistic, CS = Wbar – Bbar, represents classification strength as the difference of mean between-group similarity and mean within-group similarity in units of similarity.

Appendix 2. Description of Random Forest analysis method

Random Forest generates a large number of classification trees, specified by the user, using a bootstrapping method to select a subset of the predictors randomly chosen at each node to further classify the dataset. The best grouping from the variables is chosen at each node and the un-pruned trees are aggregated by averaging. The method chooses the majority of correctly grouped sites by summing the predictions of all iterations and predicts new data based on the majority votes. The successive iterations of trees are independently created with bootstrapped data in a "bagging" method (Liaw and Wiener 2002). An Out-of-Bag error rate (OOB) is calculated for each bootstrap iteration by predicting the dataset not in the bootstrap sample with the classification tree developed from the bootstrap sample. An overall class OOB error is calculated from combining the OOB from each bootstrap iteration (Liaw and Wiener 2002). Because the predictor variables are randomly selected, correlation among un-pruned trees is low.

Because of its predictive advantage and ease of use, Random Forest was chosen for community prediction. Compared to other modeling methods, Random Forest has better predictive capability, creates better estimates of suitable habitat, and has superior estimates of the importance of abiotic variables in models of tree species (Prasad et al. 2006). Since many un-pruned trees are aggregated, variance is relatively small compared to other methods (Prasad et al. 2006). Only two parameters are set by the user: the number of variables in the random subset at each node and the number of trees in the forest (Liaw and Wiener 2002). The measure of how useful each variable was in constructing the model is given as importance values. The importance value is equal to the change in prediction error when the variable is permuted while all other variables are constant (Liaw and Wiener 2002).

R software (Foundation for Statistical Computing, version 2.2.1) was used to develop the Random Forest models. Software is available for download at no cost (http://www.r-project.org/).

Appendix 3. Indicator Species Analysis results for Great Lakes – Ohio Basins mussel communities. Indicator values and randomized Monte-Carlo simulation mean indicator values, standard deviation, and p-values are presented. Significant indicator taxa ($p < 0.05$) are in bold type. Special Concern species designated by the PA Fish and Boat Commission were omitted from this table.

Community Name	Common Name	Scientific Name	Indicator Value	Randomized Mean Ind Value	Randomized IV Std Dev	Monte-Carlo p-value
Fatmucket Mussel Community	**Fatmucket**	***Lampsilis siliquoidea***	**53.1**	**25.4**	**6.30**	**0.004**
	Giant floater	***Pyganodon grandis***	**34.3**	**11.8**	**3.95**	**0.001**
	Three-ridge	***Amblema plicata***	**22.1**	**13.2**	**4.45**	**0.029**
	Wabash pigtoe	***Fusconaia flava***	**20.5**	**8.8**	**3.49**	**0.014**
	White heelsplitter	*Lasmigona complanata*	4.9	3.2	2.08	0.225
	Paper pondshell	*Utterbackia imbecillis*	2.7	4.4	2.42	0.812
	Eastern pondmussel	*Ligumia nasuta*	1.6	2.4	1.58	1.000
Spike Mussel Community	**Spike**	***Elliptio dilatata***	**71.8**	**23.1**	**5.68**	**0.001**
	Black sandshell	*Ligumia recta*	12.8	8.8	3.70	0.134
Fluted Shell Mussel Community	**Fluted shell**	***Lasmigona costata***	**46.6**	**20.2**	**4.90**	**0.003**
	Kidneyshell	***Ptychobranchus fasciolaris***	**46.2**	**13.9**	**4.54**	**0.001**
	Mucket	***Actinonaias ligamentina***	**43.1**	**18.7**	**5.10**	**0.002**
	Elktoe	***Alasmidonta marginata***	**40.7**	**11.1**	**4.01**	**0.001**
	Squawfoot	***Strophitus undulatus***	**29.2**	**13.5**	**4.71**	**0.014**
	Pocketbook	***Lampsilis ovata***	**27.1**	**10.8**	**3.98**	**0.006**
	Plain pocketbook	*Lampsilis cardium*	24.9	16.3	5.29	0.067
	Wavy-rayed lampmussel	***Lampsilis fasciola***	**20.9**	**10.3**	**4.09**	**0.026**
	Creek heelsplitter	***Lasmigona compressa***	**9.5**	**5.3**	**2.86**	**0.088**
	Round pigtoe	*Pleurobema sintoxia*	9.4	14.3	5.20	0.866
	Wabash pigtoe	*Fusconaia flava*	7.1	5.8	3.02	0.230
	Cylindrical papershell	*Anodontoides ferussacianus*	6.2	5.0	3.10	0.234
	Rainbow mussel	*Villosa iris*	2.3	3.1	1.98	0.601
Pink Heelsplitter Community	**Pink heelsplitter**	***Potamilus alatus***	**68.8**	**7.7**	**3.25**	**0.001**
	Mapleleaf	***Quadrula quadrula***	**14.6**	**3.1**	**2.05**	**0.004**
	Fragile papershell	***Leptodea fragilis***	**12.5**	**5.8**	**2.92**	**0.038**

Appendix 4. Indicator Species Analysis results for Susquehanna – Potomac River Basins mussel communities. Indicator values and randomized Monte-Carlo simulation mean indicator values, standard deviation, and p-values are presented. Significant indicator taxa ($p < 0.05$) are in bold type. Special Concern species designated by the PA Fish and Boat Commission were omitted from this table.

Community Name	Common Name	Scientific Name	Indicator Value	Randomized Mean Ind Value	Randomized IV Std Dev	Monte-Carlo p-value
Eastern Elliptio Community	**Eastern elliptio**	***Elliptio complanata***	**71.0**	**20.7**	**4.59**	**0.001**
	Rainbow mussel	*Villosa iris*	8.5	9.1	7.71	0.359
Squawfoot Mussel Community	**Squawfoot**	***Strophitus undulatus***	**86.1**	**18.6**	**7.39**	**0.001**
	Triangle floater	*Alasmidonta undulata*	25.3	15.4	8.35	0.121
	Eastern lampmussel	*Lampsilis radiata*	11.2	12.6	8.22	0.404
	Cylindrical papershell	*Anodontoides ferussacianus*	7.1	6.6	6.37	0.283
	White heelsplitter	*Lasmigona complanata*	2.3	5.7	6.30	0.786
Eastern Floater Community	**Eastern floater**	***Pyganodon cataracta***	**87.7**	**14.3**	**8.60**	**0.001**
Yellow Lampmussel Community	**Yellow lampmussel**	***Lampsilis cariosa***	**77.0**	**18.0**	**7.53**	**0.001**
Lanceolate Elliptio Complex Community	**Atlantic spike**	***Elliptio producta***	**99.7**	**9.7**	**7.02**	**0.001**
	Plain pocketbook	*Lampsilis cardium*	20.0	4.2	6.01	0.102
Elktoe Community	**Elktoe**	***Alasmidonta marginata***	**92.5**	**14.4**	**8.77**	**0.001**

Appendix 5. Indicator Species Analysis results for Delaware River Basin mussel communities. Indicator values and randomized Monte-Carlo simulation mean indicator values, standard deviation, and p-values are presented. Significant indicator taxa ($p < 0.05$) are in bold type. Special Concern species designated by the PA Fish and Boat Commission were omitted from this table. One community type with only Special Concern Species is not presented here.

Community Name	Common Name	Scientific Name	Indicator Value	Randomized Mean Ind Value	Randomized IV Std Dev	Monte-Carlo p-value
Eastern Elliptio Community	**Eastern elliptio**	***Elliptio complanata***	**83.2**	**34.0**	**1.28**	**0.001**
	Yellow lampmussel	*Lampsilis cariosa*	1.3	3.3	5.40	1.000
Alewife Floater Community	**Alewife floater**	***Anodonta implicata***	**97.3**	**35.7**	**10.09**	**0.001**
	Squawfoot	***Strophitus undulatus***	**65.3**	**16.7**	**8.26**	**0.001**
	Eastern floater	***Pyganodon cataracta***	**16.3**	**2.6**	**4.07**	**0.031**
	Triangle floater	*Alasmidonta undulata*	14.1	9.3	5.96	0.170

Appendix 6. Indicator Species Analysis results for Ohio – Great Lakes Basins fish communities. Indicator values and randomized Monte-Carlo simulation mean indicator values, standard deviation, and p-values are presented. Significant indicator taxa ($p < 0.05$) are in bold type. Special Concern species designated by the PA Fish and Boat Commission were omitted from this table.

Community Name	Community Name	Scientific Name	Indicator Value	Randomized Mean Ind Value	Randomized IV Std Dev	Monte-Carlo p-value
Coolwater Community	**Blacknose dace**	***Rhinichthys atratulus***	**44.8**	**13.9**	**0.75**	**0.001**
	Creek chub	***Semotilus atromaculatus***	**37.4**	**14.2**	**0.78**	**0.001**
	White sucker	***Catostomus commersoni***	**29.1**	**15**	**0.78**	**0.001**
	Redside dace	***Clinostomus elongatus***	**16.3**	**3.4**	**0.43**	**0.001**
	Longnose dace	***Rhinichthys cataractae***	**13.1**	**4.6**	**0.50**	**0.001**
	Fathead minnow	***Pimephales promelas***	**1.8**	**0.9**	**0.24**	**0.003**
	Pearl dace	*Margariscus margarita*	1.2	0.9	0.24	0.085
	American brook lamprey	*Lampetra appendix*	1.1	1.1	0.27	0.409
Warmwater Community	**Greenside darter**	***Etheostoma blennioides***	**56.9**	**6.6**	**0.59**	**0.001**
	Northern hogsucker	***Hypentelium nigricans***	**45.5**	**11.6**	**0.74**	**0.001**
	River chub	***Nocomis micropogon***	**42.1**	**5.0**	**0.51**	**0.001**
	Bluntnose minnow	***Pimephales notatus***	**40.9**	**7.7**	**0.62**	**0.001**
	Central stoneroller	***Campostoma anomalum***	**37.1**	**8.8**	**0.68**	**0.001**
	Rainbow darter	***Etheostoma caeruleum***	**35.6**	**6.3**	**0.59**	**0.001**
	Rosyface shiner	***Notropis rubellus***	**35.3**	**3.7**	**0.47**	**0.001**
	Banded darter	***Etheostoma zonale***	**33.9**	**3.0**	**0.42**	**0.001**
	Smallmouth bass	***Micropterus dolomieu***	**32.3**	**8.2**	**0.63**	**0.001**
	Common shiner	***Luxilus cornutus***	**30.1**	**6.6**	**0.60**	**0.001**
	Rock bass	***Ambloplites rupestris***	**29.5**	**5.7**	**0.57**	**0.001**
	Johnny darter	***Etheostoma nigrum***	**28.9**	**8.2**	**0.62**	**0.001**
	Fantail darter	***Etheostoma flabellare***	**27.3**	**7.7**	**0.62**	**0.001**
	Variegate darter	***Etheostoma variatum***	**21.1**	**2.1**	**0.36**	**0.001**
	Logperch	***Percina caprodes***	**20.6**	**3.9**	**0.47**	**0.001**
	Stonecat	***Noturus flavus***	**19.6**	**2.0**	**0.37**	**0.001**
	Silver shiner	***Notropis photogenis***	**18.9**	**2.2**	**0.34**	**0.001**
	Blackside darter	***Percina maculata***	**16.8**	**3.4**	**0.44**	**0.001**
	Striped shiner	***Luxilus chrysocephalus***	**16.6**	**2.2**	**0.37**	**0.001**
	Golden redhorse	***Moxostoma erythrurum***	**15.1**	**4.4**	**0.49**	**0.001**
	Sand shiner	***Notropis stramineus***	**13.7**	**2.2**	**0.35**	**0.001**
	Mimic shiner	***Notropis volucellus***	**11.9**	**2.2**	**0.36**	**0.001**
	Pumpkinseed	***Lepomis gibbosus***	**11.0**	**5.6**	**0.56**	**0.001**

Appendix 6. (Cont'd)

Warmwater Community	**Bluegill**	***Lepomis macrochirus***	**10.7**	**5.9**	**0.56**	**0.001**
	Spotfin shiner	***Cyprinella spiloptera***	**10.2**	**1.9**	**0.34**	**0.001**
	Yellow bullhead	***Ameiurus natalis***	**10**	**1.9**	**0.33**	**0.001**
	Silverjaw minnow	***Ericymba buccata***	**8.4**	**1.9**	**0.34**	**0.001**
	Largemouth bass	***Micropterus salmoides***	**8.4**	**3.6**	**0.46**	**0.001**
	Green sunfish	***Lepomis cyanellus***	**8.3**	**2.6**	**0.39**	**0.001**
	Streamline chub	***Erimystax dissimilis***	**5.7**	**0.9**	**0.26**	**0.001**
	Yellow perch	***Perca flavescens***	**5.5**	**1.9**	**0.35**	**0.001**
	Black redhorse	***Moxostoma duquesnei***	**4.9**	**0.9**	**0.25**	**0.001**
	Brown bullhead	***Ameiurus nebulosus***	**4.0**	**1.9**	**0.35**	**0.001**
	Golden shiner	***Notemigonus crysoleucas***	**4.0**	**1.3**	**0.28**	**0.001**
	Tonguetied minnow	***Exoglossum laurae***	**3.9**	**0.8**	**0.23**	**0.001**
	Spottail shiner	***Notropis hudsonius***	**3.9**	**1.3**	**0.29**	**0.001**
	Longhead darter	***Percina macrocephala***	**2.6**	**0.6**	**0.20**	**0.001**
	Grass pickerel	***Esox americanus vermiculatus***	**2.4**	**1.0**	**0.26**	**0.001**
	Trout perch	***Percopsis omiscomaycus***	**2.2**	**1.0**	**0.25**	**0.001**
	Channel darter	***Percina copelandi***	**1.7**	**0.6**	**0.20**	**0.001**
	Ohio lamprey	***Ichthyomyzon bdellium***	**1.7**	**0.7**	**0.21**	**0.004**
Coldwater Community	**Brook trout**	***Salvelinus fontinalis***	**62.2**	**5.9**	**0.56**	**0.001**
	Mottled sculpin	***Cottus bairdii***	**35.7**	**13.5**	**0.76**	**0.001**
	Brown trout	***Salmo trutta***	**26.6**	**5.3**	**0.57**	**0.001**
	Rainbow trout	***Oncorhynchus mykiss***	**1.8**	**0.9**	**0.25**	**0.007**
Large River Community	**Channel catfish**	***Ictalurus punctatus***	**36.3**	**3.1**	**0.42**	**0.001**
	Sauger	***Sander canadensis***	**27.6**	**2.5**	**0.39**	**0.001**
	Common carp	***Cyprinus carpio***	**27.1**	**5.0**	**0.53**	**0.001**
	Gizzard shad	***Dorosoma cepedianum***	**25.3**	**2.7**	**0.39**	**0.001**
	Freshwater drum	***Aplodinotus grunniens***	**21.3**	**2.1**	**0.36**	**0.001**
	Walleye	***Sander vitreus***	**19.3**	**2.2**	**0.36**	**0.001**
	White bass	***Morone chrysops***	**17.5**	**1.6**	**0.31**	**0.001**
	Shorthead redhorse	***Moxostoma macrolepidotum***	**16.5**	**2.2**	**0.38**	**0.001**
	Spotted bass	***Micropterus punctulatus***	**16.1**	**1.8**	**0.34**	**0.001**
	Silver redhorse	***Moxostoma anisurum***	**14.0**	**2.3**	**0.37**	**0.001**
	Quillback carpsucker	***Carpiodes cyprinus***	**11.6**	**1.5**	**0.31**	**0.001**
	Emerald shiner	***Notropis atherinoides***	**9.1**	**2.8**	**0.39**	**0.001**

Appendix 6. (Cont'd)

Ohio Large River Community	Flathead catfish	*Pylodictis olivaris*	9.0	1.0	0.27	0.001
	Black crappie	*Pomoxis nigromaculatus*	8.7	1.6	0.32	0.001
	Smallmouth buffalo	*Ictiobus bubalus*	6.7	0.9	0.24	0.001
	River redhorse	*Moxostoma carinatum*	6.3	0.9	0.25	0.001
	Mooneye	*Hiodon tergisus*	5.9	0.7	0.24	0.001
	White crappie	*Pomoxis annularis*	5.2	1.1	0.26	0.001
	Muskellunge	*Esox masquinongy*	4.0	0.7	0.22	0.001
	Longnose gar	*Lepisosteus osseus*	2.9	0.5	0.18	0.001
	Brook silverside	*Labidesthes sicculus*	2.0	0.6	0.20	0.001
	Northern pike	*Esox lucius*	1.8	0.6	0.21	0.002

Appendix 7. Indicator Species Analysis results for Atlantic Basin fish communities. Indicator values and randomized Monte-Carlo simulation mean indicator values, standard deviation, and p-values are presented. Significant indicator taxa (p < 0.05) are in bold type.

Community Name	Common Name	Scientific Name	Indicator Value	Randomized Mean Ind Value	Randomized IV Std Dev	Monte-Carlo p-value
Warmwater Community 1	**Central stoneroller**	***Campostoma anomalum***	**45.6**	**3.0**	**0.44**	**0.001**
	Northern hogsucker	***Hypentelium nigricans***	**43.8**	**3.9**	**0.46**	**0.001**
	River chub	***Nocomis micropogon***	**31.6**	**2.7**	**0.40**	**0.001**
	Longnose dace	***Rhinichthys cataractae***	**27.7**	**8.0**	**0.61**	**0.001**
	Cutlips minnow	***Exoglossum maxillingua***	**27.1**	**6.0**	**0.55**	**0.001**
	Mottled sculpin	***Cottus bairdii***	**21.4**	**2.5**	**0.46**	**0.001**
	Margined madtom	***Noturus insignis***	**21.3**	**4.2**	**0.52**	**0.001**
	Creek chub	***Semotilus atromaculatus***	**19.3**	**6.4**	**0.57**	**0.001**
	Rosyface shiner	***Notropis rubellus***	**18.0**	**1.9**	**0.37**	**0.001**
	Fantail darter	***Etheostoma flabellare***	**7.6**	**1.0**	**0.30**	**0.001**
	Greenside darter	***Etheostoma blennioides***	**5.3**	**0.8**	**0.30**	**0.001**
Warmwater Community 2	**Redbreast sunfish**	***Lepomis auritus***	**38.6**	**2.7**	**0.43**	**0.001**
	Rock bass	***Ambloplites rupestris***	**34.3**	**3.9**	**0.49**	**0.001**
	Spotfin shiner	***Cyprinella spiloptera***	**29.1**	**2.0**	**0.41**	**0.001**
	Fallfish	***Semotilus corporalis***	**26.8**	**4.7**	**0.50**	**0.001**
	Smallmouth bass	***Micropterus dolomieu***	**26.6**	**5.0**	**0.50**	**0.001**
	Spottail shiner	***Notropis hudsonius***	**26.4**	**3.0**	**0.45**	**0.001**
	Common shiner	***Luxilus cornutus***	**26.0**	**5.4**	**0.55**	**0.001**
	Tesselated darter	***Etheostoma olmstedi***	**22.7**	**6.4**	**0.55**	**0.001**
	Pumpkinseed	***Lepomis gibbosus***	**22.4**	**4.4**	**0.49**	**0.001**
	Bluntnose minnow	***Pimephales notatus***	**21.8**	**2.4**	**0.44**	**0.001**
	Bluegill	***Lepomis macrochirus***	**21.7**	**3.3**	**0.46**	**0.001**
	Green sunfish	***Lepomis cyanellus***	**18.7**	**2.1**	**0.39**	**0.001**
	Satinfin shiner	***Cyprinella analostana***	**18.7**	**1.2**	**0.34**	**0.001**
	Swallowtail shiner	***Notropis procne***	**17.9**	**1.1**	**0.33**	**0.001**
	Yellow bullhead	***Ameiurus natalis***	**17.0**	**1.3**	**0.34**	**0.001**
	Shield darter	***Percina peltata***	**14.2**	**1.9**	**0.38**	**0.001**
	American eel	***Anguilla rostrata***	**13.2**	**3.4**	**0.47**	**0.001**
	Largemouth bass	***Micropterus salmoides***	**12.4**	**2.2**	**0.39**	**0.001**
	Common carp	***Cyprinus carpio***	**10.5**	**1.7**	**0.39**	**0.001**
	Comely shiner	***Notropis amoenus***	**7.2**	**0.6**	**0.25**	**0.001**

Appendix 7. (Cont'd)

Warmwater Community 2	**Chain pickerel**	***Esox niger***	**5.8**	**1.6**	**0.32**	**0.001**
	Banded darter	***Etheostoma zonale***	**5.4**	**1.0**	**0.29**	**0.001**
	Brown bullhead	***Ameiurus nebulosus***	**5.0**	**2.2**	**0.37**	**0.001**
	Redfin pickerel	***Esox americanus***	**3.6**	**0.6**	**0.23**	**0.001**
	Creek chubsucker	***Erimyzon oblongus***	**3.4**	**0.8**	**0.30**	**0.001**
	Sea lamprey	***Petromyzon marinus***	**1.7**	**0.4**	**0.21**	**0.001**
	Rosyside dace	***Clinostomus funduloides***	**1.4**	**0.8**	**0.29**	**0.044**
Coolwater Community 1	**Slimy Sculpin**	***Cottus cognatus***	**14.7**	**3.7**	**0.50**	**0.001**
	Fathead minnow	***Pimephales promelas***	**1.4**	**0.6**	**0.24**	**0.029**
	Pearl dace	***Margariscus margarita***	1.0	0.6	0.25	0.059
River and Impoundment Community	**Walleye**	***Stizostedion vitreus***	**6.3**	**0.8**	**0.26**	**0.001**
	Yellow perch	***Perca flavescens***	**3.1**	**1.1**	**0.28**	**0.001**
	Black crappie	***Pomoxis nigromaculatus***	**2.8**	**0.8**	**0.27**	**0.002**
	Goldfish	***Carassius auratus***	**1.1**	**0.5**	**0.25**	**0.044**
Coolwater Community 2	**Blacknose dace**	***Rhinichthys atratulus***	**23.6**	**9.5**	**0.61**	**0.001**
	White sucker	***Catostomus commersoni***	**22.0**	**9.8**	**0.59**	**0.001**
	Golden shiner	***Notemigonus crysoleucas***	**4.2**	**1.6**	**0.35**	**0.001**
Coldwater Community	**Brook trout**	***Salvelinus fontinalis***	**64.1**	**4.6**	**0.52**	**0.001**
	Brown trout	***Salmo trutta***	**27.8**	**5.2**	**0.52**	**0.001**
	Rainbow trout	***Oncorhynchus mykiss***	0.7	0.4	0.22	0.095
Lower Delaware River Community	**White perch**	***Morone americana***	**85.5**	**0.9**	**0.29**	**0.001**
	Channel catfish	***Ictalurus punctatus***	**49.0**	**0.9**	**0.30**	**0.001**
	Blueback herring	***Alosa aestivalis***	**30.0**	**0.5**	**0.23**	**0.001**
	Eastern silvery minnow	***Hybognathus regius***	**21.2**	**0.5**	**0.23**	**0.001**
	White catfish	***Ameiurus catus***	**21.1**	**0.5**	**0.22**	**0.001**
	Striped bass	***Morone saxatilis***	**20.1**	**0.4**	**0.24**	**0.001**
	Gizzard shad	***Dorosoma cepedianum***	**18.0**	**0.6**	**0.23**	**0.001**
	American shad	***Alosa sapidissima***	**15.1**	**0.5**	**0.24**	**0.001**
	Banded killifish	***Fundulus diaphanus***	**9.8**	**1.3**	**0.33**	**0.001**

Appendix 8. Indicator Species Analysis results for genus-level macroinvertebrate communities. Indicator values and randomized Monte-Carlo simulation mean indicator values, standard deviation, and p-values are presented. Significant indicator taxa (p < 0.05) are in bold type.

Community Name	Class/Phylum	Order	Family	Genus	Indicator Value	Randomized Mean Ind Value	Randomized IV Std Dev	Monte-Carlo p-value
High Quality Small Stream Community	**Insecta**	**Ephemeroptera**	**Heptageniidae**	***Epeorus***	**22.6**	**5.4**	**1.06**	**0.001**
	Insecta	**Plecoptera**	**Perlodidae**	***Isoperla***	**21.1**	**4.8**	**1.16**	**0.001**
	Insecta	**Coleoptera**	**Elmidae**	***Oulimnius***	**21.1**	**4.2**	**1.11**	**0.001**
	Insecta	**Plecoptera**	**Pteronarcyidae**	***Pteronarcys***	**20.4**	**3.1**	**1.10**	**0.001**
	Insecta	**Trichoptera**	**Rhyacophilidae**	***Rhyacophila***	**20.2**	**5.3**	**1.10**	**0.001**
	Insecta	**Trichoptera**	**Hydropsychidae**	***Diplectrona***	**18.0**	**4.4**	**1.12**	**0.001**
	Insecta	**Plecoptera**	**Perlidae**	***Acroneuria***	**16.3**	**4.7**	**1.07**	**0.001**
	Insecta	**Ephemeroptera**	**Heptageniidae**	***Cinygmula***	**14.2**	**2.7**	**1.08**	**0.001**
	Insecta	**Plecoptera**	**Chloroperlidae**	***Sweltsa***	**13.9**	**2.6**	**1.12**	**0.001**
	Insecta	**Diptera**	**Tipulidae**	***Hexatoma***	**13.0**	**5.1**	**1.17**	**0.001**
	Insecta	**Ephemeroptera**	**Leptophlebiidae**	***Paraleptophlebia***	**12.5**	**4.3**	**1.18**	**0.002**
	Insecta	**Odonata**	**Gomphidae**	***Lanthus***	**12.3**	**2.9**	**1.01**	**0.001**
	Insecta	**Diptera**	**Tipulidae**	***Dicranota***	**12.0**	**4.3**	**1.11**	**0.001**
	Insecta	**Trichoptera**	**Uenoidae**	***Neophylax***	**11.0**	**4.1**	**1.11**	**0.001**
	Insecta	**Diptera**	**Empididae**	***Chelifera***	**9.7**	**3.1**	**1.12**	**0.001**
	Insecta	**Trichoptera**	**Hydropsychidae**	***Parapsyche***	**7.1**	**1.6**	**1.05**	**0.003**
	Insecta	**Coleoptera**	**Psephenidae**	***Ectopria***	**6.7**	**2.7**	**1.04**	**0.009**
	Insecta	**Plecoptera**	**Perlodidae**	***Yugus***	**5.1**	**1.6**	**0.99**	**0.011**
	Insecta	**Plecoptera**	**Chloroperlidae**	***Haploperla***	**5.1**	**2.4**	**1.10**	**0.021**
	Insecta	Plecoptera	Peltoperlidae	*Tallaperla*	3.5	2.0	1.04	0.085
	Insecta	Plecoptera	Peltoperlidae	*Peltoperla*	3.4	1.8	1.12	0.062
	Insecta	Plecoptera	Chloroperlidae	*Suwallia*	3.1	1.6	1.02	0.087
	Insecta	Trichoptera	Odontoceridae	*Psilotreta*	2.4	2.0	1.19	0.242
	Insecta	Ephemeroptera	Heptageniidae	*Nixe*	2.0	1.7	1.10	0.252
	Insecta	Diptera	Blephariceridae	*Blepharicera*	1.7	1.5	1.02	0.334
	Insecta	Plecoptera	Perlodidae	*Cultus*	1.0	1.6	1.15	0.709

Appendix 8 (Cont'd)

High Quality Headwater Stream Community	**Insecta**	**Plecoptera**	**Nemouridae**	***Amphinemura***	**19.5**	**6.0**	**1.06**	**0.001**
	Insecta	**Trichoptera**	**Lepidostomatidae**	***Lepidostoma***	**16.4**	**3.7**	**1.14**	**0.001**
	Insecta	**Plecoptera**	**Leuctridae**	***Leuctra***	**15.8**	**6.3**	**1.06**	**0.001**
	Insecta	**Diptera**	**Simuliidae**	***Prosimulium***	**12.4**	**3.2**	**1.09**	**0.001**
	Insecta	**Diptera**	**Tabanidae**	***Chrysops***	**8.4**	**2.6**	**1.24**	**0.008**
	Insecta	**Diptera**	**Tipulidae**	***Limnophila***	**8.0**	**1.9**	**1.12**	**0.004**
	Insecta	**Trichoptera**	**Limnephilidae**	***Pycnopsyche***	**7.2**	**3.4**	**1.08**	**0.010**
	Insecta	**Diptera**	**Tipulidae**	***Pseudolimnophila***	**7.1**	**2.3**	**0.98**	**0.005**
	Insecta	**Odonata**	**Cordulegastridae**	***Cordulegaster***	**5.4**	**2.0**	**1.10**	**0.016**
	Insecta	**Plecoptera**	**Nemouridae**	***Ostrocerca***	**4.7**	**1.7**	**1.04**	**0.016**
	Insecta	Diptera	Tabanidae	*Tabanus*	4.3	2.0	1.11	0.052
	Insecta	**Diptera**	**Tipulidae**	***Molophilus***	**3.9**	**1.8**	**1.11**	**0.040**
	Insecta	Diptera	Empididae	*Clinocera*	3.3	2.4	1.08	0.123
	Insecta	Diptera	Tipulidae	*Ormosia*	2.3	1.5	0.98	0.137
	Insecta	Plecoptera	Perlodidae	*Diploperla*	1.8	1.5	0.97	0.262
	Insecta	Trichoptera	Psychomyiidae	*Lype*	1.8	1.7	1.03	0.353
	Insecta	Ephemeroptera	Ephemeridae	*Litobrancha*	1.7	1.6	1.16	0.318
	Insecta	Odonata	Calopterygidae	*Calopteryx*	1.5	1.8	1.21	0.463
	Insecta	Coleoptera	Hydrophilidae	*Hydrobius*	0.7	1.6	1.08	0.895

Appendix 8 (Cont'd)

High Quality Large Stream Community	**Insecta**	**Ephemeroptera**	**Ephemerellidae**	***Drunella***	**30.1**	**4.6**	**1.12**	**0.001**
	Insecta	**Ephemeroptera**	**Baetidae**	***Acentrella***	**23.6**	**4.1**	**1.09**	**0.001**
	Insecta	**Ephemeroptera**	**Ephemerellidae**	***Serratella***	**17.6**	**3.4**	**1.05**	**0.001**
	Insecta	**Ephemeroptera**	**Ephemerellidae**	***Ephemerella***	**17.6**	**6.6**	**1.03**	**0.001**
	Insecta	**Ephemeroptera**	**Heptageniidae**	***Leucrocuta***	**17.2**	**2.8**	**1.17**	**0.001**
	Insecta	**Trichoptera**	**Philopotamidae**	***Dolophilodes***	**16.9**	**4.8**	**1.03**	**0.001**
	Insecta	**Trichoptera**	**Hydropsychidae**	***Ceratopsyche***	**14.1**	**3.5**	**1.10**	**0.001**
	Insecta	**Ephemeroptera**	**Baetidae**	***Baetis***	**12.8**	**6.6**	**0.97**	**0.001**
	Insecta	**Trichoptera**	**Glossosomatidae**	***Agapetus***	**10.0**	**2.6**	**1.17**	**0.002**
	Insecta	**Trichoptera**	**Polycentropodidae**	***Polycentropus***	**8.8**	**4.1**	**1.17**	**0.003**
	Insecta	**Coleoptera**	**Elmidae**	***Promoresia***	**8.4**	**3.2**	**1.15**	**0.004**
	Insecta	**Plecoptera**	**Perlidae**	***Paragnetina***	**7.9**	**2.3**	**1.11**	**0.007**
	Insecta	**Diptera**	**Tipulidae**	***Antocha***	**7.2**	**4.2**	**1.08**	**0.022**
	Insecta	Trichoptera	Philopotamidae	*Wormaldia*	4.5	2.5	1.02	0.054
	Insecta	Trichoptera	Brachycentridae	*Micrasema*	3.8	2.3	1.20	0.075
	Insecta	Ephemeroptera	Ephemerellidae	*Attenella*	3.5	1.7	1.10	0.071
	Insecta	Ephemeroptera	Heptageniidae	*Heptagenia*	3.3	2.0	1.03	0.086
	Insecta	Odonata	Gomphidae	*Ophiogomphus*	3.0	1.8	1.04	0.093
	Insecta	Ephemeroptera	Ephemerellidae	*Dannella*	2.7	1.7	1.03	0.148
	Insecta	Coleoptera	Dryopidae	*Helichus*	2.4	1.9	1.11	0.238
	Insecta	Ephemeroptera	Baetidae	*Cloeon*	1.8	1.7	1.07	0.303
	Insecta	Ephemeroptera	Baetidae	*Acerpenna*	1.7	1.5	0.99	0.318
Sluggish Headwater Stream Community	**Gastropoda**	**Basommatophora**	**Physidae**		**38.4**	**2.6**	**1.05**	**0.001**
	Hirudinea				**12.2**	**2.2**	**1.16**	**0.001**
	Gastropoda	**Basommatophora**	**Planorbidae**		**10.2**	**1.7**	**1.09**	**0.001**
	Insecta	**Diptera**	**Chironomidae**		**8.8**	**8.7**	**0.29**	**0.026**
	Insecta	**Coleoptera**	**Dytiscidae**	***Agabus***	**7.7**	**1.8**	**1.11**	**0.004**
	Gastropoda	**Basommatophora**	**Lymnaeidae**		**5.6**	**1.6**	**0.98**	**0.009**
	Insecta	Coleoptera	Haliplidae	*Peltodytes*	2.8	1.5	1.10	0.092
	Insecta	Coleoptera	Hydrophilidae	*Tropisternus*	2.0	1.6	1.06	0.210

Appendix 8 (Cont'd)

Common Large Stream Community	**Insecta**	**Coleoptera**	**Elmidae**	***Dubiraphia***	**27.9**	**3.1**	**1.10**	**0.001**
	Insecta	**Ephemeroptera**	**Caenidae**	***Caenis***	**22.3**	**2.9**	**1.10**	**0.001**
	Insecta	**Coleoptera**	**Elmidae**	***Optioservus***	**15.3**	**5.7**	**1.11**	**0.001**
	Insecta	**Ephemeroptera**	**Ephemeridae**	***Ephemera***	**15.3**	**2.9**	**1.05**	**0.001**
	Insecta	**Diptera**	**Ceratopogonidae**	***Probezzia***	**12.1**	**1.8**	**1.07**	**0.001**
	Insecta	**Plecoptera**	**Perlidae**	***Perlesta***	**10.3**	**2.5**	**1.07**	**0.002**
	Insecta	**Ephemeroptera**	**Ephemerellidae**	***Eurylophella***	**8.2**	**4.0**	**1.12**	**0.011**
	Insecta	**Ephemeroptera**	**Heptageniidae**	***Stenacron***	**7.9**	**3.7**	**1.14**	**0.008**
	Insecta	**Trichoptera**	**Hydroptilidae**	***Ochrotrichia***	**5.3**	**1.6**	**1.14**	**0.019**
	Insecta	**Diptera**	**Ceratopogonidae**	***Bezzia***	**4.9**	**1.8**	**1.02**	**0.018**
	Insecta	Ephemeroptera	Baetidae	*Centroptilum*	4.6	2.4	1.08	0.055
	Insecta	**Trichoptera**	**Helicopsychidae**	***Helicopsyche***	**4.2**	**1.5**	**1.05**	**0.033**
	Insecta	Ephemeroptera	Leptophlebiidae	*Habrophlebiodes*	3.3	2.2	1.08	0.140
	Acarina				2.5	2.3	1.04	0.326
Limestone / Agricultural Stream Community	**Crustacea**	**Isopoda**			**24.3**	**3.7**	**1.15**	**0.001**
	Oligochaeta				**11.9**	**6.9**	**0.97**	**0.001**
	Bivalvia	Veneroida	Sphaeriidae		5.1	3.0	1.18	0.062
	Insecta	Diptera	Tipulidae	*Pilaria*	2.8	1.6	1.05	0.110
	Insecta	Coleoptera	Elmidae	*Macronychus*	2.5	1.6	1.03	0.143
Small Urban Stream Community	**Insecta**	**Trichoptera**	**Hydropsychidae**	***Cheumatopsyche***	**20.8**	**5.5**	**1.11**	**0.001**
	Insecta	**Coleoptera**	**Elmidae**	***Stenelmis***	**17.4**	**5.2**	**1.13**	**0.001**
	Insecta	**Diptera**	**Simuliidae**	***Simulium***	**16.2**	**5.8**	**1.14**	**0.001**
	Insecta	**Diptera**	**Empididae**	***Hemerodromia***	**11.9**	**3.2**	**1.08**	**0.001**
	Insecta	**Coleoptera**	**Elmidae**	***Ancyronyx***	**7.1**	**1.5**	**0.94**	**0.002**
	Insecta	**Odonata**	**Aeshnidae**	***Boyeria***	**6.2**	**2.2**	**1.15**	**0.016**
	Insecta	**Odonata**	**Coenagrionidae**	***Enallagma***	**6.0**	**1.6**	**1.17**	**0.015**
	Insecta	Coleoptera	Dytiscidae	*Hydroporus*	3.4	2.1	1.11	0.119
	Oligochaeta	Tubificida	Tubificidae	*Limnodrilus*	2.3	1.6	1.05	0.142
	Bivalvia	Veneroida	Corbiculidae		1.8	1.6	0.92	0.221
Forested Headwater Stream Community	**Insecta**	**Plecoptera**	**Chloroperlidae**	***Alloperla***	**15.6**	**2.1**	**1.11**	**0.001**
	Insecta	**Diptera**	**Tipulidae**	***Tipula***	**9.9**	**4.4**	**1.09**	**0.003**
	Insecta	**Ephemeroptera**	**Ameletidae**	***Ameletus***	**8.9**	**2.9**	**1.10**	**0.004**
	Insecta	Diptera	Tipulidae	*Pedicia*	3.6	1.6	1.12	0.063
	Insecta	Odonata	Gomphidae	*Gomphus*	1.5	1.8	1.10	0.457

Appendix 8 (Cont'd)

Common Small Stream Community	**Insecta**	**Ephemeroptera**	**Heptageniidae**	***Stenonema***	**21.5**	**5.7**	**1.12**	**0.001**
	Insecta	**Coleoptera**	**Psephenidae**	***Psephenus***	**18.6**	**4.6**	**1.07**	**0.001**
	Insecta	**Trichoptera**	**Philopotamidae**	***Chimarra***	**14.5**	**3.3**	**1.13**	**0.001**
	Insecta	**Trichoptera**	**Glossosomatidae**	***Glossosoma***	**11.1**	**2.5**	**1.13**	**0.002**
	Insecta	**Ephemeroptera**	**Isonychiidae**	***Isonychia***	**10.6**	**3.7**	**1.12**	**0.003**
	Insecta	**Trichoptera**	**Hydropsychidae**	***Macrostemum***	**9.8**	**1.6**	**1.05**	**0.002**
	Insecta	**Megaloptera**	**Sialidae**	***Sialis***	**9.1**	**3.3**	**1.07**	**0.002**
	Insecta	**Odonata**	**Gomphidae**	***Stylogomphus***	**8.2**	**2.2**	**1.09**	**0.003**
	Insecta	**Megaloptera**	**Corydalidae**	***Corydalus***	**7.2**	**1.9**	**1.02**	**0.005**
	Insecta	**Odonata**	**Gomphidae**	***Arigomphus***	**4.7**	**1.6**	**1.01**	**0.025**
	Insecta	**Plecoptera**	**Perlidae**	***Eccoptura***	**4.6**	**2.0**	**1.11**	**0.041**
	Insecta	Plecoptera	Perlidae	*Agnetina*	3.6	2.3	1.25	0.093
	Gastropoda	Mesogastropoda	Pleuroceridae		3.0	1.6	1.10	0.102
	Gastropoda	Basommatophora	Ancylidae		2.3	1.9	1.21	0.243
	Insecta	Plecoptera	Nemouridae	*Prostoia*	1.9	1.5	0.91	0.235
	Insecta	Coleoptera	Ptilodactylidae	*Anchytarsus*	1.9	1.6	1.00	0.237
	Insecta	Plecoptera	Capniidae	*Allocapnia*	1.7	1.5	1.09	0.306
	Insecta	Trichoptera	Goeridae	*Goera*	1.4	1.6	1.07	0.410
	Insecta	Trichoptera	Psychomyiidae	*Psychomyia*	1.2	1.6	1.09	0.597
Ohio River Community	**Insecta**	**Trichoptera**	**Polycentropodidae**	***Cyrnellus***	**62.6**	**1.9**	**0.91**	**0.001**
	Crustacea	**Amphipoda**			**34.4**	**3.9**	**1.18**	**0.001**
	Insecta	**Trichoptera**	**Hydroptilidae**	***Hydroptila***	**30.1**	**2.8**	**1.14**	**0.001**
	Insecta	**Ephemeroptera**	**Leptohyphidae**	***Tricorythodes***	**17.3**	**1.8**	**1.05**	**0.001**
	Insecta	**Odonata**	**Coenagrionidae**	***Argia***	**6.6**	**2.0**	**1.05**	**0.010**
Mixed Land Use Stream Community	**Insecta**	**Trichoptera**	**Hydropsychidae**	***Hydropsyche***	**20.2**	**6.0**	**1.04**	**0.001**
	Insecta	**Megaloptera**	**Corydalidae**	***Nigronia***	**11.0**	**4.5**	**1.08**	**0.001**
	Insecta	**Ephemeroptera**	**Leptophlebiidae**	***Leptophlebia***	**6.2**	**1.8**	**1.21**	**0.017**
	Insecta	**Diptera**	**Athericidae**	***Atherix***	**5.7**	**2.2**	**1.19**	**0.025**
	Insecta	Ephemeroptera	Leptophlebiidae	*Habrophlebia*	1.2	1.6	1.03	0.588
Large Stream Generalist Community	No significant indicators were found with Indicator Species Analysis (Chironomidae and Oligochaeta were commonly associated with the group)							

Appendix 9. Indicator Species Analysis results for family-level macroinvertebrate communities. Indicator values and randomized Monte-Carlo simulation mean indicator values, standard deviation, and p-values are presented. Significant indicator taxa ($p < 0.05$) are in bold type.

Community Name	Phylum/Class	Order	Family	Indicator Value	Randomized Mean Ind Value	Randomized IV Std Dev	Monte-Carlo p-value
Low Gradient Valley Stream Community	**Insecta**	**Coleoptera**	**Elmidae**	**30.1**	**8.6**	**0.86**	**0.001**
	Insecta	**Coleoptera**	**Psephenidae**	**24.0**	**6.5**	**0.85**	**0.001**
	Insecta	**Trichoptera**	**Hydropsychidae**	**16.6**	**11.7**	**0.64**	**0.001**
	Bivalvia	**Veneroida**	**Corbiculidae**	**10.8**	**0.8**	**0.44**	**0.001**
	Insecta	**Odonata**	**Coenagrionidae**	**7.5**	**1.0**	**0.49**	**0.001**
	Insecta	**Ephemeroptera**	**Caenidae**	**2.9**	**0.9**	**0.50**	**0.010**
	Insecta	**Odonata**	**Calopterygidae**	**2.2**	**0.7**	**0.43**	**0.011**
	Gastropoda	**Basommatophora**	**Ancylidae**	**2.7**	**0.9**	**0.50**	**0.018**
	Bivalvia	**Veneroida**	**Sphaeriidae**	**2.8**	**1.6**	**0.59**	**0.039**
	Gastropoda	Basommatophora	Lymnaeidae	1.2	0.6	0.43	0.051
	Insecta	Diptera	Tabanidae	1.6	1.6	0.63	0.254
	Insecta	Trichoptera	Hydroptilidae	0.5	0.7	0.46	0.569
High Quality Small Stream Community	**Insecta**	**Ephemeroptera**	**Isonychiidae**	**22.7**	**3.7**	**0.75**	**0.001**
	Insecta	**Trichoptera**	**Philopotamidae**	**20.8**	**8.2**	**0.83**	**0.001**
	Insecta	**Megaloptera**	**Corydalidae**	**13.4**	**7.0**	**0.84**	**0.001**
	Insecta	**Diptera**	**Athericidae**	**4.7**	**1.5**	**0.60**	**0.005**
	Insecta	**Trichoptera**	**Glossosomatidae**	**5.5**	**2.5**	**0.71**	**0.008**
	Insecta	Ephemeroptera	Ephemeridae	3.1	2.3	0.62	0.085
	Insecta	Trichoptera	Helicopsychidae	1.3	0.7	0.50	0.094
	Insecta	Coleoptera	Gyrinidae	0.8	0.6	0.44	0.167
Common Headwater Stream Community	**Insecta**	**Trichoptera**	**Lepidostomatidae**	**3.5**	**1.3**	**0.51**	**0.010**
	Insecta	**Plecoptera**	**Capniidae**	**3.3**	**1.4**	**0.54**	**0.017**
	Insecta	Odonata	Cordulegastridae	1.5	1.0	0.53	0.093
Limestone / Agricultural Stream Community	**Crustacea**	**Amphipoda**		**21.8**	**4.0**	**0.74**	**0.001**
	Insecta	**Diptera**	**Simuliidae**	**20.3**	**7.1**	**0.87**	**0.001**
	Crustacea	**Isopoda**	**Asellidae**	**19.9**	**3.0**	**0.76**	**0.001**
	Turbellaria			**17.4**	**3.6**	**0.76**	**0.001**
	Annelida			**15.5**	**8.5**	**0.84**	**0.001**
	Gastropoda	**Basommatophora**	**Physidae**	**10.0**	**2.1**	**0.65**	**0.001**
	Insecta	**Coleoptera**	**Dytiscidae**	**7.9**	**1.7**	**0.55**	**0.001**

Appendix 9 (Cont’)

Limestone / Agricultural Stream Community	**Insecta**	**Diptera**	**Chironomidae**	**13.5**	**11.5**	**0.67**	**0.013**
	Gastropoda	**Basommatophora**	**Planorbidae**	**3.0**	**0.9**	**0.50**	**0.013**
	Insecta	**Coleoptera**	**Hydrophilidae**	**2.2**	**0.9**	**0.51**	**0.031**
High Quality Headwater Stream Community	**Insecta**	**Plecoptera**	**Leuctridae**	**25.6**	**6.6**	**0.78**	**0.001**
	Insecta	**Ephemeroptera**	**Baetidae**	**18.0**	**8.9**	**0.79**	**0.001**
	Crustacea	**Decapoda**	**Cambaridae**	**14.5**	**9.5**	**0.83**	**0.001**
	Insecta	**Trichoptera**	**Polycentropodidae**	**8.3**	**3.4**	**0.75**	**0.001**
	Insecta	Odonata	Aeshnidae	3.5	2.3	0.67	0.058
	Insecta	Diptera	Ceratopogonidae	1.4	1.0	0.52	0.108
	Arachnida	Hydracarina		0.8	0.7	0.47	0.271
Common Large Stream Community	**Insecta**	**Plecoptera**	**Nemouridae**	**19.5**	**6.1**	**0.83**	**0.001**
	Insecta	**Ephemeroptera**	**Ameletidae**	**19.3**	**2.6**	**0.66**	**0.001**
	Insecta	**Plecoptera**	**Taeniopterygidae**	**4.8**	**1.1**	**0.52**	**0.003**
	Insecta	Trichoptera	Leptoceridae	0.6	0.6	0.43	0.313
High Quality Mid-Sized Stream Community	**Insecta**	**Plecoptera**	**Chloroperlidae**	**33.5**	**4.2**	**0.81**	**0.001**
	Insecta	**Plecoptera**	**Pteronarcyidae**	**32.4**	**2.5**	**0.63**	**0.001**
	Insecta	**Ephemeroptera**	**Ephemerellidae**	**29.3**	**8.6**	**0.81**	**0.001**
	Insecta	**Ephemeroptera**	**Heptageniidae**	**27.7**	**9.9**	**0.79**	**0.001**
	Insecta	**Trichoptera**	**Rhyacophilidae**	**26.6**	**6.2**	**0.81**	**0.001**
	Insecta	**Plecoptera**	**Perlodidae**	**26.4**	**6.4**	**0.85**	**0.001**
	Insecta	**Ephemeroptera**	**Leptophlebiidae**	**26.4**	**7.3**	**0.84**	**0.001**
	Insecta	**Plecoptera**	**Perlidae**	**19.1**	**9.0**	**0.77**	**0.001**
	Insecta	**Diptera**	**Tipulidae**	**15.6**	**10.6**	**0.75**	**0.001**
	Insecta	**Plecoptera**	**Peltoperlidae**	**15.4**	**2.6**	**0.70**	**0.001**
	Insecta	**Odonata**	**Gomphidae**	**13.0**	**4.5**	**0.77**	**0.001**
	Insecta	**Trichoptera**	**Limnephilidae**	**11.2**	**4.3**	**0.73**	**0.001**
	Insecta	**Trichoptera**	**Uenoidae**	**8.1**	**2.9**	**0.74**	**0.001**
	Insecta	**Trichoptera**	**Odontoceridae**	**3.6**	**0.9**	**0.44**	**0.005**
	Insecta	Trichoptera	Brachycentridae	1.6	1.2	0.55	0.131
	Insecta	Coleoptera	Ptilodactylidae	1.0	0.7	0.49	0.161
	Insecta	Trichoptera	Psychomyiidae	0.4	0.6	0.44	0.662
AMD Stream Community	**Insecta**	**Megaloptera**	**Sialidae**	**5.2**	**2.3**	**0.72**	**0.009**
	Insecta	Diptera	Empididae	1.9	1.2	0.50	0.080

Appendix 10. Importance values of Random Forest models by model type. A community predictive model was developed for mussel communities, fish communities, and macroinvertebrate communities for family and genus-level datasets. Importance values ≥ 1 are in bold type. (Label codes are defined as follows: OhEr = Ohio – Great Lakes Basins, SP = Susquehanna – Potomac River Basins, DE = Delaware River Basin, At = Atlantic Basin, Fish = fish community, Muss = mussel community, MI = macroinvertebrate community)

Variable Code	Variable	OhErMuss	SPMuss	DEMuss	OhErFish	AtFish	MIGenus	MIFamily
ABIOCLASS	Physical stream class combination of geology class, gradient class, and watershed area class	0.22	0.62	0.64	0.44	0.65	0.55	0.61
ARBOLATE_2	Total stream miles in catchment	**1.46**	0.13	0.59	**1.27**	**1.15**	**1.06**	0.53
AVGELV	Average reach elevation	**1.79**	0.73	**1.75**	**1.00**	**1.42**	**1.35**	**1.03**
D_LINK	Number of downstream links (first order streams)	**1.78**	0.95	**1.23**	0.91	0.76	0.63	0.34
DAMACCUM	Number of upstream dams in catchment	**1.54**	0.60	0.56	0.76	0.83	0.66	0.21
DAMDENS	Density of dams in catchment	0.83	0.63	0.79	0.63	0.89	0.36	0.22
DAMS_12	Number of reach dams and reach dam storage capacity	0.07	0.00	0.00	-0.18	0.08	0.04	-0.04
DAMSTACCUM	Accumulated dam storage in upstream watershed	**1.51**	0.57	0.21	**1.15**	0.92	0.53	0.26
DAMSTDENS	Density of dams * storage capacity in catchment	0.47	0.08	-0.09	0.87	0.80	0.31	0.15
DAMSTORA_2	Dam storage in reach watershed	-0.03	0.00	0.00	-0.13	0.08	0.03	-0.01
DOMLOCGEO	Dominant geology class in reach watershed	0.26	**1.10**	0.10	0.35	0.19	0.76	0.61
DOMUPSGEO	Dominant geology class in catchment	-0.12	0.03	0.39	0.29	0.52	0.72	0.57
F_ELV	Upstream reach elevation	**1.60**	0.33	**1.36**	0.99	**1.44**	**1.34**	**1.00**
GRAD_CLASS	Class of stream reach gradient	0.04	-0.13	0.30	0.36	0.71	0.61	0.42
GRADIENT	Stream reach gradient	0.73	-0.20	-0.07	0.63	**1.06**	**1.09**	0.66
LINK	Number of upstream links (first order streams)	**1.16**	0.73	0.50	**1.22**	**1.12**	0.84	0.30

Appendix 10 (Cont'd)

Variable Code	Variable	OhErMuss	SPMuss	DEMuss	OhErFish	AtFish	MIGenus	MIFamily
LOCALGEO1	% sandstone geology class in reach watershed	-0.18	**1.03**	0.50	0.63	0.58	0.85	0.53
LOCALGEO2	% shale geology class in reach watershed	0.11	0.97	0.30	0.50	0.65	0.51	0.51
LOCALGEO3	% calcareous geology class in reach watershed	0.29	**1.00**	0.10	0.21	0.39	0.57	0.71
LOCALGEO4	% crystalline silicic geology class in reach watershed	0.00	0.00	0.00	0.00	0.20	0.20	0.15
LOCALGEO5	% crystalline mafic geology class in reach watershed	0.00	0.05	0.00	0.00	0.21	0.18	0.37
LOCALGEO6	% unconsolidated materials geology class in reach watershed	0.10	0.00	0.00	-0.02	-0.19	0.14	0.00
PC_COMMIND	% commercial/industrial/ transportation in catchment	**1.11**	0.47	0.01	0.82	**1.05**	**1.02**	0.63
PC_DECFOR	% deciduous forest in catchment	**1.47**	0.05	0.67	0.81	0.94	**1.30**	0.73
PC_EMRWET	% emergent wetland in catchment	0.63	0.65	0.09	0.67	**1.24**	0.86	0.48
PC_EVEFOR	% evergreen forest in catchment	**1.69**	0.67	-0.07	0.65	0.74	0.84	0.59
PC_GRASS	% grassland in catchment	-0.30	0.00	0.00	0.06	0.00	0.00	0.00
PC_HIGHURB	% high intensity residential in catchment	**1.53**	**1.23**	0.98	0.62	**1.03**	**1.33**	0.58
PC_LOWURB	% low intensity residential in catchment	**1.47**	0.92	**1.10**	0.70	**1.08**	**1.42**	0.90
PC_MIXFOR	% mixed forest in catchment	0.83	0.57	**1.16**	0.69	0.96	0.69	0.48
PC_NONRCAG	% non-row crop agriculture in catchment	**1.99**	0.10	0.18	0.65	**1.09**	**1.08**	0.83
PC_OPNWATR	% open water in catchment	**1.07**	0.42	0.08	0.73	**1.16**	0.74	0.46
PC_ORCH	% orchard in catchment	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Appendix 10 (Cont'd)

Variable Code	Variable	OhErMuss	SPMuss	DEMuss	OhErFish	AtFish	MIGenus	MIFamily
PC_PASTURE	% pasture/hay in catchment	**1.99**	-0.10	0.18	0.85	**1.09**	0.79	0.86
PC_QUARMN	% quarries/stripmines/gravel pits in catchment	**1.21**	0.33	0.09	0.56	0.71	0.84	0.77
PC_ROCK	% bare rock/sand/clay in catchment	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PC_ROWCROP	% agriculture in row crops in catchment	0.61	0.85	0.40	0.71	0.91	0.72	0.72
PC_SCRUB	% scrubland in catchment	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PC_SMGRAIN	% small grains in catchment	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PC_TOTAG2	% agriculture in catchment	**1.30**	0.10	0.18	0.84	**1.08**	0.83	0.90
PC_TOTFOR2	% forest in catchment	0.76	-0.17	0.65	0.87	**1.15**	**1.98**	0.81
PC_TOTURB2	% urban in catchment	**1.49**	0.35	0.05	0.74	**1.07**	**1.38**	0.88
PC_TRANS	% transitional in catchment	0.74	-0.20	**1.11**	0.50	0.89	0.48	0.47
PC_URBREC	% urban/recreational grasses in catchment	**1.30**	**1.57**	0.00	0.28	0.72	0.39	0.32
PC_WDYWET	% woody wetland in catchment	**1.14**	0.12	0.47	0.64	**1.09**	0.46	0.39
PCTOTWETL2	% wetland in catchment	0.35	-0.43	0.05	0.70	**1.20**	0.76	0.48
PS_ACCUM	Number of point sources in catchment	0.00	0.50	0.74	0.00	**1.04**	0.64	0.40
PSDENSITY	Density of point sources in the reach watershed	0.85	0.00	0.56	0.70	0.78	0.68	0.50
PTSOURCE_2	Number of point sources in the reach watershed	**1.38**	-0.33	0.00	0.21	0.00	0.00	0.11

Appendix 10 (Cont'd)

Variable Code	Variable	OhErMuss	SPMuss	DEMuss	OhErFish	AtFish	MIGenus	MIFamily
RDSTR_DENS	Density of road – stream crossings in the reach watershed	0.95	0.80	**1.23**	0.58	0.97	0.97	0.47
RDSTRXINGS	Number of reach road – stream crossings, density of reach road – stream crossings	-0.13	-0.30	0.48	-0.09	0.35	0.19	0.29
REFSEG2	Class of stream quality (reference condition or non-ref condition)	0.50	0.23	0.00	0.45	0.49	0.37	0.26
RIP_AG	% agriculture in reach riparian zone	**1.10**	-0.42	0.00	0.61	0.70	0.39	0.83
RIP_BARREN	% barren in reach riparian zone	0.18	0.00	0.00	0.13	-0.06	0.13	0.10
RIP_DEVEL	% developed in reach riparian zone	0.75	0.05	0.16	0.38	0.27	0.49	0.43
RIP_FOREST	% forest in reach riparian zone	**1.94**	-0.70	-0.10	0.62	**1.04**	0.74	0.80
RIP_WATER	% open water in reach riparian zone	**1.29**	**1.20**	-0.37	0.54	0.87	0.37	0.19
RIP_WETL	% wetland in reach riparian zone	**1.24**	0.05	-0.43	0.08	0.59	0.26	0.21
RSC_ACCUM	Road – stream crossings in catchment	0.00	0.62	0.77	0.00	**1.13**	**1.28**	0.61
RSC_DENSIT	Density of road – stream crossings in reach watershed	0.00	0.45	**1.11**	0.00	**1.05**	**1.21**	0.48
SQMI	Watershed area (mi^2)	**1.41**	0.12	**1.09**	**1.18**	**1.07**	**1.12**	0.40
STRORDER	Strahler stream order of reach	**1.39**	-0.20	0.48	0.70	0.86	0.57	0.28
T_ELV	Downstream reach elevation	**1.35**	0.47	**1.72**	0.87	**1.31**	**1.28**	0.91
TOT_BARERO	Area of bare rock/sand/clay in catchment	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TOT_COMM_I	Area of commercial/industrial/transportation in catchment	**1.33**	0.03	0.18	0.86	**1.02**	0.97	0.65

Appendix 10 (Cont'd)

Variable Code	Variable	OhErMuss	SPMuss	DEMuss	OhErFish	AtFish	MIGenus	MIFamily
TOT_DECFOR	Area of deciduous forest in catchment	**2.00**	0.20	0.29	**1.20**	**1.05**	**1.15**	0.51
TOT_EMERWE	Area of emergent wetland in catchment	0.94	0.37	0.53	0.99	**1.26**	**1.06**	0.54
TOT_EVEFOR	Area of evergreen forest in catchment	**1.99**	0.42	0.57	0.86	0.96	**1.05**	0.51
TOT_GRASS	Area of grassland in catchment	-0.18	0.00	0.00	0.03	0.00	0.04	0.00
TOT_HIGHIN	Area of high intensity residential in catchment	**1.52**	0.40	0.84	0.74	**1.08**	**1.20**	0.63
TOT_LOWINT	Area of low intensity residential in catchment	**1.33**	0.08	**1.08**	0.78	**1.13**	**1.06**	0.82
TOT_MIXFOR	Area of mixed forest in catchment	**1.80**	0.03	**1.11**	0.90	**1.03**	0.74	0.39
TOT_OPENWA	Area of open water in catchment	**1.52**	**1.03**	0.65	0.99	**1.26**	**1.05**	0.45
TOT_ORCH	Area of orchard in catchment	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TOT_PASTUR	Area of pasture/hay in catchment	**1.55**	-0.18	-0.24	**1.32**	**1.31**	**1.27**	0.76
TOT_QUARMI	Area of quarries/stripmines/gravel pits in catchment	**2.01**	0.28	-0.38	0.63	0.77	**1.04**	0.70
TOT_ROWCRO	Area of agriculture in row crops in catchment	**1.62**	0.55	0.55	**1.08**	**1.14**	0.99	0.65
TOT_SCRUB	Area of scrubland in catchment	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TOT_SMGRAI	Area of small grains in catchment	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TOT_TRANS	Area of transitional landcover in catchment	**1.35**	-0.60	0.50	0.85	0.98	0.82	0.52
TOT_URBREC	Area of urban/recreational grasses in catchment	**1.07**	0.38	0.46	0.39	0.88	0.52	0.33
TOT_WOODYW	Area of woody wetland in catchment	**1.58**	0.80	**1.22**	0.90	**1.22**	0.68	0.44

Appendix 10 (Cont'd)

Variable Code	Variable	OhErMuss	SPMuss	DEMuss	OhErFish	AtFish	MIGenus	MIFamily
UPSTRDSTR	Number of catchment road – stream crossings	**1.63**	0.30	**1.01**	**1.14**	**1.12**	0.71	0.63
UPSTRGEO1	% sandstone geology class in catchment	**1.35**	0.45	0.65	0.69	0.91	0.95	0.72
UPSTRGEO2	% shale geology class in catchment	0.80	-0.08	**1.36**	0.67	0.80	0.75	0.60
UPSTRGEO3	% calcareous geology class in catchment	**1.36**	**1.97**	0.00	0.33	0.62	0.69	0.66
UPSTRGEO4	% crystalline silicic geology class in catchment	0.00	0.00	0.00	0.00	0.49	0.34	0.16
UPSTRGEO5	% crystalline mafic geology class in catchment	0.00	0.20	0.00	0.00	0.39	0.49	0.46
UPSTRGEO6	% unconsolidated materials geology class in catchment	0.58	0.00	0.00	0.14	0.12	0.18	0.08
WSHEDCLASS	Watershed size class	0.49	0.00	0.84	0.86	0.58	0.51	0.15

Appendix 11 (a-g). Confusion matrices from Random Forest models of community occurrence for classifications of a) Ohio – Great Lakes Basins mussels, b) Susquehanna – Potomac River Basins mussels, c) Delaware River Basin mussels, d) Ohio – Great Lakes Basins fish, e) Atlantic Basin fish, f) genus-level macroinvertebrates, and g) family-level macroinvertebrates. The number of reaches correctly classified as community presence for each of the model communities and the percent class error are listed for each model. Class error rates < 40% are in bold type.

	Community name	Fatmucket	Spike	Fluted shell	Pink Heelsplitter	Class Error
a) Ohio – Great Lakes Basins Mussels	Fatmucket	33	2	8	2	**26.7%**
	Spike	9	9	6	0	62.5%
	Fluted shell	10	5	29	0	**34.1%**
	Pink Heelsplitter	3	0	0	0	100.0%

	Community name	Eastern Elliptio	Squawfoot	Eastern Floater	Yellow Lampmussel	Elktoe	Lanceolate Elliptio	Class Error
b) Susquehanna – Potomac River Basins Mussels	Eastern Elliptio	40	5	0	2	0	0	**14.9%**
	Squawfoot	16	4	2	0	0	0	81.8%
	Eastern Floater	1	2	0	1	0	0	100.0%
	Yellow Lampmussel	7	2	1	3	0	0	76.9%
	Elktoe	2	1	0	1	0	0	100.0%
	Lanceolate Elliptio	0	2	0	0	0	0	100.0%

Appendix 11 (a-g). (Cont’d)

c) Delaware River Basin Mussels

Community name	Eastern Elliptio	Alewife Floater	Other	Class Error
Eastern Elliptio	99	1	0	**1.0%**
Alewife Floater	1	0	0	100.0%
Other	1	0	1	50.0%

d) Ohio – Great Lakes Basins Fish

Community name	Coolwater	Warmwater	Coldwater	Large River	Class Error
Coolwater	528	35	71	3	**17.1%**
Warmwater	79	182	6	16	**35.7%**
Coldwater	99	2	168	0	**37.5%**
Large River	48	32	1	130	**38.4%**

Appendix 11 (a-g). (Cont’d)

	Community name	Warmwater 1	Warmwater 2	Coolwater 1	River & Impoundment	Coolwater 2	Coldwater	Lower Delaware River	Class Error
e) Atlantic Basin Fish	Warmwater 1	523	32	16	5	60	47	0	**23.4**%
	Warmwater 2	73	146	6	39	62	6	0	56.0%
	Coolwater 1	39	4	117	8	77	104	0	66.5%
	River & Impoundment	23	41	21	157	40	26	6	50.0%
	Coolwater 2	68	36	45	15	275	49	0	43.6%
	Coldwater	49	1	28	0	51	518	0	**19.9%**
	Lower Delaware River	1	0	0	7	0	0	25	**24.2%**

Appendix 11 (a-g). (Cont'd)

	Community name	High Quality Small	High Quality Headwater	High Quality Large	Sluggish Headwater	Common Large	Limestone / Agricultural	Small Urban Stream	Large Stream Generalist	Forested Headwater	Common Small	Ohio River	Mixed Land Use	Class Error
f) Macroinvertebrate – Genus	High Quality Small	73	12	26	0	1	0	0	0	0	7	0	0	**38.7%**
	High Quality Headwater	29	13	7	1	4	2	0	2	0	1	0	0	78.0%
	High Quality Large	26	4	101	1	1	2	0	0	0	6	0	0	**28.4%**
	Sluggish Headwater	1	3	2	1	1	7	6	1	0	3	0	0	96.0%
	Common Large	8	2	15	0	13	5	0	2	0	6	1	0	75.0%
	Limestone / Agricultural	5	4	9	2	2	30	2	0	0	8	0	0	51.6%
	Small Urban Stream	0	0	3	0	1	3	11	0	0	5	0	0	52.2%
	Large Stream Generalist	5	1	3	0	3	5	0	6	0	2	2	0	77.8%
	Forested Headwater	5	2	0	0	1	0	0	0	0	0	0	0	100.0%
	Common Small	4	0	12	1	3	2	0	1	0	37	0	0	**38.3%**
	Ohio River	0	0	0	0	1	0	0	1	0	0	21	0	**8.7%**
	Mixed Land Use	3	1	3	0	2	0	0	0	0	2	0	2	84.6%

Appendix 11 (a-g). (Cont'd)

	Community name	Low Gradient Valley	High Quality Small	Common Headwater	Limestone / Agricultural	High Quality Headwater	Common Large	High Quality Mid-Sized	AMD	Class Error
g) Macroinvertebrate – Family	Low Gradient Valley	164	103	0	22	12	32	3	0	51.2%
	High Quality Small	35	507	13	11	52	39	39	1	**27.3%**
	Common Headwater	12	40	75	6	22	18	83	0	70.7%
	Limestone / Agricultural	43	42	4	98	4	13	3	0	52.7%
	High Quality Headwater	14	111	26	3	124	50	43	0	66.6%
	Common Large	28	80	27	4	42	128	48	0	64.1%
	High Quality Mid-Sized	4	98	32	0	32	36	176	0	53.4%
	AMD	4	10	6	7	2	7	4	0	100.0%