APPALACHIAN LANDSCAPE CONSERVATION COOPERATIVE GRANT 2013 MILESTONE REPORT

Grant Number: 2012-03

Grant Title:
Development of a hydrologic foundation and flow-ecology relationships for monitoring riverine resources in the Marcellus Shale region

Phase 1 Project Report

Submitted by:

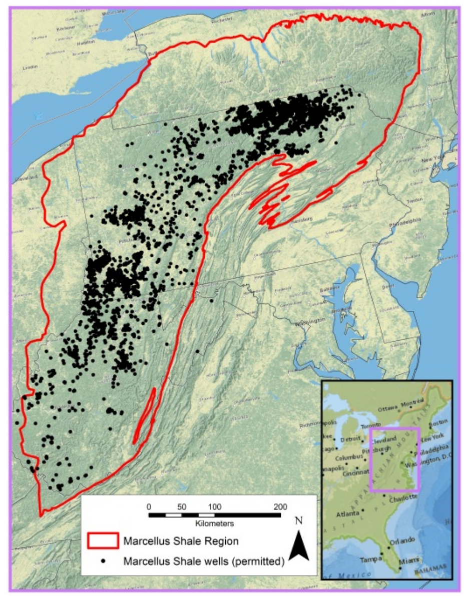
William L. Fisher
U. S. Geological Survey
New York Cooperative Fish and Wildlife Research Unit
Cornell University, Ithaca, NY 14853
Contact information: wlf9@cornell.edu, 607-255-2839

Jason Taylor and Maya Weltman-Fahs
New York Cooperative Fish and Wildlife Research Unit
Cornell University, Ithaca, NY 14853

30 June 2013 (revised 25 July 2013, final 19 August 2013)

Project Narrative: This project provided information on models that predict ecological responses to flow alteration within the Marcellus Shale region of the Appalachian Landscape Conservation Cooperative. The project involves using the Ecological Limits of Hydrologic Alteration (ELOHA) approach to develop a hydrologic foundation, develop flow-ecology relationships, and predict future impacts associated with increased water withdrawals within the Marcellus Shale region. The 1st phase of the project will involve reviewing existing tools and gathering available data within the project area. The 2nd phase of the project will require applying appropriate hydrologic modeling tools to build a hydrologic foundation and estimate flow alteration, followed by relating existing biological data to flow alteration metrics to develop flow-ecology relationships. The hydrologic foundation and flow-ecology relationships will serve as a useful tool for predicting future biological changes associated with increased water withdrawals in the Marcellus Shale region.

Important Background Information: Horizontal hydraulic fracturing has led to rapid expansion of natural gas drilling in the Marcellus Shale deposit in portions of West Virginia and Pennsylvania (see accompanying figure), and is expected to continue and expand into Ohio and New York. Two to seven million gallons of water are needed per hydraulic fracturing ‘stimulation’ event, a single natural gas well can be fractured several times over its lifespan, and a well pad site can host multiple wells. This large volume of water needed per well, multiplied by the distributed nature of development across the region, suggests that hydraulic fracturing techniques for natural gas development will put substantial strain on regional water supplies (Rahm and Riha 2012). Surface water is the primary source for hydraulic fracturing related water withdrawals in the Susquehanna River basin within the Marcellus Shale region, but groundwater, which has been a major water source in other natural gas deposits, is also a potential water source. Water consumption related to natural gas drilling, whether surface or subsurface, combined with existing concerns over climate change and future non-drilling water resource needs, have sparked concern among hydrologists and aquatic biologists about the sourcing of water within the region. Changes in stream flow may alter available habitat for freshwater biodiversity and other ecological processes in adjacent freshwater ecosystems. This concern highlights the need for the development of region-wide environmental flow policies, including the Marcellus Shale region, that are protective of stream ecosystems well into the future.



Environmental flows can be defined as the flow of water in a natural river or lake that sustains healthy ecosystems and the goods and services that humans derive from them (Poff et al. 1997). A number of measures have proven useful for quantitatively describing the flow of water in a water body: *magnitude* or the amount of water flowing, in cubic feet per second, or some other unit of measure; *duration* of a hydrologic condition, such as high or low flow events; *timing* of flows; *frequency* of occurrence; and the *rate of change* between one type of flow and another. Each of these measures can be characterized by a range of natural variability, with particular emphasis on inter-annual variability. The process of defining environmental flows seeks to preserve enough of the natural variability in these hydrologic measures to protect the ecological functions essential to diverse, healthy communities of aquatic organisms. For example, natural floods are necessary to scour river channels, maintain floodplains, and provide access to floodplains for organisms that depend on them; on the other hand, aquatic biota may not be adversely impacted with some reduction in the natural frequency and duration of flooding. Prescriptions for environmental flows, which seek to balance ecological and economic needs, have been developed for a number of river systems around the globe, including partnerships between the Army Corps of Engineers and The Nature Conservancy for the Savannah River and other rivers.

While river-specific approaches have contributed substantially to the field of flow restoration, the global pace of human modification of river flow regimes, and the growing threat to freshwater biodiversity, demand a framework that can develop flow recommendations for the rivers of an entire region. The Ecological Limits of Hydrologic Alteration (ELOHA) framework seeks to fill this need, beginning with: 1) *developing a hydrologic foundation* based on modeled baseline and developed hydrographs; 2) *classifying stream types* using baseline hydrology and geomorphic characteristics to facilitate generalizations that can apply to all the streams within a class; 3) *analysis of flow alteration*; and 4) *development of flow-ecology linkages,* which provide testable relationships "that can serve as a starting point for empirically based flow management at a regional scale" (Poff et al. 2010). This framework incorporates best professional judgment with quantitative analysis of existing data, and has been applied at the watershed level for the Susquehanna, Connecticut, and Potomac rivers, at the statewide level in Massachusetts, Michigan, Maine, and Florida, and efforts are currently underway in the Great Lakes portion of New York and the upper Ohio River region of Pennsylvania.

Goal/Purpose Statement: Flow-ecology hypotheses developed through previous (Ecosystem Flow Recommendations for the Susquehanna River Basin) and current (Upper Ohio/Great Lakes Tributaries) ELOHA projects within or adjacent to the Marcellus Shale region may serve as a framework for developing empirical relationships between hydrologic alteration and ecological responses and making predictions about future scenarios. This will require adequate flow models that can be used to explore flow-ecology relationships to enhance long-term management of aquatic resources across the Marcellus Shale region. Many models have been developed at spatial or temporal scales that do not match existing invertebrate and fish data, model only high or low flows, or were developed by groups who wish to keep them proprietary. Therefore, Phase I of this project will involve an inventory of flow models and the underlying, or potential, data sources from instream monitoring networks to: 1) Determine what ecological flow models that predict both low and high flows and are in use or are applicable to the Marcellus Shale region; and 2) Recommend suitable model(s) for instream flow predictions both dependent and independent of ecological/biological data. In Phase II of this project we will: 3) Apply a predictive model(s) that assesses how existing permitted and non-permitted water uses and future water use will alter critical hydrologic and hydraulic forces that maintain aquatic habitats; and finally, 4) forecast how biological communities or target species will respond to predicted changes in hydrology.

Specific Deliverables:

Phase I

Two deliverables were identified for phase I of this project:

1) A report that assesses the availability of hydrologic and ecological flow models suitable for the Marcellus Shale region that predict discharge thresholds and frequency of both high and low flow events and the vulnerabilities these extremes will create for conservation targets, then recommends one or more models for use in the Marcellus Shale region.

2) A georeferenced summary assessment of the adequacy of available ecological data to inform ecological flow model(s) for streams within the Marcellus Shale, including a summary assessment of critical information gaps.

We produced these deliverables by accomplishing four objectives: 1) a literature review of hydrologic models currently used within the Marcellus Shale region, 2) development of a geo-referenced stream gage database, 3) contact and coordination with users and developers of stream flow modeling tools, and 4) development of a geo-referenced stream biological database for the Marcellus Shale region. Below are the results of our efforts by objective.

Review of Hydrologic Models

Forty-nine hydrologic flow models were reviewed. Information compiled for each models included: model name, description, landscape generalization, processing style, developing agency, technical contacts, website availability, temporal scale, geographic scale, inputs and outputs. To evaluate the utility of these models for the Marcellus Shale region, we used the following criteria to evaluate each model and tabulated the sum of the individual category rankings to produce an overall ranking for each model. The higher the model rank, the greater utility for this project. When a particular criterion for the model was not available from the literature or documentation for that model, the model received a zero for that category. The zero score reflected the unavailability of information with respect to how a model functions and not the quality of the model. Without that information, the relevance and usefulness of that model for this project would be significantly limited. We thank Brian Buchanan, Cornell University, for developing these model evaluation criteria.

Model Criteria

Temporal Scale (Monthly=5, Weekly=3, Daily=1)
Explanation: Longer time steps (monthly) are preferable to shorter time steps (daily) when modeling hydrologic patterns over large areas (e.g., regions, HUC4-HUC8 drainage basins). Conversely, modeling hydrologic patterns at shorter temporal steps (weekly or daily) are preferable if modeling is conducted over small areas (e.g., subwatersheds, HUC12 drainage basins).

Continuous/Event-based (Continuous=5, Event Based=1).
Explanation: Continuous simulation is preferable because modeling will focus on base/low flows and not flood/peak flows.

Spatial Scale (Macro=5 [>10,000 km^2], Meso=3 [<10,000 km^2, 1 km^2], Micro=1 [<1 km^2])
Explanation: Because we are dealing with a large region (Marcellus Shale), the scale at which the model is applicable must be regional to reliably produce estimates for of different rivers/stream types.

Spatial Resolution (Lumped=5, Semi-lumped/Combined lumped/distributed=3, Distributed=1)
Explanation: A fully distributed model would be too computationally complicated at the scale of the Marcellus Shale region.

Parsimony (Low=1 [≥10 parameters], Medium=3 [10>≤5], High=5 [<5])
Explanation: Because there are a number of different states and regional agencies collecting data that will need to be used to represent the entire Marcellus Shale region, and consistency is vital across the data for creation of relevant input datasets; therefore, the model chosen must have few input requirements (must be parsimonious). Further, the model must be computationally efficient to run with the large datasets.

Inter-agency Collaboration (Low=1 [not currently used by other relevant agencies], High=5 [widely adopted and actively used by other relevant agencies and collaborators])
Explanation: Interagency collaboration is vital to cooperative research, given the size of the region.

Code and Model Support Availability (Publically Available/Code is actively maintained/User manuals exist/Website maintained=5, Commercially Available/Some website and code maintenance= 3, Hard to Obtain Research Model/Code is not maintained/Little user support exists =1)
Explanation: Model must be updated and supported to be useful, otherwise it will be very difficult to initiate the modeling process and any problems encountered while modeling could be slow down or halt the project.

Results

Data for each model and the ranks are in the HydroModel-ranked spreadsheet file included with this report. Based on our evaluation of the models, the ABCD monthly water balance model had the highest rank with a score of 33 out of a possible 35. The remaining top ten models that ranked high were: AVGWLF (ArcView Generalized Water Loading Functional model, now known as MapShed) and MIKE 11RR (Rainfall Runoff model) both with a score of 27; GEFC (Global Environmental Flow Calculator), OASIS (Options Analysis in Irrigation Systems) and SWAT (Soil and Water Assessment Tool) all with a score of 25; the TPWBM (Two-Parameter Water Balance Model) with a score of 24; and HEC-HMS (Hydrologic Engineering Center-Hydrologic Modeling System), Thornthwaite Monthly Water Balance Model, and WaterFall (Watershed Flow and Allocation system), which all scored 23. Twenty-eight models ranged in rank between 22 and 11. The KINEROS2 (Kinematic Runoff and Erosion Model v.2) had the lowest score of 5, largely because of missing information.

The ABCD model is a monthly water balance model with a continuous processing style that generalizes the landscape. The model runs on a monthly time-step, can be used at a continental to watershed geographic scale, requires climate data (average annual precipitation), potential evapotranspiration (average monthly temperature, solar radiation), streamflow (average monthly) and outputs monthly streamflow. We recommend using this model for modeling flows in rivers and streams in the Marcellus Shale region. This model is applicable across the whole study area and the North Atlantic LCC and the Northeast Climate Science Center projects are currently using it, so it should integrate well across LCC boundaries with other projects.

Our model evaluation included a streamflow estimator tool, StreamStats. Although this model type did not rank in the top ten (score equaled 19), primarily be of inadequate information for some criteria, it would be useful for working at small spatial scales such as subwatersheds. We recommend using this model and other similare streamflow estimator tools that have been or are being developed by the U.S. Geological Survey. The USGS PA Water Science Center has developed the Baseline Streamflow Estimator (BaSE) tool (http://pa.water.usgs.gov/projects/surfacewater/flow_estimation/), and the Sustainable Yield Estimator (SYE) tool is currently being developed for New York (Chris Gazoorian, USGS, NY Water Science Center, personal communication). A similar tool was developed for Massachusetts (U.S Geological Society Scientific Investigations Report 2011-5193)

Georeferenced Stream Gage Database

A database of 187 stream gages in the Marcellus Shale region was provided by Dr. Ryan MacManamay of the Oak Ridge National Laboratory. Dr. MacManamay has completed hydrologic modeling and stream classification research for the Southeast LCC. Information provided for each model includes: location, station ID, station name, drainage area, type, river distance, Water Resources Report notes, screening notes, and stream flow for three time periods: 1900-2009, 1950-2009, 1990-2009. The spreadsheet with the gage information is included in this report.

Coordination with Regional Streamflow Modelers and Users

During phase 1 we identified and/or made contact with several people in the region who are working with streamflow models and/or ecological (mostly fish) databases. Below is a table of those people and their organizations.

Name	**Organization**	**Contacted? Y/N**	**Comments**
Cara Campbell	Research Fish Biologist, U.S. Geological Society, Leetown Science Center, Northern Appalachian Research Laboratory	Y	Met at USGS NARL and discussed fish mapping project.

John Arway	Executive Director, Pennsylvania Fish and Boat Commission	Y	Requested information on PAFBC fish data. Referred us to Leroy Young.
Leroy Young	Pennsylvania Fish and Boat Commission	Y	Refered to Rod Klime
Rod Kime	Pennsylvania Department of Environmental Preservation	Y	Inquired about PAFBC data via emails and phone message but got no reply-PAFBC data available through MARIS
Nevin Welte	Western Pennsylvania Conservancy	N	Did not contact
Tyler Wagner	Assistant Unit Leader-Fisheries of the Pennsylvania Cooperative Fish and Wildlife Research Unit	Y	Inquired about PA fish databases.
Dan Cincotta	West Virginia Department of Natural Resources	N	Did not contact. WV data available through MARIS.
Scott Morrison	West Virginia Department of Natural Resources	Y	Inquired about WV fish databases. Referred us to Dan Cincotta.
Brian Carr	West Virginia Department of Environmental Preservation	N	Did not contact. WV data available through MARIS.
John Wirts	West Virginia Department of Environmental Preservation	N	Did not contact. WV data available through MARIS.
Terry Messing	United States Geological Survey – West Virginia	N	Did not contact. WV data available through MARIS.
Andy Loftus	Andrew Loftus Consulting and MARIS fish database developer	Y	Had several phone conversations about merit of using MARIS-also provided advice on inporting new data into MARIS
Stuart Welsh	Assistant Unit Leader-Fisheries, of the West	Y	Inquired about WV fish databases.

	Virginia Cooperative Fish and Wildlife Research Unit		Referred us to Dan Cincotta.
Ruth Thornton	The Nature Conservancy – West Virginia	N	Did not contact
Lou Reynolds	Environmental Protection Agency	N	Did not contact
Sam Dinkins	ORSANCO	Y	Talked briefly about project at Upper Ohio River environmental flow workshop (TNC)
Arlene Olivero and Mark Anderson	The Nature Conservancy	Y	Multiple phone calls and conference call about stream classification- TNC agreed to incorporate extra area in data layer development for their LCC stream classification project. This stream classification will be available to flow project for stratifying future ecological response models across stream types.
Ryan MacManamay	Oak Ridge National Laboratory	Y	Contacted and had several discussions with Ryan- Ryan is willing to provide his least-altered gage database for modelling purposes. Ryan is also working on a flow classification for the other APP LCC stream classfication project. Taylor and MacManamay are working on a finer scale Marcellus classification as a side project.
Chris Gazoorian	USGS New York Water Science Center	Y	Met with Chris to discuss the NY streamflow estimator tool.
Stacey Archfield	USGS MA-RI Water Science Center	N	Did not contact, but work was referenced at AtlLCC and NECSC meeting.
Ben Letcher	USGS Silvio Conte Fish Center and Univ. of Massachusetts	Y	Met at NA and APP LCC aquatics meeting- Discussed potential applicability of ABCD model to

			our project and fish database structure
Bob Miltner	Ohio EPA	Y	Requested and aquired Ohio EPA fish community data
Doug Carlson	NY Department of Environmental Conservation	Y	Met at NY Sustainable Flow workshops and on multiple 1 on 1 meetings (hosted by TNC and NY COOP)-discussed application of NY data
Fred Henson	NY Department of Environmental Conservation	Y	Met at NY Sustainable Flow workshops and discussed through email conversations (hosted by TNC and NY COOP)-discussed application of NY data
Mark Hartle	Pennsylvania Fish and Boat Commission	Y	Met at Upper Ohio River Environmental Flow workshops-Inquired about applicability of PA fish data to environmental flow relationships

Georeferenced Stream Biological Database

Summary

A fish database representing the four states (NY, PA, WVA, OH) comprising the majority of the Marcellus Shale region was created using available data from state and federal agencies. Data from the Ohio Environmental Protection Agency (OEPA), the United States Geological Survey (USGS) (NAQWA) program, and the United States Environmental Protection Agency (USEPA) Mid-Atlantic EMAP program were reformatted and combined with existing compiled state agency data for NY, PA, and WVA in the Multistate Aquatic Resource Information System (MARIS). We used the MARIS database structure to format all data in an Access database. The fish database will be linked to flow modeling efforts and used to assess flow-ecology relationships in the next phase of this project. The database with the fish information is included in this report.

Objectives

One of the objectives of phase 1 of this project was to develop a georeferenced summary assessment of the adequacy of available ecological data to inform ecological flow models for streams within the Marcellus Shale region. This involved acquiring georeferenced fish data from multiple agencies who collect data within the Marcellus Shale region (NY, PA, WVA, OH), formatting data to a standard database structure, and assessing the types of data available and the adequacy of different data types for modeling flow ecology relationships in the future.

Database development

We used the Multistate Aquatic Resource Information System (MARIS) as a platform for building the Marcellus Shale region fish database. MARIS is a platform hosted by the USGS to share existing state fisheries data across the US. Types of data that can be incorporated into the database include:

- geo-referencing data,
- event information including
- collection gear
- total catch and weight by species

We chose the MARIS platform as a template for the Marcellus Shale region fish database because it provided distinct advantages including:

1. a standardized data template for combining fish data from various sources that has been vetted by the National Fish Habitat Action Plan,
2. a considerable amount of fish data from within the Marcellus Shale region has already been compiled in MARIS, and
3. using the MARIS platform as a database template provides future opportunities for our data acquisition efforts from this project to be incorporated into the online MARIS platform. Currently we are only using the MARIS platform as a database template for this project. The Marcellus Shale Fish database has not been incorporated into the "official" MARIS platform publically available online at http://www.marisdata.org/. However, we consulted with Andy Loftus (MARIS Coordinator) during the database development, and have kept detailed notes of how new data was formatted to load into the MARIS platform (Appendix X), to facilitate potentially moving the database into the larger public MARIS platform in the future.

We used the MARIS developer template to combine existing MARIS data (NY, PA, WVA) with additional fish data from state and federal sources. We acquired additional fish data from:

- Ohio collected by OEPA (contact: Bob Miltner, bob.miltner@epa.state.oh.us);
- US EPA's Mid Atlantic Streams Data Sets downloaded from (http://www.epa.gov/emap/html/data/surfwatr/data/index.html); and
- USGS NAQWA data downloaded from the USGS BioData website (https://aquatic.biodata.usgs.gov/landing.action).

Several factors had to be addressed for successful incorporation of these datasets into the MARIS platform. Additional taxa, inconsistencies in naming, and non fish vertebrates were assessed against the MARIS species lookup table (tbl_fish_species_lookup). Twenty-two new entries that comprised primarily new hybrid combinations (20), a subspecies, and an exotic species record were added to the species lookup table from the Ohio dataset. In follow-up conversations with Bob Miltner concerning hybrids, he suggested aggregating hybrids at the genus level. However, to maintain consistency with the MARIS platform, we maintained hybrids, which can always be aggregated later for future analyses. Eight additional taxa were added to the species lookup table from the US EPA dataset. Additionally, 14 taxa comprising miscellaneous records for snakes, turtles, salamanders, frogs, invertebrates and unidentified taxa were removed from the US EPA dataset. Five taxa representing hybrids or higher taxonomic status were added to the species lookup table from the USGS dataset.

Fields that corresponded with fields in the MARIS location table template (tbl_location) were identified in the OEPA, USEPA and USGS databases and changed to correct field ids for merging into the MARIS location table. These changes are summarized in Table 1. Some fields were calculated through spatial joins in GIS to update geo-referenced data fields. This process was also performed to update fields and append new data to the tbl_fish_info table in the MARIS platform (Table 2).

Tables representing locations, fish records, species info, and dataset and originator ids from each database were appended to a new MARIS template. All new species in the species lookup table were assigned 7 digit maris_fishspecies_id numbers that start with 999 to clearly differentiate them from “official” MARIS fishspecies ids so that the MARIS folks could decide on numbers later if this dataset is incorporated into MARIS at a later time.

Table 1. Corresponding MARIS and dataset fields for location table

Attribute #	MARIS	MARIS NY	MARIS PA	MARIS WVA	OHIO EPA	EMAP	NAQUWA
1	maris_join_id	X	X	X	X	X	
2	State	X	X	X	=”OH”	STATE	StateAbb (SiteInfo)
3	originator_id	X	X	X	=”50”	=”51”	=”52”
4	dataset_id	X	X	X	=”50”	=”51”	=”52”
5	maris_water_id	X	X	X	=”OH-OEPA-[originator_water_id]”		
6	originator_water_id	X	X	X	RIVERCODE		
7	originator_station_id	X	X	X	STORET	STRM_ID	SiteNumber (SiteInfo)
8	originator_join_id						
9	water_name	X	X	X		DRAINB (93-96)	
10	station_name		X	X		STRMNAME	
11	originator_station_desc				NAME		
12	water_type	X	X	X	SITE_TYPE	=”STREAM”	=”STREAM”
13	wt_code	X	X	X	Coded based on MARIS #s	Coded based on MARIS #s	Coded based on MARIS #s
14	coll_loc_type	X	X	X	=”SMALL AREA”	=”SMALL AREA”	=”SMALL AREA”

15	lat	X	X	X	LATITUDE	LAT_DD	Latitude_dd (SIteInfo)
16	lon	X	X	X	LONGITUDE	LON_DD	Longitude_dd(SiteInfo)
17	coord_loc_cd	X	X	X			
18	coll_acc_desc	X	X	X			
19	upstream_lat						
20	upstream_lon						
21	fips_county_maris	X	X	X			
22	fips_state	X	X	X	Added from GIS	Added from GIS	StateFIPSCode (SiteInfo)
23	county_name	X	X	X	Added from GIS	COUNTY	County
24	cong_dist	X	X	X	Added from GIS	Added from GIS	Added from GIS
25	plss_section						
26	plss_township						
27	plss_range						
28	usgs_huc_8	X	X	X	Added	Added from GIS	HUCCODE (SiteInfo)
29	usgs_huc_10	X	X	X	Added	Added from GIS	Added from GIS
30	usgs_huc_12	X	X	X	HUC	Added from GIS	Added from GIS

31	usgs_huc_10_name	X	X	X	Added from GIS	Added from GIS	Added from GIS
32	usgs_huc_12_name	X	X	X	Added from GIS	Added from GIS	Added from GIS
33	nhd_reach						
34	nhdplus_v1_reach						
35	nhdplus_v2_reach						
36	originator_fips_county		X		Added from GIS	Added from GIS	CountyFIPSCode (SiteInfo)
37	comment			X			

Table 2. Corresponding MARIS and dataset fields for fish info table

Attribute #	MARIS	MARIS NY	MARIS PA	MARIS WVA	OHIO EPA	USEPA EMAP	USGS NAWQA
1	fishinfo_id	X	X	X			
2	maris_join_id	X	X	X			
3	state	X	X	X	"OH"	STATE	StateAbb (SiteInfo)
4	originator_id	X	X	X	"33"	"34"	"52"
5	dataset_id	X	X	X	"33"	"34"	"52"
6	originator_water_id	X	X	X	Linked from Species lookup table		
7	originator_station_id	X	X	X	STORET	STRM_ID	SiteNumber (FishCount)
8	originator_join_id						
9	sample_begin_date	X	X	X	TDATE	DATE_COL	CollectionDate (FishCount)
10	sample_end_date	X			TDATE	DATE_COL	CollectionDate (FishCount)
11	target_species	X	X	X	"ALL"	"ALL"	"ALL"
12	target_std	X	X	X	"ALL"	"ALL"	"ALL"
13	maris_fishspecies	X	X	X	Link from species lookup	Link from species	Link from species

	_id				table	lookup table	lookup table
14	originator_species _id		X	X	FINCODE	VERTCOD E	PublishedSortOrder (FishCount)
15	originator_itis_tsn	X	X				IT IS_TSN (FishCOunt)
16	maris_itis_tsn	X	X	X	Linked from Species lookup table	Linked from Species lookup table	Linked from Species lookup table
17	originator_sci_na me			X	Link to species table	GENUS + SPECIES	PublishedTaxonName (FishCount)
18	te_species_flag						
19	originator_sample _id	Combined originator_samp le_id and sample begin date	Combined originator_samp le_id and sample begin date	X	Combined originator_id and sample begin date columns to create a unique id code.	STRM_ID + VISIT_NO	SiteVisitSampleNumber (FishCount)
20	gear_type_1	X	X	X	TYPE-Changed OEPA codes to MARIS codes	"EL"	Update codes
21	gear_desc_1	X	X	X	Added narrative of	"Backpack Electrofishi	Gear (FishMethodANd SubreachInfo)

					OEPA sampling types which includes several different Electrofishing methods-These include the original TYPE codes from OEPA	ng”	
22	gear_type_2						
23	gear_desc_2						
24	sampling_methodology					“Backpack Electrofishing”	Pass (FishMethodAndSubreachInfo)
25	total_catch	X	X	X	COUNTED	ABUND	Abundance (FishCount)
26	total_weight		X*		TOTAL_WEIGHT (changed from grams to Kilograms)		

Table 2 continued.

Attribute #	**MARIS**	**MARIS NY**	**MARIS PA**	**MARIS WVA**	**OHIO EPA**	**USEPA EMAP**	**USGS NAWQA**
27	effort_time		X*	X	TIME_FISHED		Seconds Shock Time(FishMethodAndSubreachInfo)
28	time_units		X*	X	Changed from seconds to hours		
29	effort_area_dist	X	X*		DISTANCE_FISHED		
30	area_dist_units	X	X*		METERS		
31	cpue_time		X*	x	Calculated as total_catch/effort_time		
32	cpue_space		X*		Calculated as total_catch/effort_area_dist		
33	bpue_time		X*		Calculated as total_weight/effort_time		
34	bpue_space		X*		Calculate as total_weight/effort_area_dist		
35	pop_est						
36	pop_est_method						
37	pop_est_model						

38	pop_est_area						
39	pop_est_measure						
40	pop_est_measure_units						
41	biomass_est						
42	sample_desc						
43	Comment						

*not calculated for all samples.

Database structure

The Marcellus Shale Fish database consists of five main tables. These tables include:

- tbl_datasets_marcellus: provides info on original dataset and states represented for each dataset_id
- tbl_originators_marcellus: provides information on data collecting agency and states representd for each originator_id
- tbl_fish_species_lookup_marcellus: provides unique ids (maris_fishspecies_id), common names and scientific names at family, genus and species level (Table 3)
- tbl_loc_info_marcellus: provides unique ids for collection sites (originator_station_id) and associated site information, including latitude and longitude, which can be used to link location info to fish collection info in tbl_fish_info_marcellus. Additional queries were run to create refined location tables (Table 4):
    - tbl_location_marcellus_state_stream_sites: all stream fish collection sites within states or ecoregions that overlap the Marcellus boundary
    - tbl_location_marcellus_all_sites: all fish collection sites within the Marcellus boundary
- tbl_fish_info_marcellus: provides unique ids for each collection event (originator_sample_id) which can be used to link collection information (date, collection methods, effort, species, abundance) with site information in the tbl_loc_info_marcellus table (Table 5).

Information from the last three tables (tbl_fish_species_lookup_marcellus, tbl_location_info_marcellus (or any of the refined location tables), and tbl_fish_info_marcellus) can be combined based on unique ids (highlighted in green in Tables 3-5) and queried based on criteria in the tables (i.e. collection method, targeted sampling verses community sampling, etc.) to develop fish datasets for different analyses in the future.

Table 3. Fields included in the tbl_fish_species_lookup_marcellus table. Green highlighted row is unique id.

Field Name	Data Type	Description
maris_fishspecies_id	Number	Unique identifier for tbl_fish_species_ lookup
itis_tsn	Number	Integrated Taxonomic Information System, Taxonomic Serial Number
comm_name	Text	common name
hybrid_genus_spp_group	Text	categorization for hybrids, genus, species, or species groups
sci_name	Text	scientific name - genus and species
family	Text	family name
genus	Text	genus name
species	Text	species name

Table 4. Fields included in the tbl_location_info_marcellus table. Green highlighted row is unique id.

Field Name	Data Type	Description
state	Text	Mandatory - State Postal Code Abbreviation
originator_water_id	Text	If populated, originator's Foreign Key to their sampling locations in tbl_location when coll_loc_type = "ENTIRE"
originator_station_id	Text	If populated, originator's Foreign Key to their sampling locations in tbl_location when coll_loc_type = "SMALL AREA"
originator_join_id	Text	Originator's Foreign Key to their sampling locations in tbl_location
sample_begin_date	Date/Time	Mandatory - Date of the beginning of data collection for the sampling event in MM/DD/YYYY format

sample_end_date	Date/Time	Mandatory - Date of the end of data collection for the sampling event in MM/DD/YYYY format
target_species	Text	Species or species group code of fish targeted during the sampling effort If the target is unknown, enter "UNKNOWN", if all species are targeted (ie, a fish community estimate), enter "ALL"
target_std	Text	General standardized target group to separate fish community sampling from other sampling ("ALL", "TARGET", or "UNKNOWN")
maris_fishspecies_id	Number	Foreign key to tbl_fish_species_ lookup
originator_species_id	Text	Mandatory - The code used by the originator to designate the species in the database
originator_itis_tsn	Number	Integrated Taxonomic Information System, Taxonomix Serial Number (if provided by originator)
originator_sci_name	Text	The scientific name provided by the originator
originator_sample_id	Text	Sample ID from the originator dataset
gear_type_1	Text	FN = Fyke Net, TN=Trap Net, GN=Gill Net, PN=Pound Net, EL=Electrofishing, SE=Seine, TR=Trawl, EP=Eel pot, FP=Fish Pot,CP=Crab Pot, SN=snorkel, HL=Hook & Line, KN=Kick Net, HN = Hoop Net, RO=rotenone, OT=Other
gear_desc_1	Text	Detailed description of primary gear used in fish collection (originator specific, not standardized)
gear_type_2	Text	FN = Fyke Net, TN=Trap Net, GN=Gill Net, PN=Pound Net, EL=Electrofishing, SE=Seine, TR=Trawl, EP=Eel pot, FP=Fish Pot,CP=Crab Pot, SN=snorkel, HL=Hook & Line, KN=Kick Net, HN = Hoop Net, RO=rotenone, OT=Other
gear_desc_2	Text	Detailed description of the secondary gear used in fish collection (originator specific, not standardized)
sampling_methodology	Text	Supplemental information on the method used to sample, such as single pass electrofishing, multi-pass electrofishing, etc Note: if multipass, only the first pass of

		data for "count" or "CPUE" should be included
total_catch	Number	Total number of fish caught in sample If species occurrence is noted but not enumerated, this field should be blank
total_weight	Number	Total weight (kilograms) of fish caught in sample for surveys where all fish were weighed If species occurrence is noted but not enumerated, this field should be blank
effort_time	Text	Total duration of sampling effort, not including processing time
time_units	Text	Originator standard unit of time for sampling effort (HOURS, DAYS, NETNIGHTS, HAULS)

Table 4 continued.

Field Name	**Data Type**	**Description**
effort_area_dist	Number	Total area or distance sampled
area_dist_units	Text	originator standard unit of space (area or distance) for sampling effort (METERS, MILES, HECTARES, KILOMETERS, FEET, or ACRES)
cpue_time	Number	Catch Per Unit Effort Time The total number of fish caught per standard unit of time - TOTAL_CATCH/EFFORT_TIME for a single gear type DO NOT FILL IN IF GEAR TYPE2 IS POPULATED
cpue_space	Number	Catch Per Unit Effort Space The total number of fish caught per standard unit of space (TOTAL_CATCH/EFFORT_AREA_DIST) for a single gear type DO NOT FILL IN IF GEAR TYPE2 IS POPULATED
bpue_time	Number	Catch Per Unit Effort Time Biomass The total weight of fish caught per standard unit of time (TOTAL_WEIGHT/EFFORT_TIME) for a single gear type DO NOT FILL IN IF GEAR TYPE2 IS

		POPULATED
bpue_space	Number	Catch Per Unit Effort Space Biomass The total weight of fish caught per standard unit of space (TOTAL_WEIGHT/EFFORT_AREA_DIST) for a single gear type DO NOT FILL IN IF GEAR TYPE2 IS POPULATED
pop_est	Number	Population Estimate for sample reach Sampling technique and estimator used is specific to the originator Consult the originator's metadata
pop_est_method	Text	Sampling method used to estimate population abundance (SCMR, MCMR, DEP, or OTHER) SCMR = Single Census Mark-Recapture MCMR = Multiple Census Mark-Recapture, DEP = Depletion, or OTHER
pop_est_model	Text	Population abundance estimator used Choices include CHAPMAN, PETERSON, SCHNABEL, DE LURY, CORMACK JOLLY SEBER, or OTHER
pop_est_area	Text	Population estimate is for the entire waterbody or a smaller area of the waterbody Choices ENTIRE, or SMALL AREA
pop_est_measure	Number	For subsections of the waterbody, linear or areal distance for which the population estimate is measured (eg, 1000)
pop_est_measure_units	Text	Units used to measure POP_EST_MEASURE (METERS, MILES, HECTARES, KILOMETERS, ACRES)
biomass_est	Number	Biomass Estimate for sample reach Sampling technique andmethod is specific to the originator Consult the originator's metadata
sample_desc	Text	A brief description of sampling event
comment	Text	General Comment
te_species_flag	Text	Flag provided by originator to indicate species is threatened or endangered, and location information should be withheld from query results

Table 5. Fields included in the tbl_fish_info_marcellus table. Green highlighted row is unique id and light green rows are unique ids from species lookup and location tables.

Field Name	Data Type	Description
state	Text	Mandatory - State Postal Code Abbreviation
originator_water_id	Text	If populated, originator's Foreign Key to their sampling locations in tbl_location when coll_loc_type = "ENTIRE"
originator_station_id	Text	If populated, originator's Foreign Key to their sampling locations in tbl_location when coll_loc_type = "SMALL AREA"
originator_join_id	Text	Originator's Foreign Key to their sampling locations in tbl_location
sample_begin_date	Date/Time	Mandatory - Date of the beginning of data collection for the sampling event in MM/DD/YYYY format
sample_end_date	Date/Time	Mandatory - Date of the end of data collection for the sampling event in MM/DD/YYYY format
target_species	Text	Species or species group code of fish targeted during the sampling effort If the target is unknown, enter "UNKNOWN", if all species are targeted (ie, a fish community estimate), enter "ALL"
target_std	Text	General standardized target group to separate fish community sampling from other sampling ("ALL", "TARGET", or "UNKNOWN")
maris_fishspecies_id	Number	Foreign key to tbl_fish_species_ lookup
originator_species_id	Text	Mandatory - The code used by the originator to designate the species in the database
originator_itis_tsn	Number	Integrated Taxonomic Information System, Taxonomix Serial Number (if provided by originator)
originator_sci_name	Text	The scientific name provided by the originator
originator_sample_id	Text	Sample ID from the originator dataset
gear_type_1	Text	FN = Fyke Net, TN=Trap Net, GN=Gill Net, PN=Pound Net, EL=Electrofishing, SE=Seine, TR=Trawl, EP=Eel

		pot, FP=Fish Pot,CP=Crab Pot, SN=snorkel, HL=Hook & Line, KN=Kick Net, HN = Hoop Net, RO=rotenone, OT=Other
gear_desc_1	Text	Detailed description of primary gear used in fish collection (originator specific, not standardized)
gear_type_2	Text	FN = Fyke Net, TN=Trap Net, GN=Gill Net, PN=Pound Net, EL=Electrofishing, SE=Seine, TR=Trawl, EP=Eel pot, FP=Fish Pot,CP=Crab Pot, SN=snorkel, HL=Hook & Line, KN=Kick Net, HN = Hoop Net, RO=rotenone, OT=Other
gear_desc_2	Text	Detailed description of the secondary gear used in fish collection (originator specific, not standardized)
sampling_methodology	Text	Supplemental information on the method used to sample, such as single pass electrofishing, multi-pass electrofishing, etc Note: if multipass, only the first pass of data for "count" or "CPUE" should be included
total_catch	Number	Total number of fish caught in sample If species occurrence is noted but not enumerated, this field should be blank
total_weight	Number	Total weight (kilograms) of fish caught in sample for surveys where all fish were weighed If species occurrence is noted but not enumerated, this field should be blank
effort_time	Text	Total duration of sampling effort, not including processing time
time_units	Text	Originator standard unit of time for sampling effort (HOURS, DAYS, NETNIGHTS, HAULS)

Table 5 continued

Field Name	Data Type	Description
effort_area_dist	Number	Total area or distance sampled
area_dist_units	Text	originator standard unit of space (area or distance) for sampling effort (METERS, MILES, HECTARES, KILOMETERS, FEET, or ACRES)
cpue_time	Number	Catch Per Unit Effort Time The total number of fish caught per standard unit of time - TOTAL_CATCH/EFFORT_TIME for a single gear type DO NOT FILL IN IF GEAR TYPE2 IS POPULATED
cpue_space	Number	Catch Per Unit Effort Space The total number of fish caught per standard unit of space (TOTAL_CATCH/EFFORT_AREA_DIST) for a single gear type DO NOT FILL IN IF GEAR TYPE2 IS POPULATED
bpue_time	Number	Catch Per Unit Effort Time Biomass The total weight of fish caught per standard unit of time (TOTAL_WEIGHT/EFFORT_TIME) for a single gear type DO NOT FILL IN IF GEAR TYPE2 IS POPULATED
bpue_space	Number	Catch Per Unit Effort Space Biomass The total weight of fish caught per standard unit of space (TOTAL_WEIGHT/EFFORT_AREA_DIST) for a single gear type DO NOT FILL IN IF GEAR TYPE2 IS POPULATED
pop_est	Number	Population Estimate for sample reach Sampling technique and estimator used is specific to the originator Consult the originator's metadata
pop_est_method	Text	Sampling method used to estimate population abundance (SCMR, MCMR, DEP, or OTHER) SCMR = Single Census Mark-Recapture MCMR = Multiple Census Mark-Recapture, DEP = Depletion, or OTHER
pop_est_model	Text	Population abundance estimator used Choices include CHAPMAN, PETERSON, SCHNABEL, DE LURY,

		CORMACK JOLLY SEBER, or OTHER
pop_est_area	Text	Population estimate is for the entire waterbody or a smaller area of the waterbody Choices ENTIRE, or SMALL AREA
pop_est_measure	Number	For subsections of the waterbody, linear or areal distance for which the population estimate is measured (eg, 1000)
pop_est_measure_units	Text	Units used to measure POP_EST_MEASURE (METERS, MILES, HECTARES, KILOMETERS, ACRES)
biomass_est	Number	Biomass Estimate for sample reach Sampling technique andmethod is specific to the originator Consult the originator's metadata
sample_desc	Text	A brief description of sampling event
comment	Text	General Comment
te_species_flag	Text	Flag provided by originator to indicate species is threatened or endangered, and location information should be withheld from query results

Fish Field Collection Information Summary

Sample locations

The Marcellus Shale fish database includes existing MARIS fish data for NY (1976-2007), PA (1975-2007), and WVA (1997-2010) with additional data from Ohio EPA (1978-2012), the USEPA EMAP program (1993-1998), and the USGS NAWQA program (1993-2012). There are 35512 locations represented within the database (tbl_location_marcellus_all_sites). There are 14707 unique stream fish collection locations within the Marcellus Shale boundary (tbl_loc_marcellus_fish_locations, Figure 1). In total, there are 437045 fish records within the database (tbl_fish_info_marcellus) with 151151 individual species counts recorded from sites within the Marcellus Shale boundary.

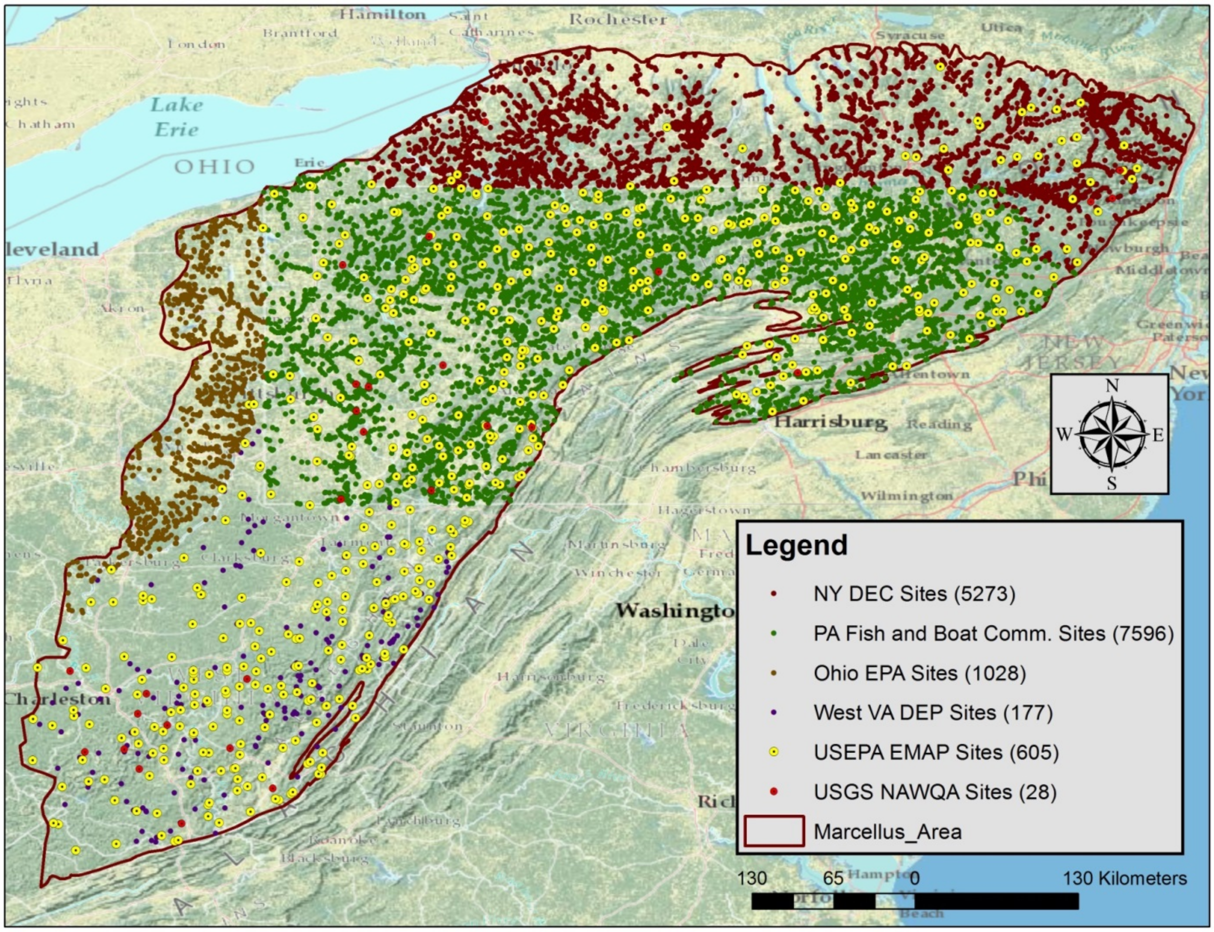


Figure 1. Distribution of sampling site according to source extracted from the Marcellus Shale Fish database that fall within the Marcellus Shale boundary (Maya add source info for boundary here).

Unique taxons

There were 287 unique Maris fish species ids represented by collections from within the Marcellus Shale region, including new additions from this project (Table 6). These taxa ids represent 27 families and 81 genera of fish. The top 25 most frequently collected taxa in rank order are white suckers, creek chubs, blacknose dace, brown trout, central stonerollers, mottled sculpin, northern hog suckers, smallmouth bass, Johnny darters, brook trout, longnose dace, greenside darters, bluntnose minnows, common shiners, blue gill, fantail darters, rock bass, pumpkinseed, largemouth bass, rainbow darters, cutlips minnows, green sunfish, river chub, common carp, and logperch. Forty-two are hybrids and an additional 26 are higher level taxonomic ids (family or genus). Future development of relationships between flow metrics and fish responses will have to determine the best approach for incorporating hybrid or family/genus data, but this will be dependent on the response measures. Additional taxa may need to be condensed for future analyses (subspecies, different strains). Two darter species not known to occur within the region were extracted from the database. There are 1077 records for the barrens darter from PA. This is likely to be the banded darter and needs to be resolved. Additionally there is a record for the Tennessee darter in the database that must be an error.

Table 6. Taxonomic ids extracted from sampling events collected within the Marcellus Shale region.

Common name	**Genus species**	**# of records**
Bowfin	*Amia calva*	8
American eel	*Anguilla rostrata*	330
Brook silverside	*Labidesthes sicculus*	167
River carpsucker	*Carpiodes carpio*	62
Quillback	*Carpiodes cyprinus*	436
Highfin carpsucker	*Carpiodes velifer*	9
Suckers (Family)	*Catostomidae spp.*	83
Longnose sucker	*Catostomus catostomus*	37
White sucker	*Catostomus commersonii*	9193
Suckers (Genus)	*Catostomus spp.*	3
Blue sucker	*Cycleptus elongatus*	1
Creek chubsucker	*Erimyzon oblongus*	65
Lake chubsucker	*Erimyzon sucetta*	1
Northern hog sucker	*Hypentelium nigricans*	5083
Smallmouth buffalo	*Ictiobus bubalus*	253
Bigmouth buffalo	*Ictiobus cyprinellus*	1
Black buffalo	*Ictiobus niger*	22
Spotted sucker	*Minytrema melanops*	162
Silver redhorse	*Moxostoma anisurum*	586

smallmouth redhorse	*Moxostoma breviceps*	164
River redhorse	*Moxostoma carinatum*	138
Black Jumprock	*Moxostoma cervinum*	1
Black redhorse	*Moxostoma duquesnii*	560
Golden redhorse	*Moxostoma erythrurum*	1714
Shorthead redhorse	*Moxostoma macrolepidotum*	375
Redhorses	*Moxostoma spp.*	18
Greater redhorse	*Moxostoma valenciennesi*	7
Torrent Sucker	*Thoburnia rhothoeca*	3
Centrarchidae hybrid	*Centrarchidae hybrid*	1
Rock bass	*Ambloplites rupestris*	3109
Unspecified Centrarchid spp.	*Centrarchid spp.*	3
Warmouth	*Chaenobryttus gulosus*	184
Bluespotted sunfish	*Enneacanthus gloriosus*	25
Banded sunfish	*Enneacanthus obesus*	4
Redbreast sunfish	*Lepomis auritus*	278
Green sunfish	*Lepomis cyanellus*	2108
Pumpkinseed	*Lepomis gibbosus*	2929
Orangespotted sunfish	*Lepomis humilis*	24
Green sunfish X Unknown	*Lepomis cyanellus x centrarchidae*	136
Green sunfish X Pumpkinseed	*Lepomis cyanellus x lepomis gibbosus*	64
Green sunfish x Warmouth hybrid	*Lepomis cyanellus x lepomis gulosus*	3
Green sunfish X Bluegill	*Lepomis cyanellus x lepomis macrochirus*	194

Table 6 continued.

Common name	**Genus species**	**# of records**
Green sunfish X Longear sunfish	*Lepomis cyanellus x lepomis megalotis*	39
Pumpkinseed x Warmouth hybrid	*Lepomis gibbosus x lepomis gulosus*	1
Pumpkinseed X Orangespotted sunfish	*Lepomis gibbosus x lepomis humilis*	1
Pumpkinseed x Bluegill hybrid	*Lepomis gibbosus x lepomis macrochirus*	27
Pumpkinseed X Longear sunfish	*Lepomis gibbosus x lepomis megalotis*	1
Warmouth x Bluegill hybrid	*Lepomis gulosus x lepomis macrochirus*	1
Lepomis hybrids	*Lepomis hybrids*	42
Bluegill x Orangespotted sunfish hybrid	*Lepomis macrochirus x l. humilis*	1
Longear sunfish x Bluegill hybrid	*Lepomis megalotis x l. macrochirus*	9
Hybrid Sunfishes	*Lepomis spp. (hybrid)*	1
Sunfish hybrid	*Lepomis spp. X Lepomis spp.*	62
redbreast x green sunfish	*Lepomis auritus x Lepomis cyanellus*	1
Bluegill	*Lepomis macrochirus*	3266
Longear sunfish	*Lepomis megalotis*	356
Redear sunfish	*Lepomis microlophus*	53
Sunfishes	*Lepomis spp.*	9
Smallmouth bass	*Micropterus dolomieu*	4590
Spotted bass	*Micropterus punctulatus*	534
Largemouth bass	*Micropterus salmoides*	2263

White crappie	*Pomoxis annularis*	431
Pomoxis hybrids	*Pomoxis spp. (hybrids)*	9
Black crappie	*Pomoxis nigromaculatus*	626
Black crappie (blacknose)	*Pomoxis nigromaculatus (blacknose)*	1
Blueback herring	*Alosa aestivalis*	11
Skipjack herring	*Alosa chrysochloris*	64
Alewife	*Alosa pseudoharengus*	34
American shad	*Alosa sapidissima*	20
Herrings	*Alosa spp.*	1
Gizzard shad	*Dorosoma cepedianum*	1043
Oriental weatherfish	*Misgurnus anguillicaudatus*	2
Mottled sculpin	*Cottus bairdii*	5116
Blue Ridge sculpin	*Cottus caeruleomentum*	2
Banded sculpin	*Cottus carolinae*	3
Slimy sculpin	*Cottus cognatus*	1366
Potomac sculpin	*Cottus girardi*	2
Kanawha Scuplin	*Cottus kanawhae*	1
Checkered Sculpin	*Cottus n. sp.*	42
Sculpins	*Cottus spp.*	584
Hybrid x Minnow	*Minnow hybrid*	13
Central stoneroller	*Campostoma anomalum*	6087
Goldfish	*Carassius auratus*	119
Redside dace	*Clinostomus elongatus*	1115
Rosyside dace	*Clinostomus funduloides*	36
Redside dace X Creek chub	*Clinostomus elongatus x semotilus*	1

	atromaculatus	
Redside Dace x Striped Shiner	*Clinostomus elongatus x luxilus chrysocephalus*	1
Lake chub	*Couesius plumbeus*	5

Table 6 continued.

Common name	**Genus species**	**# of records**
Grass carp	*Ctenopharyngodon idella*	2
Triploid grass carp	*Ctenopharyngodon idella (triploid)*	2
Satinfin shiner	*Cyprinella analostana*	77
Whitetail shiner	*Cyprinella galactura*	6
Cyprinella Hybrid	*Cyprinella Sp. x Cyprinella Sp.*	1
Spotfin shiner	*Cyprinella spiloptera*	1240
Satinfin shiners	*Cyprinella spp.*	24
Steelcolor shiner	*Cyprinella whipplei*	14
Undetermined CYPRINID	*Cyprinidae spp. (Undetermined)*	2
Sheepshead minnow	*Cyprinodon variegatus*	1
Common carp	*Cyprinus carpio*	1908
Carp x Goldfish hybrid	*Cyprinus carpio x carassius auratus*	64
Streamline chub	*Erimystax dissimilis*	80
Gravel chub	*Erimystax x-punctatus*	2
Tonguetied minnow	*Exoglossum laurae*	107
Cutlips minnow	*Exoglossum maxillingua*	2222
Brassy minnow	*Hybognathus hankinsoni*	4
Eastern silvery minnow	*Hybognathus regius*	8

Bigeye chub	*Hybopsis amblops*	42
White Shiner	*Luxilus albeolus*	2
Crescent Shiner	*Luxilus cerasinus*	3
Striped shiner	*Luxilus chrysocephalus*	1302
Common shiner	*Luxilus cornutus*	3303
Common shiner X Striped shiner	*Luxilus cornutus x luxilus chrysocephalus*	8
Striped shiner X River chub	*Luxilus chrysocephalus x nocomis micropogon*	15
Striped Shiner x Rosyface Shiner	*Luxilus chrysocephalus x notropis rubellus*	56
Striped Shiner x Creek Chub	*Luxilus chrysocephalus x semotilus atromaculatus*	2
Striped Shiner x Stoneroller	*Luxilus chrysocephalus x campostoma anomalum*	2
Striped Shiner x Silver Shiner	*Luxilus chrysocephalus x notropis photogenis*	1
Striped Shiner x Southern Redbelly Dace	*Luxilus chrysocephalus x phoxinus erythrogaster*	1
Striped Shiner x Redfin Shiner	*Luxilus chrysocephalus x lythrurus umbratilis*	1
highscale shiners	*Luxilus spp.*	2
Rosefin shiner	*Lythrurus ardens*	3
Scarletfin shiner	*Lythrurus fasciolaris*	3
Redfin shiner	*Lythrurus umbratilis*	133
Silver chub	*Macrhybopsis storeriana*	118
Pearl dace	*Margariscus margarita*	249
Hornyhead chub	*Nocomis biguttatus*	57

River Chub x Stoneroller	*Nocomis micropogon x campostoma anomalum*	3
Bluehead chub	*Nocomis leptocephalus*	5
River chub	*Nocomis micropogon*	2103
Bigmouth Chub	*Nocomis platyrhynchus*	14
Golden shiner	*Notemigonus crysoleucas*	833
Comely shiner	*Notropis amoenus*	25
Popeye shiner	*Notropis ariommus*	2
Emerald shiner	*Notropis atherinoides*	802

Table 6 continued.

Common name	**Genus species**	**# of records**
Bridle shiner	*Notropis bifrenatus*	6
River shiner	*Notropis blennius*	17
Silverjaw minnow	*Notropis buccatus*	993
Ghost shiner	*Notropis buchanani*	5
Ironcolor shiner	*Notropis chalybaeus*	1
Bigmouth shiner	*Notropis dorsalis*	19
Blackchin shiner	*Notropis heterodon*	4
Blacknose shiner	*Notropis heterolepis*	4
Spottail shiner	*Notropis hudsonius*	685
Pugnose shiner X Blackchin shiner	*Notropis anogenus x notropis heterodon*	16
Rosyface Shiner x Silver Shiner	*Notropis rubellus x notropis photogenis*	2

Sand Shiner x Silver Shiner	*Notropis stramineus x notropis photogenis*	1
Silver shiner	*Notropis photogenis*	659
Swallowtail shiner	*Notropis procne*	71
Rosyface shiner	*Notropis rubellus*	1519
New River Shiner	*Notropis scabriceps*	3
subspecies of mimic shiner	*Notropis sp cf volucellus*	6
Eastern shiners	*Notropis spp. (Eastern shiners)*	26
Sand shiner	*Notropis stramineus*	851
Telescope shiner	*Notropis telescopus*	14
Mimic shiner	*Notropis volucellus*	498
Channel shiner	*Notropis wickliffi*	86
Suckermouth minnow	*Phenacobius mirabilis*	18
Kanawha Minnow	*Phenacobius teretulus*	1
Blackside dace	*Phoxinus cumberlandensis*	1
Northern redbelly dace	*Phoxinus eos*	55
Southern redbelly dace	*Phoxinus erythrogaster*	218
mountain redbelly dace	*Phoxinus oreas*	5
Bluntnose minnow	*Pimephales notatus*	3712
Fathead minnow	*Pimephales promelas*	375
Bullhead minnow	*Pimephales vigilax*	30
Blacknose dace	*Rhinichthys atratulus*	7127
Longnose dace	*Rhinichthys cataractae*	4223
Blacknose Dace Hybrid (Rhinichthys atratulus x R.obtusus)	*Rhinichthys atratulus x R. obtusus*	1

Western blacknose dace	*Rhinichthys obtusus*	1065
Rudd	*Scardinius erythrophthalmus*	1
Creek chub	*Semotilus atromaculatus*	7652
Fallfish	*Semotilus corporalis*	1265
Creek chubs	*Semotilus spp.*	1
Grass pickerel	*Esox americanus*	90
Redfin pickerel (subspecies)	*Esox americanus americanus*	472
Grass pickerel (vermiculatus)	*Esox americanus vermiculatus*	58
Northern pike X Grass pickerel	*Esox lucius x esox americanus*	3
Tiger muskellunge	*Esox lucius x esox masquinongy*	92
Northern pike	*Esox lucius*	177
Muskellunge	*Esox masquinongy*	147

Table 6 continued.

Common name	**Genus species**	**# of records**
Chain pickerel	*Esox niger*	438
Pikes	*Esox spp.*	11
killifishes and Topminnows	*Fundulidae spp.*	1
Banded killifish	*Fundulus diaphanus*	54
Western Banded Killifish (subspecies)	*Fundulus diaphanus menona*	2
Mummichog	*Fundulus heteroclitus*	2
Blackstripe topminnow	*Fundulus notatus*	13
Burbot	*Lota lota*	91
Brook stickleback	*Culaea inconstans*	67

Grand Total	*Grand Total*	151151
Goldeye	*Hiodon alosoides*	4
Mooneyes	*Hiodon spp.*	1
Mooneye	*Hiodon tergisus*	62
White catfish	*Ameiurus catus*	14
Black bullhead	*Ameiurus melas*	120
Yellow bullhead	*Ameiurus natalis*	1341
Brown bullhead	*Ameiurus nebulosus*	1074
Channel catfish	*Ictalurus punctatus*	845
Catfishes	*Ictalurus spp.*	2
Mountain madtom	*Noturus eleutherus*	1
Slender madtom	*Noturus exilis*	1
Yellowfin madtom	*Noturus flavipinnis*	1
Stonecat	*Noturus flavus*	652
Tadpole madtom	*Noturus gyrinus*	5
Margined madtom	*Noturus insignis*	275
Brindled madtom	*Noturus miurus*	87
Madtoms	*Noturus spp.*	7
Flathead catfish	*Pylodictis olivaris*	233
Alligator gar	*Atractosteus spatula*	348
Spotted gar	*Lepisosteus oculatus*	1
Longnose gar	*Lepisosteus osseus*	149
White perch	*Morone americana*	35
White bass	*Morone chrysops*	388
Hybrid Striped Bass - Female	*Morone saxatilis x Morone americana*	5

sbass x male wperch		
Striped bass x White bass hybrid (Wiper)	*Morone saxatilis x morone chrysops*	109
Striped bass	*Morone saxatilis*	5
Rainbow smelt	*Osmerus mordax*	1
Greenside darter	*Etheostoma blennioides*	3799
Rainbow darter	*Etheostoma caeruleum*	2254
Bluebreast darter	*Etheostoma camurum*	23
Fantail darter	*Etheostoma flabellare*	3144
Barrens darter	*Etheostoma forbesi*	1077
Spotted darter	*Etheostoma maculatum*	7
Johnny darter	*Etheostoma nigrum*	4576
Tessellated darter	*Etheostoma olmstedi*	1785

Table 6 continued.

Common name	**Genus species**	**# of records**
Candy Darter	*Etheostoma osburni*	3
Eastern sand darter	*Etheostoma pellucidum*	24
Orangethroat darter	*Etheostoma spectabile*	15
Etheostoma	*Etheostoma spp.*	2
Tennessee darter	*Etheostoma tennesseense*	1
Tippecanoe darter	*Etheostoma tippecanoe*	5
Variegate darter	*Etheostoma variatum*	512
Banded darter	*Etheostoma zonale*	1280
Yellow perch	*Perca flavescens*	977

Logperch	*Percina caprodes*	1867
Channel darter	*Percina copelandi*	58
Gilt darter	*Percina evides*	16
Appalachia Darter	*Percina gymnocephala*	4
Longhead darter	*Percina macrocephala*	89
Blackside darter	*Percina maculata*	981
Sharpnose darter	*Percina oxyrhyncha*	6
Shield darter	*Percina peltata*	416
Slenderhead darter	*Percina phoxocephala*	22
Roanoke Darter	*Percina roanoka*	3
Dusky darter	*Percina sciera*	13
River darter	*Percina shumardi*	4
Roughbelly darters	*Percina spp.*	1
Sauger	*Sander canadensis*	636
Walleye X Sauger (Saugeye)	*Sander vitreus x sander canadensis*	54
Walleyes and Saugers	*Sander spp.*	102
Walleye	*Sander vitreus*	977
Trout-perch	*Percopsis omiscomaycus*	245
Ohio lamprey	*Ichthyomyzon bdellium*	48
Northern brook lamprey	*Ichthyomyzon fossor*	1
Northern brook lamprey (ammocoete)	*Ichthyomyzon fossor (ammocoete)*	2
Mountain brook lamprey	*Ichthyomyzon greeleyi*	29
Lampreys	*Ichthyomyzon spp.*	15
Lampreys (ammocoetes 1)	*Ichthyomyzon spp. (ammocoetes 1)*	7

Least brook lamprey	*Lampetra aepyptera*	121
American brook lamprey	*Lampetra appendix*	65
American brook lamprey (ammocoete)	*Lampetra appendix (ammocoete)*	68
Lampetra	*Lampetra spp.*	11
Sea lamprey (ammocoete)	*Petromyzon marinus (ammocoete)*	6
Sea lamprey	*Petromyzon marinus*	29
Unidentified lamprey	*Petromyzontidae spp.*	24
Western Mosquitofish	*Gambusia affinis*	7
Paddlefish	*Polyodon spathula*	1
Coho salmon	*Oncorhynchus kisutch*	9
Rainbow trout	*Oncorhynchus mykiss*	811
Rainbow trout (strain 3)	*Oncorhynchus mykiss (strain 3)*	403
Rainbow trout (strain 2)	*Oncorhynchus mykiss (strain 2)*	15

Table 6 continued.

Common name	**Genus species**	**# of records**
Rainbow trout (strain 1)	*Oncorhynchus mykiss (strain 1)*	19
Rainbow trout (domestic)	*Oncorhynchus mykiss (domestic)*	590
Atlantic salmon	*Salmo salar*	39
Brown trout (domestic)	*Salmo trutta (domestic)*	732
Brown trout	*Salmo trutta*	6109
trout hybrid	*Salmonidae spp.*	35
Brook trout (domestic)	*Salvelinus fontinalis (domestic)*	1419
Brook trout	*Salvelinus fontinalis*	4370
Tiger trout	*Salvelinus fontinalis x salmo trutta*	44
Lake trout	*Salvelinus namaycush*	1
Trouts	*Salvelinus and salmo spp.*	2
Freshwater drum	*Aplodinotus grunniens*	649
Central mudminnow	*Umbra limi*	237
Eastern mudminnow	*Umbra pygmaea*	23
UNKNOWN	*UNKNOWN*	15
No fish collected.		107

Sampling events

The distribution of fish sampling events varies across states with NY, PA and OH having significantly more sampling effort than WVA. A few sampling events from federal programs occurred in MD and VA but state agency data has not been included in the database at this time. Fish data is collected for a variety of reasons using different collection techniques. Collection purposes can range from surveys for rare or endangered species, general fish species distributions, calculating water quality indices such as the index of biotic integrity (IBI), and monitoring and managing sport fisheries. Within the Marcellus Shale fish database there are two fields that can be used to sort or select collection records by purpose (target_std) and method (gear_type_1). In general, targe_std is broken up into three categories, ALL, TARGET, or UNKNOWN. TARGET surveys represent collection events focused on a particular species or group of gamefish (trout, smallmouth bass). When the purpose of the survey was to describe the assemblage, it is assigned as ALL. UNKNOWN is self explanatory but can be assumed to represent general fish assemblage surveys but may not have been conducted with standard methods. Fish sampling recorded within the database was conducted with several types of sampling gear, with the dominant gear type being electro-fishing (Table 7). However, all electro-fishing efforts are not equal. New York and PA only use standard methods for conducting targeted game fish surveys (Table 7) (personal communication with Doug Carlson (dmcarlso@gw.dec.state.ny.us), Fred Henson (fghenson@gw.dec.state.ny.us), and Mark Hartle (mhartle@state.pa.us)). Thus targeted electro-fishing sampling events within these states may provide adequate abundance estimates for target species (trout), but not necessarily the whole fish assemblage. In contrast, Ohio EPA, WV DEP, USEPA and USGS collect fish with comparable standard electro-fishing methods designed to collect data for calculating biotic indices and can also be used to describe assemblage structure based on relative abundance estimates or presence/absence data (Table 8). Additionally, non target data from NY and PA will provide reliable presence/absence data which combined with the other data sources provides a considerable amount of data for developing individual species occupancy models or measures of assemblage structure or functional traits based on presence/absence data (Table 8).

Table 7. Distribution of sampling events within the Marcellus Shale boundary by agency, collection type, and collection method for each state. EL = electrofishing, GN = gillnet, OT=, SE = seinge, TN = trapnet, HL = , UN = unknown. Rows highlighted in green represent sample types that provide community information. Rows highlighted in orange represent targeted game fish sampling (primarily trout).

	MD	NY	OH	PA	VA	WVA	Total
NYDEC		**2960**					**2960**
All		1663					1663
EL		1414					1414
GN		78					78
OT		36					36
SE		118					118
TN		17					17
Target		1297					1297
EL		1272					1272
GN		22					22
SE		2					2
TN		1					1
PAFABC				**9790**			**9790**
All				10			10
OT				3			3
TN				7			7
Target				5060			5060
EL				5060			5060
Unknown				4720			4720
EL				4219			4219

GN				258			258
OT				4			4
SE				150			150
TN				7			7
UN				82			82
WVADEP						**196**	**196**
All						73	73
EL						73	73
Target						95	95
EL						9	9
HL						2	2
OT						84	84
Unknown						28	28
EL						27	27
OT						1	1
OEPA			**2221**				**2221**
All			2221				2221
EL			2219				2219
SE			2				2
USEPA	**7**	**22**		**154**	**6**	**104**	**293**
All	7	22		154	6	104	293
EL	7	22		154	6	104	293
USGS		**10**		**25**		**7**	**42**
All		10		25		7	42
EL		10		25		7	42

Total	7	2992	2221	9969	6	307	15502

Table 8. Distribution of non-targeted electro-fishing sampling events within the Marcellus Shale boundary by agency for each state. Number of sampling events that can be used to describe community structure or individual species occupancy based on presence/absence are summarized in P-A Data row. Number of sampling events that may be used for computing biotic indices or describing changes in community structure based on relative abundance are summarized in Rel Abd Data row.

	MD	NY	OH	PA	VA	WVA	Total
NYDEC							
All-EL		1414*					1414
PAFABC							
Unknown-EL				4219*			4219
WVDEP							
All-EL						73	73
Unknown-EL						27	27
OEPA							
All-EL			2219				2219
USEPA							
All-EL	7	22		154	6	104	293
USGS							
All-EL		10		25		7	42
P-A Data	**7**	**1446**	**2219**	**4398**		**211**	**8287**
Rel Abd Data	**7**	**32**	**2219**	**179**	**6**	**211**	**2654**

*Not necessarily collected with standard methods.

Potential Applications and Limitations

The large number of sampling events available for analyses should provide opportunities for several different kinds of analyses that relate fish response to modeled flow metrics. First target electro-fishing data can be used to relate trout abundance to landscape and local habitat, including modeled flow metrics. Additionally, this data can be combined with other data to model occupancy instead of abundance. Second, fish based response metrics or ecological trait measures (% fluvial fish) can be calculated from available relative abundance data (OEPA, WVDEP, USEPA, USGS) collected for biomonitoring purposes. This data may also be useful for conducting multivariate analysis (nMDS ordinations) to relate assemblage structure to modeled flow metrics across a range of sites.

This dataset is not without limitations including unequal distribution of sampling effort, a diversity of sampling methods, and differences in taxonomic resolution among sampling events. Unequal distribution of sampling effort among states is a common problem in trans-boundary datasets. Overcoming this may require taking random subsets of sampling events from data rich regions to even out the spatial distribution of data. Acquiring additional data may help fill in data sparse areas. Data from Maryland and Virginia state agencies have not been incorporated into this database and may provide additiona data in the small portions of the Marcellus Shale region that these states occupy. There is more fish data present in WVA, however WV Dept. of Natural Resources is not willing to release fish data to open source databases (Kauffman 2012). Additional sources of brook trout data (Trout Unlimited) could be added to strengthen brook trout datasets. While there are a diversity of sampling methods present in the dataset, electro-fishing is by far the dominant gear used to collect fish in the region. Limiting datasets to those that use electro-fishing gear should not limit available data very much. However, limiting datasets to those that electro-fish with standard methods for calculating biotic indices limits the amount of data substantially. Still, there are enough sampling events (2654) that even with subsetting Ohio data, should result in a fairly substantial dataset (~300-500 events) for the region. Additionally, this data may be augmented with new data from the recently released US EPA National Stream Survey. There are 244 NSS sites within states that overlap the Marcellus Shale boundary (MD, NY, OH, PA, VA, WVA). Sites that fall directly within the boundary should be extracted and added to the database. Differences in taxonomic resolution are also a common problem in databases representing a wide range of agencies, methods, and collecting abilities. Overcoming differences in taxonomic resolution will require making tough choices about removing taxa, or sites with taxa not identified to species, or splitting counts of unidentified taxa across abundances of taxa within the same classification (genus). Despite these limitations, the database presented in this report should provide a useful tool for developing flow-ecology models when combined with future hydrologic modeling efforts in phase 2 of this project.

References

Poff, N. L., J. D. Allan, M. B. Bain, J. R. Karr, K. L. Presegaard, B. D. Richter, R. E. Sparks, and J. C. Stromberg. 1997. The natural flow regime. Bioscience 47:769-784.

Poff, N. L., B. D. Richter, A. H. Arthington, S. E. Bunn, R. J. Naiman, E. Kendy, M. Acreman, C. Apse, B. P. Bledsoe, M. C. Freeman, J. Henriksen, R. B. Jacobson, J. G. Kennen, D. M. Merritt, J. H. O'Keeffe, J. D. Olden, K. Rogers, R. E. Tharme, and A. Warner. 2010. The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards. Freshwater Biology 55:147-170.

Rahm, B. G. and S. J. Riha. 2012. Toward strategic management of shale gas development: Regional, collective impacts on water resources. Environmental Science and Policy 17:12-23.