**PiSCES: Pi(scine) Stream Community Estimation System**

Mike Cyterski, Justin Babendreier, Craig Barber, Mike Galvin, John M. Johnston, Lourdes Prieto, Luis Suarez and Kurt Wolfe

USEPA National Exposure Research Laboratory

Ecosystem Research Division

National Exposure Research Laboratory

Athens, GA

**Abstract**

The Piscine Stream Community Estimation System (PiSCES) provides users with a reasonable fish community for any stream reach in the conterminous United States. PiSCES utilizes distributional information obtained from Nature Serve, the USGS, and the Peterson Field Guide to Freshwater Fishes of North America for over 1,000 native and non-native freshwater fish species. In addition to these geographic distributions, users of PiSCES can filter potential fish assemblages based on species rarity, stream size preferences (as measured by mean stream width), and genera-specific occurrence envelopes for a variety of physicochemical stream properties (pH, conductivity, turbidity, dissolved oxygen, watershed area, mean width, depth, and slope). For a user-specified total number of fishes in the stream reach, PiSCES will create an abundance distribution for the posited community using a size spectrum approach that relates abundance to maximum body size. The PiSCES interface allows users to readily access its database to see ancillary characteristics of each species, such as habitat preferences (e.g., if the species prefers riffles, runs, or pools, sand or gravel substrate, etc.) and HUC-based distribution maps. PiSCES has been used in conjunction with an integrated environmental modeling system to simulate dynamics of stream fish communities in the Coal River, WV watershed.

**Introduction**

There are over 3.5 million miles of rivers and streams in the United States, and many contain fish. Fish abundance and diversity have been used as biological indicators of stream condition (e.g., indices of biotic integrity), in addition to macrobenthos metrics (e.g., EPT taxa) and standard water quality measurements (e.g., DO, pH, conductivity) (Karr 1981; Karr and Dudley 1981; Karr et al. 1986; Miller et al. 1988; Simon and Lyons 1995; Davis et al. 1996; Hughes and Oberdorff 1999). Because fish are longer lived than aquatic macroinvertebrates, they integrate a wide array of stream and watershed factors related to the quality of lotic systems on the order of years (Plafkin et al. 1989; Karr and Chu 1999; Wesche and Isaak 1999; Kovacs et al. 2002). Fish species are grouped into tolerant and intolerant categories based on known habitat affinities, with integrity indices developed for use within states and regions of interest (Fausch et al. 1990; Stepenuck et al. 2002; Simon 1998; Stoddard et al. 2006).

Given that a fish community reflects on the state of the system it occupies, government agencies (federal, state, local), non-governmental groups and academic researchers routinely conduct field sampling to ascertain what fishes are present (Yoder and Smith 1998; Angermeier and Winston 1999; Weaver and Garman 1994; Barbour et al. 1999; Wang et al. 2000; USEPA 2013). Sampling methods vary, but some of the most popular are backpack electrofishing (multiple pass and depletion), seining (electric and otherwise) and boat shocking.

Many factors determine where fish species are found, including physical stream habitat (e.g., flow, depth, woody debris presence and substrate quality), ecological interactions (including introduced and invasive species), levels of stressors (sedimentation, pH, temperature, nutrients) and historical biogeography (Fausch et al. 1990; Simon 1998; Jackson et al. 2001; Quist et al. 2004; Rashleigh et al. 2005; Bertolo and Magnan 2006; Kennard et al. 2007; Dauwalter et al. 2008; Pool et al. 2010). Fish distributions are not static but change through time as functions of life history, disturbance, seasonal and prevailing hydrologic regimes, species introductions, climate change and food availability (Shuter and Post 1990; Rahel 2000; Taylor and Warren Jr. 2001; Taylor et al. 2006; Taylor et al. 2008; Schaefer et al. 2012; Eros et al. 2014). Field sampling, however, can be time consuming, expensive, biased by gear selectivity and compromised by the dynamic nature of fish assemblages in lotic systems. Alternative methods of estimating fish species and abundance are needed to augment field sampling efforts, and a variety of models have been developed for this purpose (Baxter and Hauer 2000; Oakes et al. 2005; Fransen et al. 2006; McCleary and Hassan 2011). An empirical modeling approach derived from field data would have wide geographic application and utility for stream condition assessment.

Multivariate statistical methods have been used to estimate fish assemblages at a variety of locations (Kelso and Johnson 1991; Madejczyk et al. 1998; Ornellas and Coutinho 1998; Angermeier and Winston 1999; Saiki and Martin 2001; Kendrick and Francis 2002; Schweizer and Jager 2012). Wadeable stream sampling data collected by the USEPA in the Mid-Atlantic Highlands region of the eastern US (Herlihy et al. 2000) was used for analyses presented in Cyterski and Barber (2006) and McCormick et al. (2000). These methods and data have also been used to develop a decision support tool for fisheries management and stream habitat restoration (WHAT-IF, USEPA 2006). However, fish community data are not available in all locations, and not always internet-accessible even when present. Data availability, accessibility and quality are therefore barriers to the widespread use of fish as biological indicators.

To overcome the reliance on field data and the resources needed to compile and interpret fish sampling databases, an alternative approach to stream fish assemblage modeling was developed: the Piscine Stream Community Estimation System, or PiSCES. PiSCES is a tool for determining credible fish communities for streams and rivers across the conterminous US. Historic and current range data for over 1,000 native and non-native freshwater fish species were combined with information on species characteristics and habitat preferences (Page and Burr 2011; NatureServe 2010; FishBase.Org). PiSCES incorporates the results of analyses conducted on physicochemical stream data to determine genera-specific “occurrence envelopes,” which are used to modify the foundational geographic distributions for each species.

**PiSCES Background**

Motivation for the development of PISCES came from the need for its services in assisting integrated modeling approaches that attempt to predict effects of land use change on ecological function in aquatic systems. At the time of its inception, the authors were also engaged in development of two additional, large-scale software modeling projects. The first was development of USEPA's OpenTERRAworks Software System (OTW), a component system within USEPA’s iemTechnologies modeling platform which provides interoperability of software systems used to model environmental systems. OTW consumes baseline geospatial landscape datasets (e.g., hydrology, terrain, soils, land use, etc.), and produces modified landscape datasets that other "downstream" models and analytics consume. In this approach, OTW obtains baseline data for a site within a watershed context, and allows users to modify and set future landscapes. It is ideal for 2D/3D landscape design, offering users an ability to easily substitute futures into modeling applications that already consume similarly-formatted landscape datasets. The second development endeavor involved expanding the USEPA's integrated watershed modeling approach described by Johnston et al. (2011). Together, these projects were undertaken to more readily facilitate "Anywhere, USA" watershed assessment of landscape futures, based on web-served data and 2D/3D landscape design performed by users.

The need for PiSCES reflected the limitations of an integrated approach in which ecology models (e.g., BASS, HSI; Barber et al. 1988, Rashleigh et al. 2005), linked to hydrology models, lacked general transferability to other watershed systems. This was tied to the need for the ecological modelers to manually analyze data available for a given site to provide an initial estimate of a likely fish community. A reliable community description is essential for the modeling system to predict fish biomass in a given watershed (deriving watershed-specific habitat suitability models was another component of that approach). PiSCES design, as described here, solves that problem by allowing users an ability to quickly estimate a probable fish community for a wide variety of lotic systems throughout the US, while also accounting for habitat suitability constraints. Thus, all three components (BASS, HSI, PiSCES) are integrated to predict the effects of land use change on ecological services. Development of PiSCES and OpenTERRAworks complemented the existing watershed assessment capability (Johnston et al., 2011). The project was driven by the need for USEPA to enhance its cumulative impact assessment capabilities under the National Environmental Policy Act (NEPA, http://www.epa.gov/compliance/nepa/index.html), and other regulatory frameworks (e.g., Clean Water Act) associated with surface coal mining permits issued in the eastern United states (e.g., assessing mountain top removal and contour mines).

In developing PiSCES, we strove to combine the best available data on geographic fish distributions with knowledge of species and genera-specific habitat preferences, and package this information in a useful software tool. PiSCES can be used for research purposes in support of ecological modeling (Johnston et al. 2015) or simply to provide information to an interested citizen-scientist curious about what fish species may be present in their local streams.

**Methods**

The PiSCES core is written in C#, a Microsoft .Net programming language.  It is compiled using the .Net Framework version 4.0.  The fish properties and HUC8 distributional ranges are stored in a SQLite relational database.  The software is deployed as a web application using the Microsoft ASP.Net framework for server side processing.  The graphical user interface (GUI) is built using HTML 5, CSS and Javascript.  The PiSCES core functionality is exposed through a collection of RESTful web services that can be accessed independently of the browser-based GUI.

Scientific and common names for fish species were adopted from Page et al. (2013). Stream segmentation for the conterminous US was derived from the NHDPlus version 1 dataset (USEPA and USGS 2005). In order to develop PiSCES’ capabilities (i.e., prediction of a probable fish community for any NHDPlus stream segment, subject to user-defined physicochemical stream characteristics), we first obtained the following information for each fish species from Page and Burr (2011), the online NatureServe Explorer (http://explorer.natureserve.org/) and FishBase (http://www.fishbase.org/):

* The mapped distribution of the species (based on historic collection records)
* The stream size (mean wetted width) where the species is most commonly found
* The rarity of each species inside its range
* The maximum size (body length) that each species attains
* Assorted habitat preferences for each species

When the PiSCES GUI is first encountered, users see a navigable map of the US, where they can pan, zoom, and then select an 8-digit hydrologic basin (HUC) of interest. These basins are defined by the Watershed Boundary Dataset (WBD), obtained from the USGS (http://nhd.usgs.gov/wbd.html). This screen also allows a user to interact with the PiSCES database and display the geographic distribution (both native and introduced) for any species of interest. Filters can be set to reveal all species that fit a chosen profile, such as having some level of rarity, attaining some maximum body size, or belonging to some larger group of fishes. Once a HUC is selected on the map, the NHDPlus stream segments within that HUC are revealed. Satellite imagery and map features (e.g., roads) are shown to assist in stream identification. Once a user selects an individual stream segment, the community prediction tab is revealed, which will allow the derivation of a plausible community for that stream segment. There are a variety of informational components in PiSCES, which are discussed in the following paragraphs. Figure 1 is a schematic diagram of its primary features and functions, including how user input impact model output.

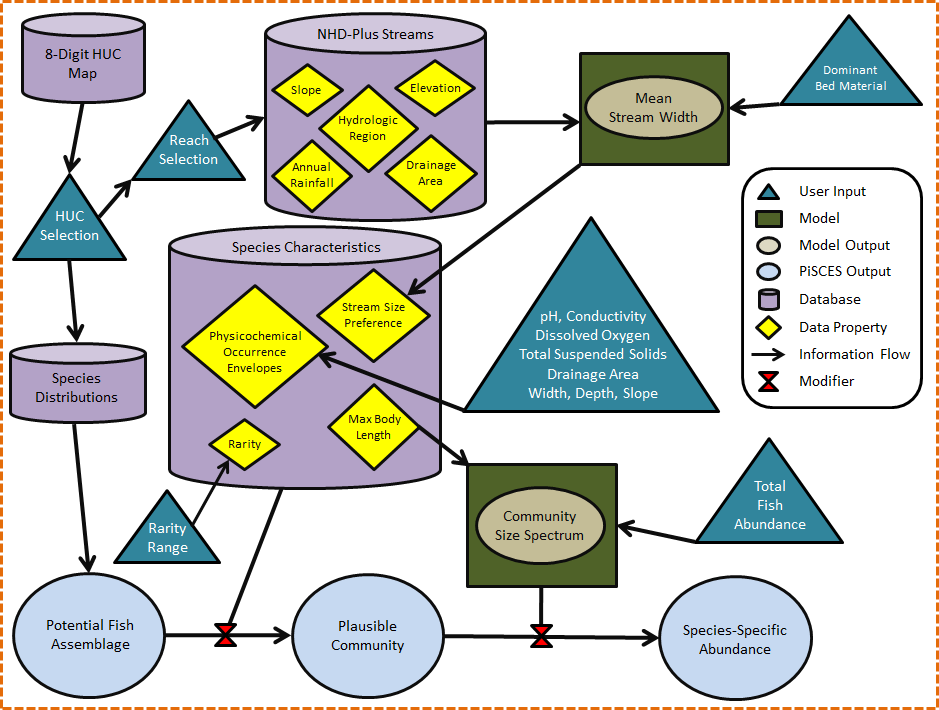


Figure . Schematic depiction of PiSCES internal processes, user inputs, databases, and information flow.

*Fish Distributions*. The basis of PiSCES fish assemblage predictions are known current geographic distributions of fish species, which were obtained primarily from two sources:

* GIS shapefiles of 8-digit HUC-based species distributions; obtained from NatureServe (2010)
* 8-digit HUC-based records of species introductions for native and non-native fishes; from the USGS Nonindigenous Aquatic Species Program.

USGS ichthyologists communicate with NatureServe personnel to keep species distributions (primarily for those species introduced to areas outside of their native ranges) up to date and quality assured. Larry Page (Florida Museum of Natural History) provided GIS shapefiles for the distributional polygons that appear in the Peterson Field Guide to Freshwater Fishes of North America (Page and Burr 2011). These polygons are imprecise at the boundaries of species distributions, so whenever possible, we chose to use the NatureServe/USGS data. The Peterson Field Guide, however, provided distributions for 16 species/sub-species not found in the NatureServe or USGS databases:

*Campostoma spadiceum* Highland Stoneroller

*Cottus hubbsi* Columbia Sculpin

*Erimyzon claviformis*  Western Creek Chubsucker

*Etheostoma atripinne*  Cumberland Snubnose Darter

*Etheostoma erythrozonum* Meramec Saddled Darter

*Etheostoma occidentale* Westrim Darter

*Etheostoma orientale*  Eastrim Darter

*Etheostoma planasaxatile* Duck Darter

*Etheostoma spilotum* Cumberland Plateau Darter

*Etheostoma tennesseense* Tennessee Darter

*Lepomis peltastes* Northern Sunfish

*Oncorhynchus mykiss newberrii*  Great Basin Rainbow Trout

*Oncorhynchus mykiss gilberti* Kern Rainbow Trout

*Oncorhynchus mykiss stonei* Sacramento Rainbow Trout

*Oncorhynchus clarkii macdonaldi* Yellowfin Cutthroat Trout

*Percina apristis*  Guadalupe Darter

For these fish, the Peterson Field Guide polygons were used. In order to convert these polygons to hydrologically-based areas (and thus more ecologically reasonable), we examined the overlap between the polygons and the map of 8-digit HUCs, taking into account HUC drainage patterns in order to determine what basins should be defined as possible locations for the species of interest. HUCs with at least 50% of their area inside the Peterson distributional polygon were considered to contain that species. We also included HUCs in a species’ distribution with overlap greater than 0%, but less than 50%, if the HUC was directly upstream or downstream of a HUC with greater than 50% overlap. When a distributional polygon was relatively small (intersecting or wholly contained within only a few HUCs), then each HUC intersecting the polygon was included in the species’ distribution.

*Stream Width*. PiSCES can filter the fish assemblage defined by species distributional information (i.e., the superset of fishes that could be found in various streams across an entire HUC) by stream size preferences for each species. Page and Burr (2011) provides this information for each species under mean flow conditions. Some species are generalists and can be found in headwaters to large rivers, and all points in between. Other species are much more selective in the size of streams they inhabit. The Peterson Guide defines lotic systems using a narrative term tied to a mean stream width metric:

* Headwater/Spring: 0-1 m
* Creek: 1-5 m
* Small River: 5-25 m
* Medium River: 25-50 m
* Large River: > 50 m

However, the NHDPlus dataset does not include information on the mean width of each stream reach. To calculate stream mean width, we used regression equations (Faustini et al. 2009) developed for nine US hydrologic regions (Figure 2). These equations predict the logarithm of bankfull width (m) based on the basin area (km2), mean annual precipitation (m), mean slope (m/m), mean elevation (m), and whether the stream bed is predominantly fine or coarse sediments. Not all of the independent variables are used in every regional equation (Table 1). These independent variables are available in the NHDPlus database for each reach, except for the dominant bed substrate. If the user cannot specify that parameter in PiSCES, then both regressions (fine and coarse bed material) are calculated, and the two results are averaged.

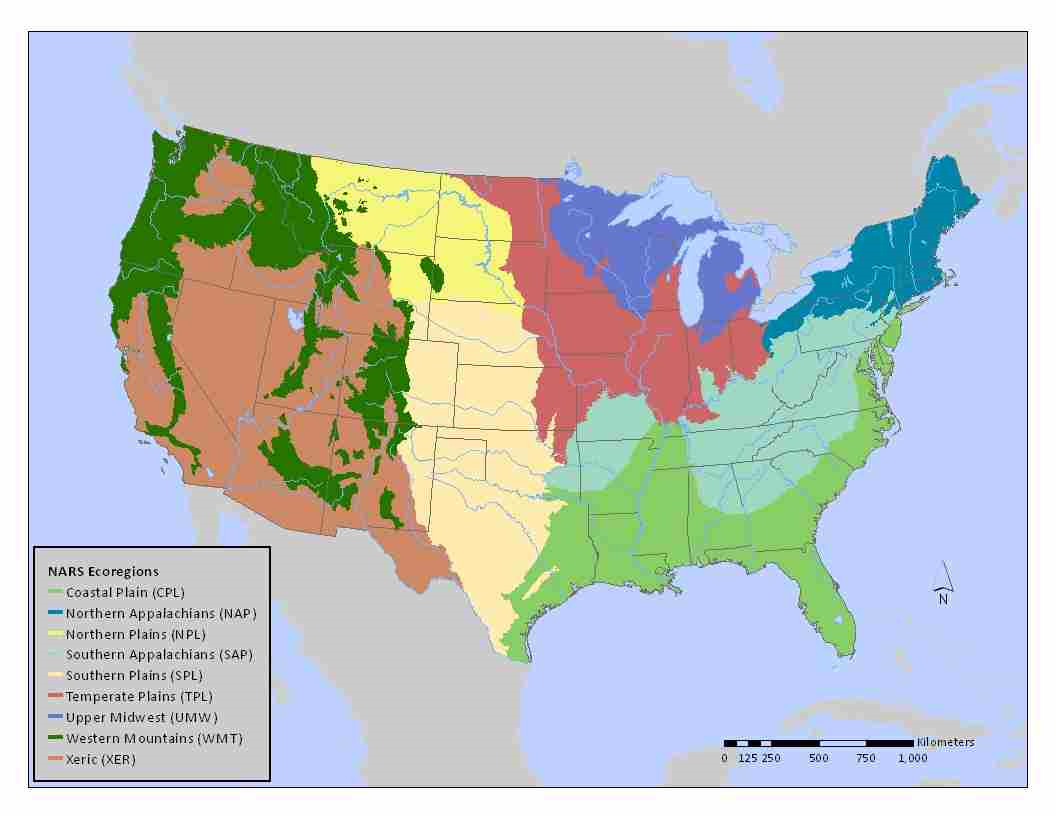


Figure 2. Nine hydrologic regions of the US, as shown in the National Rivers and Streams Assessment (USEPA 2013).

Table . Regression equations for predicting the log10 of bankfull width (m) for the nine hydrologic regions shown in Figure 2.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| EcoRegion | Sediment | Intercept | log(Area) | log(Precip) | log(Slope) | Elevation (m) |
| N. Appalachians | Fine | 0.1195 | 0.3702 | -2.6518 | -0.1895 |  |
| Course | 0.1195 | 0.3702 |  | -0.1895 |  |
| S. Appalachians | Fine | 0.3058 | 0.3935 |  |  |  |
| Course | 0.6053 | 0.3935 |  | 0.0852 |  |
| Coastal Plains | Fine | 0.4332 | 0.2604 |  |  |  |
| Course | 0.8927 | 0.2604 |  |  | -0.00557 |
| Upper Midwest | Fine | 0.2356 | 0.411 | 2.5732 | 0.0876 | 0.00137 |
| Course | 0.2356 | 0.411 |  |  |  |
| Temperate Plains | Fine | 0.3957 | 0.3498 | 1.173 |  |  |
| Course | 0.8318 | 0.1516 | 0.5172 |  |  |
| Northern Plains | Fine | 0.4041 | 0.2788 |  |  | -0.00034 |
| Course | 0.2607 | 0.2788 |  |  |  |
| Southern Plains | Fine | 0.7568 | 0.1666 |  | 0.1296 |  |
| Course | 0.2851 | 0.492 |  | 0.1296 | -0.00024 |
| Western Mountains | Fine | 0.1825 | 0.3986 | 0.6167 |  |  |
| Course | 0.2524 | 0.3986 | 0.6167 |  | 0.000033 |
| Xeric | Fine | 0.1168 | 0.3223 |  |  |  |
| Course | 0.3383 | 0.3223 | 0.3048 |  |  |

We converted the estimated bankfull width derived from Table 1 regressions to width under mean flow conditions by multiplying the bankfull width by 0.75, which is an estimate in line with stream survey data presented in Cassie (2006). When the stream size filter is invoked, PiSCES examines each potential species to ensure that the estimated mean stream width falls within the species’ stated preference range. Although there is unquantified uncertainty on multiple fronts -- the calculated mean stream widths and the Page and Burr (2011) stream size preferences -- this filter should, at its worst, keep fish like sturgeon *Acipenser transmontanus*, flathead catfish *Pylodictis olivaris*, and muskellunge *Esox masquinongy* out of predicted assemblages in headwaters and creeks, while groups like darters and madtom are kept to smaller streams.

*Species Rarity*. As noted in Page and Burr (2011), rarity is not synonymous with the spatial extent of a species’ range. A species can be very abundant within a tiny range, like a few pools or springs in the case of certain desert pupfish; or a species can be uncommon/rare, yet have a widespread distribution across a large geographic expanse. For our purposes, rarity relates to how likely a species would be found at a suitable location within its range. We converted the rarity descriptors in Page and Burr (2011) into a numeric scale:

1: Abundant 6: Uncommon

2: Abundant/Common 7: Uncommon/Rare

3: Common 8: Rare

4: Fairly Common 9: Extremely Rare

5: Common/Uncommon 10: Extinct

To use PiSCES’ rarity filter, the user sets the upper and lower bounds of the rarity criterion using the provided sliders. Doing so will filter out all species with a designated rarity above the upper bound and below the lower bound. To get a sense of what species are most likely to be seen if streams in a given HUC were sampled, the user may choose rarity ratings 1-5. However, if the user were curious about what rare species may potentially be found in the HUC, they might use rarity ratings 7-9. A rarity of 10 indicates the species was historically seen in the HUC, but is currently extinct.

*Occurrence Envelopes*. We examined density data (individuals/m2) for fishes from stream surveys of the US mid-Atlantic Highlands to derive “occurrence envelopes” for sampled fishes. These envelopes define the ranges for various characteristics of the stream where each species is most likely to be found. We used the 1993-1996 and 1997-1998 USEPA Environmental Monitoring and Assessment (EMAP) surveys (<http://www.epa.gov/emap2/html/data/surfwatr/data/index.html>) and the 2001-2002 Regional EMAP (REMAP) survey in West Virginia. Along with density estimates from 795 site visits, stream characteristics were recorded that can broadly define the types of lotic systems where each species occurs. We chose the following parameters to investigate: specific conductance (μS/cm), pH, stream width (m), stream slope (%), drainage area (ha), stream depth (cm), dissolved oxygen (% saturated), and total suspended solids (mg/l). PiSCES imposes these envelopes on raw distributional data to more accurately predict fish communities for locations lacking field data. The method is similar to the hierarchical filtering approach introduced by Smith and Powell (1971) and refined by others (Tonn 1990; Jackson et al. 2001; Stranko et. al 2005). The left panel of Figure 3 illustrates the approach for calculating species-specific environmental parameter “preferences” as detailed in Stranko et al. (2005). Our method (right panel of Figure 3) for calculating occurrence envelopes (which hypothetically should be wider than preference bounds) can be described as follows:

1) Sort every site by a given parameter value (e.g., conductivity), smallest to largest

2) For each species, raise the recorded density at each site by an exponent ≥ 0. A value of 1 gives full weight to the measured densities, and the resulting envelopes will be influenced by sites where density of the species was highest. A value close to 0 transforms density data into presence/absence data, i.e., every site where the species was captured will be weighted equally.

3) For a species of interest, use the sorted site list to calculate a cumulative proportion (Pi) of the transformed densities (Di) created in the previous step:

***Equation 1***

where *n* is the total number of sites in the dataset.

4) Identify the parameter values at the sites where Pi surpasses some threshold (e.g., the Lower Occurrence Envelope at 0.05; the Upper Occurrence Envelope at 0.95)

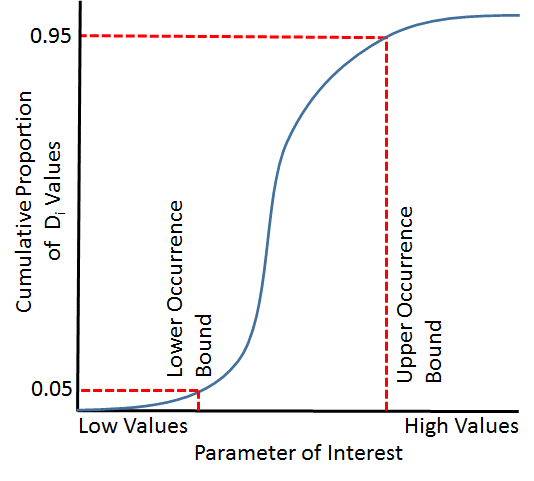
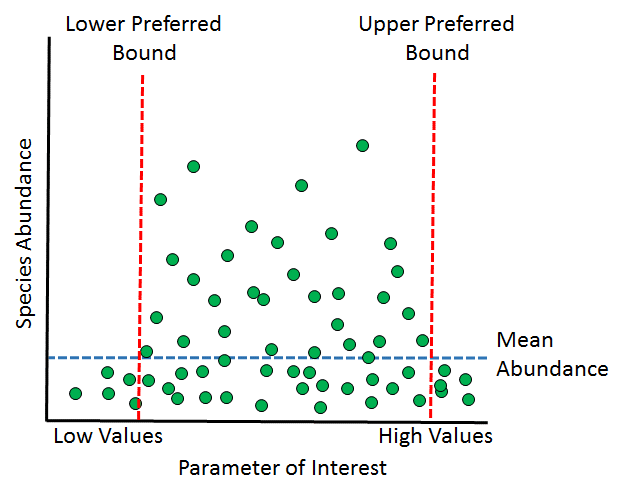


Figure 3. Two approaches to calculating environmental parameter/species preferences. The left diagram illustrates the approach of Stranko et al. (2005), using the abundance of a species across various sites; the right diagram shows our approach, using the cumulative proportion of transformed densities.

Our method is thus based on an analysis of the cumulative proportions of the Di values predicated on ranking the sites in order of their environmental parameter values, least to greatest. The preference bounds of Stranko et al. (2005) could be widened if some smaller fraction of mean abundance (e.g., 1/2 or 1/3) were used. How robust preference or occurrence envelopes are for a species would depend on the geographic extent of the analyzed dataset, and the number of sites where each species is found. In the EMAP/REMAP dataset, few species were identified in more than 50 samples, so we chose to combine species of the same genus to develop genus-specific envelopes. Even though there are intra-genus differences in habitat preferences, we hypothesize that intra-genus species would generally prefer more similar streams than inter-genus species.

For genera that occurred in more than 20 samples, Lower and Upper Occurrence bounds were set using Pi values of 0.05 and 0.95. For genera that were found in 5-19 samples, these limits were set using Pi values of 0.0001 and 0.9999 (in essence, using the minimum and maximum parameter values based on each site where the genus was found). If a genus was found in less than 5 samples, no occurrence envelope was computed. We found that atransformation exponent of 0.25 (for calculation of Di values) produced a reasonable weighting scheme across sites for most genera. However, for some genera found at more than 20 sites, using this exponent did not result in a calculable lower or upper bound (i.e., the CDF of Di values started above 0.05 or ended below 0.95 a result of very high densities at the site with either the lowest or highest parameter value across all sites where the genus was found). In these cases, we iteratively reduced the transformation exponent by 0.05 units until a lower or upper bound could be estimated. For the most part, reducing the exponent to 0.1 sufficed, but for a few genera, smaller exponents were required. Again, note that very small exponents essentially convert the density data into presence/absence information. This process basically removes a small portion of the sites at both ends of the parameter spectrum, and uses the range of parameter values across the remaining sites to estimate an occurrence envelope.

Acknowledging all the forms of uncertainty and potential error inherent in our occurrence envelope approach, we also employed a revised method that utilized the 10th and 90th percentiles of each parameter across the entire range of records in the EMAP/REMAP dataset. This re-examination of the occurrence envelopes was an attempt to use the entire dataset to identify parameters to which each genus was relatively insensitive. The lower and upper envelope bounds for every genus, as originally determined, were examined; if a lower bound was less than the 10th percentile for that parameter, the lower bound was set to the lowest value of the parameter in the dataset. In the same way, if the upper bound was greater than the 90th percentile for that parameter, the upper bound was set to the largest value in the dataset.

*Fish Abundance*. Once the initial stream fish assemblage (based only on HUC-based distributional information) has been reduced to a more probable community (a collection of species that could likely be found in a certain stream given its physicochemical characteristics), a PiSCES user may want to estimate the abundance of individuals of each species in the stream reach. For the integrated modeling system and the fish community simulation model that PiSCES was developed to support (Johnston et al. 2015), this was a necessary initial condition to estimate. PiSCES uses two pieces of information to perform this task. One is the user-specified total fish abundance in the stream reach. The default value is 2000 individuals. The second is a general relationship between the body size of an organism (maximum lengths for each species are given in Page and Burr 2011) and its abundance in a community, called the *community size spectrum* (Sheldon et al. 1972; Pope and Knights 1982; Han and Straškraba 1998; Boicourt et al. 2004):

log2(Abundance) = -α \* log2(Maximum Length, cm) ***Equation 2***

The user sets α, a tuning parameter (default = 1). Smaller α values lead to a flatter spectrum, where there are almost as many large fish as small ones. Larger α values produce a steeper decline in abundance with increasing size, as is often the case in heavily exploited fisheries (Duplisea and Castonguay 2006). We suggest keeping this parameter between 0.5 and 2.0.

Table 2 shows how this approach works for a hypothetical community of seven fish species ranging in length from 10 to 60cm, with α = 1.5. The second column is log2Abundance as determined by Equation 2. Column three is the anti-base2 logarithm (2^) of these values, and column four is each abundance divided by the sum of column three, so the values in Column 4 sum to 1. This relative abundance for each species is then multiplied by the user-specified total abundance (1,500 in this example) to provide an estimate of the absolute number of individuals of each species in the reach (column five).

Table 2. Calculations for determining a hypothetical community spectrum. Total abundance is 1,500 fish and α=1.5.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Max  Length (cm) | log2Abundance | Abundance | Relative  Abundance | N |
| 10 | -4.983 | 0.032 | 0.421 | 843 |
| 15 | -5.860 | 0.017 | 0.229 | 459 |
| 20 | -6.483 | 0.011 | 0.149 | 298 |
| 30 | -7.360 | 0.006 | 0.081 | 162 |
| 40 | -7.983 | 0.004 | 0.053 | 105 |
| 50 | -8.466 | 0.003 | 0.038 | 75 |
| 60 | -8.860 | 0.002 | 0.029 | 57 |

*Tolerance*. For a subset of species in the PiSCES database, sensitivity to watershed disturbance and stream siltation were added using information in Barbour et al. (1999). This parameter is available as an assemblage filter in the PiSCES GUI. Species are categorized as Intolerant, Moderately Tolerant, or Tolerant.

*Fish Groups*. The user can sort the PiSCES database to show groups of species that share evolutionary commonality (essentially corresponding to taxonomic families): Anchovy, Black Bass, Bowfin, Burbot, Catfish, Cavefish, Characin, Cichlid, Clupeid, Cod, Darter, Drum, Eel, Flounder, Gar, Goby, Knifefish, Lamprey, Livebearer, Loach, Madtom, Minnow, Mooneye, Mudminnow, Mullet, Needlefish, Paddlefish, Perch, Pike, Pipefish, Pirate Perch, Pupfish, Pygmy Sunfish, Salmonid, Sculpin, Silverside, Smelt, Snakehead, Splitfin, Stickleback, Stingray, Sturgeon, Sucker, Sunfish, Surfperch, Topminnow, Trout-Perch, White Bass, and Whitefish.

*Ancillary Characteristics*. PiSCES also has the following information for each species in its database:

* Origin: Native to US or Introduced
* Human Use: Sport Fish, Non-Game, Subsistence
* Typical Systems Occupied: Caves, Springs, Headwaters, Creeks, Small Rivers, Medium River, Large Rivers, Lakes/Impoundments/Ponds/Canals/Ditches, Swamp/Marsh/Bayou, Coastal/Ocean
* Preferred Lotic Habitat: Riffles, Runs/Flowing Pools, Pools/Backwaters
* Preferred Location within the System: Benthic, Surface, Nearshore/Littoral, Pelagic
* Preferred Substrate: Mud/Silt/Detritus, Sand, Gravel, Rocks/Rubble/Boulders, Vegetation, Woody Debris/Brush
* Other Preferred Water Characteristics: Clear, Turbid, Warm, Cool, Cold, Lowland (low gradient), Upland (high gradient)

These descriptors were taken from information in Page and Burr (2011), NatureServe.com, and FishBase.org. Information on subsistence species was found in Kappen et al. (2012). For most fish groups, species whose maximum body size was over 25cm were considered sport fishes unless their rarity measure was 7 or greater. For Salmonids, this threshold was 20cm, and for Sunfish and Black Bass, the threshold was 15cm. Species under these thresholds were designated as non-game.

**Results and Discussion**

The finalized PiSCES database contains information on 1015 fish species representing 200 genera. Table 3 shows categorization of the 49 groups in terms of sport fishes, non-game fishes, subsistence species, and those entirely exotic to the US. Note that some groups can be found in more than one column, and a group was classified according to the majority of the species comprising that group.

Table 3. Categorization of the 49 groups of fishes in PiSCES.

|  |  |  |  |
| --- | --- | --- | --- |
| Sportfish | Subsistence | Non-Game | Exclusively Exotic |
| Black Bass | Anchovy | Cavefish | Cichlid (Exception: Rio Grande Cichlid) |
| Bowfin | Bowfin | Characin | Knifefish |
| Burbot | Burbot | Darter | Loach |
| Catfish | Catfish | Goby | Snakehead |
| Clupeid | Clupeid | Livebearer |  |
| Cod | Cod | Madtom |  |
| Drum | Drum | Minnow |  |
| Eel | Eel | Mudminnow |  |
| Flounder | Gar | Needlefish |  |
| Gar | Lamprey | Pipefish |  |
| Mooneye | Minnow | Pirate Perch |  |
| Paddlefish | Mooneye | Pupfish |  |
| Perch | Mullet | Pygmy Sunfish |  |
| Pike | Paddlefish | Sculpin |  |
| Salmonid | Perch | Silverside |  |
| Snakehead | Pike | Splitfin |  |
| Stingray | Salmonid | Stickleback |  |
| Sturgeon | Smelt | Surfperch |  |
| Sucker | Sturgeon | Topminnow |  |
| Sunfish | Sucker |  |  |
| White Bass | Trout-Perch |  |  |
| Whitefish | White Bass |  |  |
|  | Whitefish |  |  |

*Occurrence Envelopes.* There were a total of 51 genera for which occurrence envelopes could be calculated using the mid-Atlantic EMAP and West Virginia REMAP data (Table 4). For most, we were able to calculate lower and upper bounds for the eight parameters we investigated. Missing values for cells in Table 4 arise because some parameters values were not recorded at some of the surveyed sites, so certain genus/parameter combinations dropped below a sample size of 5, stated earlier as our threshold for envelope calculation.

Table 4. Occurrence envelopes for 51 genera based on the mid-Atlantic EMAP and WV REMAP datasets for eight parameters: Con = conductivity (µS/cm), pH, W = stream width (m), Slp = stream slope (%), Ar = drainage area (ha), Dp = stream depth (cm), DO = dissolved oxygen (% saturation), TSS = total suspended solids (mg/l).



To fill in gaps in Table 4, we used envelope values from designated surrogate genera that shared ecological, morphological and/or physiological similarities to the genus with missing data (Table 5). We also used this approach to add four genera that were found in the EMAP/REMAP data, but for which no envelopes could be calculated because they were not found at more than 5 sites.

Table 5. Surrogate genera used to fill in the missing envelope estimates in Table 4, as well as an additional four genera found in the EMAP/REMAP data for which no envelopes could be calculated.

|  |  |
| --- | --- |
| Incomplete Data | Surrogate Genus |
| Amia | Dorosoma/Aphredoderus |
| Aplodinotus | Percina |
| Dorosoma | Aphredoderus |
| Enneacanthus | Lepomis |
| Erimystax | Notropis |
| Gambusia | Notropis |
| Hybognathus | Notropis |
| Ictalurus | Catostomus |
| Ictiobus | Catostomus |
| Lepisosteus | Dorosoma/Aphredoderus |
| Morone | Dorosoma/Aphredoderus |
| Perca | Percina |
| Pomoxis | Lepomis |
| Pylodictis | Ictalurus/Catostomus |
| Sander | Percina |
|  |  |
| No Data | Surrogate Genus |
| Ammocrypta | Etheostoma |
| Carpiodes | Hypentelium |
| Hybopsis | Notropis |
| Labidesthes | Notropis |

Table 6 shows the 10th and 90th percentiles of the eight investigated parameters from the 795 EMAP/REMAP survey sites. As discussed in the Methods, these were used to re-formulate the genera occurrence envelopes given in Table 4 based on a genus’ interpreted intolerance to extreme values of each parameter. For example, the 10th percentile of specific conductance across all examined sites was 45 µS/cm. In the second column of Table 4, each genus with a specific conductance lower bound less than 45 would be defined as “insensitive” to low specific conductance, and its lower envelope bound would be instead set to the lowest value of specific conductance found in the dataset (given in parentheses). When filtering a potential fish assemblage using stream characteristics, PiSCES gives the user the option of using the originally-formulated envelopes shown in Table 4, or these sensitivity-adjusted envelopes, which are less restrictive and will result in larger potential communities.

Table 6. 10th and 90th percentiles for the eight parameters in the EMAP/REMAP dataset used for revisiting the upper and lower bounds of the occurrence envelopes shown in Table 4. Minimum and maximum dataset values are shown in parentheses.

|  |  |  |
| --- | --- | --- |
| Parameter | 10% (min value) | 90% (max value) |
| Specific Conductance (µS/cm) | 45 (10) | 472 (4000) |
| pH | 6.82 (4.64) | 8.23 (9.3) |
| Stream Width (m) | 2.2 (0.03) | 28 (456) |
| Stream Slope (%) | 0.33 (0.01) | 2.36 (17.3) |
| Drainage Area (ha) | 220 (11) | 64576 (10.3 million) |
| Water Depth (cm) | 13 (0.333) | 59 (106) |
| Dissolved Oxygen (%) | 53 (7) | 88 (139) |
| Total Suspended Solids (mg/l) | 1.0 (0) | 19 (180) |

There is no doubt variability in tolerance and habitat preferences between species of the same genus (Barbour et al. 1999, Page and Burr 2011), but this variability is generally not as great as between-genera variability, thus these derived envelopes can be informative. Being genus-specific, the envelopes are wider than what would be expected for individual species, which will lead PiSCES to predict the possibility of more species than would commonly be found in a routine fish survey.

*Community Abundance.* The species-specific abundance estimates are presented in PiSCES with greater precision than the estimation process can support, and without confidence limits; thus they should be used primarily to get a logarithmic sense of species’ abundance, i.e., would the species occur in the thousands, hundreds, or tens of individuals.

*PiSCES Implementation.* In summarizing PiSCES development outcomes and the introductory discussion on motivation, design, and intended use, three major benefits are derived from its final design and functionality. In stand-alone mode, PISCES allows users to develop reasonably reliable estimates of fish communities in lotic systems across the US. This functionality has numerous applications to serve a multitude of current assessment programs and research endeavors. Secondly, within an integrated environmental modeling framework (Johnston et al. 2011), PiSCES provides a service necessary to perform hydroecological assessments which link mechanistic hydrology models with ecological models to achieve prediction goals. Finally, PiSCES' general flexibility allows users to modify a community “best estimate” based on additional, waterbody-specific data. This functionality, established as an important design requirement, enhances the capabilities of both standalone use and integrated modeling applications for which it was created.

*Availability.* PiSCES can be found on the web at: qed.epa.gov/pisces/

**Acknowledgements:**

We are indebted to the following people: Larry Page, Florida Museum of Natural History, for fish distribution maps and advice on interpreting the information contained in Page and Burr (2011); Matthew Cannister, Pamela Fuller, and Matthew Nielson, USGS Nonindigenous Aquatic Species Program, for distributional information on native and introduced species; Daniel Cassie, Canadian Department of Fisheries and Oceans, for providing additional data from his research; Jason McNees and Lynn Kutner, NatureServe, for assistance in obtaining fish species distributional information; Scott Stranko, Maryland Department of Natural Resources, and Bob Hilderbrand, University of Maryland Center for Environmental Science, for advice on analytical methods to calculate species habitat preferences.

*Disclaimer: This document has been reviewed and approved in accordance with U.S. Environmental Protection Agency policy. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.*

**References**

Angermeier, P. and Winston, M. 1999. Characterizing fish community diversity across Virginia landscapes: prerequisite for conservation. Ecological Applications, 9(1): 335-349.

Barber, M., Suarez, L. and Lassiter, R. 1998. Modeling bioconcentration of nonpolar organic pollutants by fish. Environmental Toxicology and Chemistry, 7(7): 545-558.

Barbour, M., Gerritsen, J., Snyder, B. and Stribling, J. 1999. Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish, Second Edition. EPA 841-B-99-002. U.S. Environmental Protection Agency; Office of Water; Washington, D.C.

Baxter, C. and Hauer, F. 2000. Geomorphology, hyporheic exchange, and selection of spawning habitat by bull trout (*Salvelinus confluentus*). Canadian Journal of Fisheries and Aquatic Sciences, 57(7): 1470–1481.

Bertolo, A. and Magnan, P. 2006. Spatial and environmental correlates of fish community structure in Canadian Shield lakes. Canadian Journal of Fisheries and Aquatic Sciences, 63(12): 2780-2792.

Boicourt, W., Gallegos, C., Harding Jr., L., Houde, E., Mallonee, M., McClain, C., and Roman, M. 2004. Trophic indicators of ecosystem health in Chesapeake Bay. 2004 Progress Report. USEPA Grant R828677C002.

<http://cfpub.epa.gov/ncer_abstracts/index.cfm/fuseaction/display.abstractDetail/abstract/6126/report/2004>

Caissie, D. 2006. River discharge and channel width relationships for New Brunswick Rivers. Canadian Technical Report of Fisheries and Aquatic Sciences 2637. Canadian Department of Fisheries and Oceans, Moncton, NB.

<http://www.dfo-mpo.gc.ca/Library/320466.pdf>

Dauwalter, D., Splinter, D., Fisher, W. and Marston, R. 2008. Biogeography, ecoregions, and geomorphology affect fish species composition in streams of eastern Oklahoma, USA. Environmental Biology of Fishes, 82(3), 237-249.

Davis, W., Snyder, B., Stribling, J. and Stoughton, C. 1996. Summary of State biological assessment programs for streams and rivers. EPA 230-R-96-007. U. S. Environmental Protection Agency, Office of Planning, Policy, and Evaluation, Washington, DC.

Duplisea, D. and Castonguay, M. 2006. Comparison and utility of different size-based metrics of fish communities for detecting fishery impacts. Canadian Journal of Fisheries and Aquatic Sciences, 63: 810-820.

Eros, T., Saly, P., Takacs, P., Higgins, C., Biro, P. and Schmera, D. 2014. Quantifying temporal variability in the metacommunity structure of stream fishes: the influence of non-native species and environmental drivers. Hydrobiologia, 722(1): 31-43.

Fausch, K., Lyons, J., Karr, J. and Angermeier, P. 1990. Fish communities as indicators of environmental degradation. Pages 123-144 in: *Biological indicators of stress in fish, Symposium 8* (Editor: Adams, S.). American Fisheries Society, Bethesda, Maryland.

Faustini, J., Kaufmann, P. and Herlihy, A. 2009. Downstream variation in bankfull width of wadeable streams across the conterminous United States. Geomorphology 108: 292-311.

Fransen, B., Duke, S., McWethy, G., Walter, J. and Bilby, R. 2006. A logistic regression model for predicting the upstream extent of fish occurrence based on geographical information systems data. North American Journal of Fisheries Management, 26: 960–975.

Han, B. and Straškraba, M. 1998. Size dependence of biomass spectra and size intervals: the effects of size scales and size intervals. Journal of Theoretical Biology, 191: 259–265.

Herlihy, A., Larsen, D., Paulsen, S., Urquhart, N. and Rosenbaum, B. 2000. Designing a spatially balanced, randomized site selection process for regional stream surveys: the EMAP Mid-Atlantic pilot study. Environmental Monitoring and Assessment, 63(1): 95-113.

Hughes, R. and Oberdorff, T. 1999. Applications of IBI concepts and metrics to waters outside the United States and Canada. Pages 79-83 in: *Assessing the Sustainability and Biological Integrity of Water Resources Using Fish Communities* (Editor: Simon, T.). Lewis Press, Boca Raton, Florida.

Jackson, D., Peres-Neto, P. and Olden, J. 2001. What controls who is where in freshwater fish communities the roles of biotic, abiotic, and spatial factors. Canadian Journal of Fisheries and Aquatic Sciences, 58(1): 157-170.

Johnston, J., McGarvey, D., Barber, M., Laniak, G., Babendreier, J., Parmar, R., Wolfe, K., Kraemer, S., Cyterski, M., Knightes, C., Rashleigh, B., Suarez, L. and Ambrose, R. 2011. An integrated modeling framework for performing environmental assessments: Application to ecosystem services in the Albemarle-Pamlico basins (NC and VA, USA). Ecological Modelling 222(14): 2471-2484.

Kappen, A., Allison, T. and Verhaaren, B. 2012. Treaty rights and subsistence fishing in the US waters of the Great Lakes, Upper Mississippi River, and Ohio River basins. USACE, Chicago, IL.

<http://glmris.anl.gov/documents/docs/Subsistence_Fishing_Report.pdf>

Karr, J. 1981. Assessment of biotic integrity using fish communities. Fisheries, 6(6): 21-27.

Karr, J. and Dudley, D. 1981. Ecological perspective on water quality goals. Environmental management, 5(1): 55-68.

Karr, J., Fausch, K., Angermeier, P., Yant, P. and Schlosser, I. 1986. Assessing biological integrity in running waters: a method and its rationale. Special Publication 5. Illinois Natural History Survey, Champaign, IL.

Karr, J. and Chu, E. 1999. *Restoring life in running waters: better biological monitoring*. Island Press, Washington, D.C.

Kelso, J. and Johnson, M. 1991. Factors related to the biomass and production of fish communities in small, oligotrophic lakes vulnerable to acidification. Canadian Journal of Fisheries and Aquatic Sciences, 48(12): 2523-2532.

Kendrick, T. and Francis, M. 2002. Fish assemblages in the Hauraki Gulf, New Zealand. New Zealand Journal of Marine and Freshwater Research, 36(4): 699-717.

Kennard, M., Olden, J., Arthington, A., Pusey, B. and Poff, N. 2007. Multiscale effects of flow regime and habitat and their interaction on fish assemblage structure in eastern Australia. Canadian Journal of Fisheries and Aquatic Sciences, 64(10): 1346-1359.

Kovacs, T., Martel, P. and Voss, R. 2002. Assessing the biological status of fish in a river receiving pulp and paper mill effluents. Environmental pollution, 118(1): 123-140.

Madejczyk, J., Mundahl, N. and Lehtinen, R. 1998. Fish assemblages of natural and artificial habitats within the channel border of the upper Mississippi River. The American Midland Naturalist, 139(2): 296-310.

McCleary, R. and Hassan, M. 2011. Predictive modeling and spatial mapping of fish distributions in small streams of the Canadian Rocky Mountain foothills. Canadian Journal of Fisheries and Aquatic Sciences, 65: 319–333.

McCormick, F., Peck, D. and Larsen, D. 2000. Comparison of geographic classification schemes for Mid-Atlantic stream fish assemblages. Journal of the North American Benthological Society, 19(3): 385-404.

Miller, D., Hughes, R., Karr, J., Leonard, P., Moyle, P., Schrader, L. and Orth, D. 1988. Regional applications of an index of biotic integrity for use in water resource management. Fisheries, 13(5): 12-20.

NatureServe. 2010. Digital Distribution Maps of the Freshwater Fishes in the Conterminous United States. Version 3.0. Arlington, VA. U.S.A.

Oakes, R., Gido, K., Falke, J., Olden, J. and Brock, B. 2005. Modelling of stream fishes in the Great Plains, USA. Ecology of Freshwater Fish, 14: 361–374.

Ornellas, A. and Coutinho, R. 1998. Spatial and temporal patterns of distribution and abundance of a tropical fish assemblage in a seasonal Sargassum bed, Cabo Frio Island, Brazil. Journal of Fish Biology, 53(sA): 198-208.

Page, L. and Burr, B. 2011. *Peterson Field Guide to Freshwater Fishes of North America North of Mexico*. 2nd Edition. Houghton Mifflin Harcourt, Boston.

Page, L., Espinosa-Pérez, H., Findley, L., Gilbert, C., Lea, R., Mandrak, N., Mayden, R. and Nelson, J. 2013. *Common and Scientific Names of Fishes from the United States, Canada, and Mexico.* 7th Edition. American Fisheries Society, Special Publication 34, Bethesda, MD.

Plafkin, J., Barbour, M., Porter, K., Gross, S., and Hughes, R. 1989. Rapid Bioassessment Protocols for use in Streams and Rivers: Benthic Macroinvertebrates and Fish. U.S. Environmental Protection Agency. EPA 440/4-89/001. USEPA, Washington, DC.

Pool, T., Olden, J., Whittier, J. and Paukert, C. 2010. Environmental drivers of fish functional diversity and composition in the Lower Colorado River Basin. Canadian Journal of Fisheries and Aquatic Sciences, 67(11): 1791-1807.

Pope, J. and Knights, B. 1982. Comparisons of length distributions of combined catches of all demersal fishes in surveys in the North Sea and Faroe Bank. Pages 116-118 in: *Multispecies Approaches to Fisheries Management* (Editor: Mercer, M.). Canadian Special Publication of Fisheries and Aquatic Sciences, 59.

Quist, M., Hubert, W. and Rahel, F. 2004. Elevation and stream-size thresholds affect distributions of native and exotic warmwater fishes in Wyoming. Journal of Freshwater Ecology, 19(2): 227-236.

Rahel, F. 2002. Homogenization of freshwater faunas. Annual Review of Ecology and Systematics, 33: 291-315.

Rashleigh, B., Parmar, R., Johnston, J. and Barber, M. 2005. Predictive habitat models for the occurrence of stream fishes in the Mid-Atlantic Highlands. North American Journal of Fisheries Management, 25(4): 1353-1366.

Saiki, M. and Martin, B. 2001. Survey of fish and environmental conditions in Abbotts Lagoon, Point Reyes National Seashore, California. California Fish and Game, 87(4): 123–138.

Schaefer, J., Clark, S. and Warren Jr, M. 2012. Diversity and stability in Mississippi stream fish assemblages. Freshwater Science, 31(3): 882-894.

Schweizer, P. and Jager, H. 2011. Modeling regional variation in riverine fish biodiversity in the Arkansas–White–Red River basin. Transactions of the American Fisheries Society, 140(5): 1227-1239.

Sheldon, R., Prakash, A. and Sutcliffe, W. 1972. The size distribution of particles in the ocean. Limnology and Oceanography, 18: 719–733.

Shuter, B. and Post, J. 1990. Climate, population viability, and the zoogeography of temperate fishes. Transactions of the American Fisheries Society, 119(2): 314-336.

Simon, T. and Lyons, J. 1995. Application of the index of biotic integrity to evaluate water resource integrity in freshwater ecosystems. Pages 245-262 in: *Biological assessment and criteria—Tools for water resource planning and decision making* (Editors: Davis, W. and Simon, T.)*.* Lewis Publishers, Boca Raton, Florida.

Simon, T. 1998. *Assessing the Sustainability and Biological Integrity of Water Resources using Fish Communities*. CRC Press, Boca Raton, Florida.

Smith, C. and Powell, C. 1971. The summer fish communities of Brier Creek, Marshall County, Oklahoma. American Museum Novitates, **2458**, 1–30.

Stepenuck, K., Crunkilton, R. and Wang, L. 2002. Impacts of Urban Landuse on Macroinvertebrate Communities in Southeastern Wisconsin Streams. Journal of the American Water Resources Association, 38(4):1041-1052.

Stoddard, J., Larsen, D., Hawkins, C., Johnson, R. and Norris, R. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. Ecological Applications, 16(4): 1267-1276.

Stranko, S., Hurd, M. and Klauda, R. 2005. Applying a large, statewide database to the assessment, stressor diagnosis, and restoration of stream fish communities. Environmental Monitoring and Assessment, 108: 99–121.

Taylor, C. and Warren Jr, M. 2001. Dynamics in species composition of stream fish assemblages: environmental variability and nested subsets. Ecology, 82(8): 2320-2330.

Taylor, C., Holder, T., Fiorillo, R., Williams, L., Thomas, R. and Warren Jr, M. 2006. Distribution, abundance, and diversity of stream fishes under variable environmental conditions. Canadian Journal of Fisheries and Aquatic Sciences, 63(1): 43-54.

Taylor, C., Millican, D., Roberts, M. and Slack, W. 2008. Long‐term change to fish assemblages and the flow regime in a southeastern US river system after extensive aquatic ecosystem fragmentation. Ecography, 31(6): 787-797.

Tonn, W. 1990. Climate change and fish communities: A conceptual framework. Transactions of the American Fisheries Society, **119**: 337–352.

USEPA Office of Research and Development. 2006. Watershed Health Assessment Tools Investigating Fisheries - WHAT IF version 2.0: A Manager's Guide to New Features. EPA/600/R-06/109. USEPA, Washington, DC.

USEPA Office of Wetlands, Oceans, and Watersheds. 2013. National Rivers and Streams Assessment 2008-2009. EPA/841/D-13/001. USEPA, Washington, DC.

USEPA and USGS. 2005. National Hydrography Dataset Plus – NHDPlus, Version 1.0.

<http://www.horizon-systems.com/NHDPlus/NHDPlusV1_home.php>.

Wang, L., Lyons, J., Kanehl, P., Bannerman, R. and Emmons, E. 2000. Watershed urbanization and changes in fish communities in southeastern Wisconsin streams. Journal of the American Water Resources Association 36: 1173–1175.

Weaver, L. and Garman, G. 1994. Urbanization of a watershed and historical changes in a stream fish assemblage. Transactions of the American Fisheries Society, 123(2): 162-172.

Wesche, T. and Isaak, D. 1999. Watershed management and land use practices. Pages 217-248 in: *Inland fisheries management in North America, 2nd* edition (Editors: Kohler, C. and Hubert, W). American Fisheries Society, Bethesda, Maryland.

Yoder, C. and Smith, M. 1998. Using fish assemblages in a state biological assessment and criteria program: essential concepts and considerations. Pages 17-56 in: *Assessing the Sustainability and Biological Integrity of Water Resources Using Fish* Communities (Editor: Simon, T.). Lewis Press, Boca Raton, FL.