USGS
science for a changing world

Science Analytics and Synthesis Program

Prepared in cooperation with Department of Fisheries and Wildlife, Michigan State University

Developing Fluvial Fish Species Distribution Models Across the Conterminous United States—A Scientific Framework to Support Management and Conservation

Scientific Investigations Report 2023–5088

U.S. Department of the Interior
U.S. Geological Survey

Developing Fluvial Fish Species Distribution Models Across the Conterminous United States—A Scientific Framework to Support Management and Conservation

By Hao Yu, Arthur R. Cooper, Jared Ross, Alexa McKerrow, Daniel J. Wieferich, and Dana M. Infante

Science Analytics and Synthesis Program

Prepared in cooperation with Department of Fisheries and Wildlife, Michigan State University

Scientific Investigations Report 2023–5088

U.S. Department of the Interior
U.S. Geological Survey

U.S. Geological Survey, Reston, Virginia: 2023

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit https://www.usgs.gov or call 1–888–392–8545.

For an overview of USGS information products, including maps, imagery, and publications, visit https://store.usgs.gov/ or contact the store at 1–888–275–8747.

Suggested citation:
Yu, H., Cooper, A.R., Ross, J., McKerrow, A., Wieferich, D.J., and Infante, D.M., 2023, Developing fluvial fish species distribution models across the conterminous United States—A framework for management and conservation: U.S. Geological Survey Scientific Investigations Report 2023–5088, 41 p., https://doi.org/10.3133/sir20235088.

Associated data for this publication:
Wieferich, D.J., McKerrow, A., Cooper, A.R., Yu, H., Ross, J., and Infante, D.M., 2022, Aquatic Gap Analysis Project (Aquatic GAP) aquatic species distribution modeling on the National Hydrography Dataset Plus version 2.1: U.S. Geological Survey data release, https://doi.org/10.5066/P94XM9XV.

ISSN 2328-0328 (online)

Acknowledgments

We acknowledge the U.S. Geological Survey (USGS) Aquatic Gap Analysis Project for funding most of this effort (agreement numbers G17AC00185 and G21AC00013). We also acknowledge support from the Michigan Department of Natural Resources and from the U.S. Department of Agriculture National Institute of Food and Agriculture through Michigan State University AgBioResearch.

Fish data compiled specifically for this effort came from the Connecticut Department of Energy and Environmental Protection; Delaware Department of Natural Resources and Environmental Control; Florida Fish and Wildlife Conservation Commission; Idaho Department of Fish and Game; Illinois Department of Natural Resources; Indiana Department of Environmental Management; Iowa Department of Natural Resources; Kentucky Department of Fish and Wildlife Resources; Maine Department of Inland Fisheries and Wildlife; Maryland Department of Natural Resources; Massachusetts Department of Fisheries and Wildlife; Michigan Department of Natural Resources; Montana Department of Fish, Wildlife and Parks; Multistate Aquatic Resources Information System; New Hampshire Fish and Game; New Jersey Division of Fish and Wildlife; North Carolina Inland Fisheries Division; Oklahoma Conservation Commission; South Dakota Game, Fish and Parks; Tennessee Wildlife Resources Agency; Texas Parks and Wildlife; USGS BioData; USGS Lower Mississippi-Gulf Water Science Center; Virginia Department of Game and Inland Fisheries; and Washington State Department of Ecology. Additional data and approaches for managing data for this effort were supported by the U.S. Fish and Wildlife Service with funding for the 2015 National Assessment of Stream Fish Habitats. A list of fish data providers who supported that effort can be found in Crawford and others, 2016 (table 2 therein; “Stream fish data providers for 2015 national assessment of stream fish habitats”).

Others who have made important contributions to this project include Yin Phan Tsang (University of Hawaii), John Young (USGS Eastern Ecological Science Center), Elizabeth Sellers (data manager, USGS Science Analytics and Synthesis Program), and Wes Daniel and Matthew Neilson (USGS Nonindigenous Aquatic Species Program). Additionally, a team of individuals helped establish the need for national-scale efforts to model aquatic species distributions including Andrea Ostroff, Emmanuel Frimpong, William A. Gould, Robert Hughes, Andrew Loftus, and James E. McKenna. We also wish to thank Kyle Herreman for assistance in managing data used for this effort.

Contents

Figures

Tables

Conversion Factors

International System of Units to U.S. customary units

Multiply	By	To obtain
	Length	
millimeter (mm)	0.03937	inch (in.)
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
meter (m)	1.094	yard (yd)
	Area	
square kilometer (km^2)	247.1	acre
square kilometer (km^2)	0.3861	square mile (mi^2)
	Volume	
cubic meter (m^3)	0.0002642	million gallons (Mgal)
	Mass	
kilogram (kg)	2.205	pound avoirdupois (lb)

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

°F = (1.8 × °C) + 32.

Datum

Vertical coordinate information is referenced to the North American Vertical Datum of 1988 (NAVD 88).

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

Abbreviations

AUC	area under the receiver operating characteristic curve
BRT	boosted regression trees
GAP	Gap Analysis Project
GBIF	Global Biodiversity Information Facility
HUC	hydrologic unit code
NFHP	National Fish Habitat Partnership
NLCD	National Land Cover Database
SDM	species distribution model
TSS	True Skill Statistic
USGS	U.S. Geological Survey

Developing Fluvial Fish Species Distribution Models Across the Conterminous United States—A Scientific Framework to Support Management and Conservation

By Hao Yu,[1] Arthur R. Cooper,[1] Jared Ross,[1] Alexa McKerrow,[2] Daniel J. Wieferich,[2] and Dana M. Infante[1]

Abstract

This report explains the steps and specific methods used to predict fluvial fish occurrences in their native ranges for the conterminous United States. In this study, boosted regression tree models predict distributions of 271 ecologically important fluvial fish species using relations between fish presence/absence and 22 natural and anthropogenic landscape variables. Models developed for the freshwater portions of the ranges for species represented 28 families. *Cyprinidae* was the family with the most species (87 of 271) modeled for this study, followed by *Percidae* (34) and *Ictaluridae* (17). Model predictive performance was evaluated using four metrics: area under the receiver operating characteristic curve, sensitivity, specificity, and True Skill Statistic, which are all from tenfold cross-validation results. The relative importance of the predictor variables in the boosted regression tree models was calculated and ranked for each species. The three strongest natural predictors of fish distributions were network catchment area, the mean annual air temperature of the local catchment, and the maximum elevation of the local catchment, while the three strongest anthropogenic predictors were downstream main stem dam density, distance to downstream main stem dam, and the percentage of pasture/hay land use area within network catchment boundaries. Study results showed 61 fish species were sensitive to climate variables, and 40 fish species were sensitive to anthropogenic stressors. The models developed in this study can be used to derive critical information regarding habitat protection priorities, anthropogenic threats, and potential effects of climate change on habitat suitability, aiding in efforts to conserve fluvial fishes now and into the future.

[1]Department of Fisheries and Wildlife, Michigan State University.

[2]U.S. Geological Survey.

Introduction

An overarching mission of the U.S. Geological Survey (USGS) Aquatic Gap Analysis Project (GAP) is to support national and regional assessments of the conservation status of vertebrate species and plant communities by providing information on the most common and abundant aquatic species found in the United States, while also advancing knowledge on distributions and habitat suitability of rarer, poorly characterized aquatic species. To meet these needs, Aquatic GAP uses spatial analyses and species distribution models (SDMs) to assess aquatic biodiversity and habitats to identify gaps in species protection or threats to habitats. Products of these analyses contribute to conservation planning and prioritization efforts throughout the United States. However, data characterizing habitat suitability and key landscape factors limiting species distributions are lacking for many fluvial fishes in the United States. Development of fluvial fish SDMs provides an opportunity to fill these knowledge gaps.

SDMs are widely used as a management tool to analyze freshwater species distributions and quantify habitat suitability (Bouska and others, 2015). Regression-based approaches are commonly used in SDM development (Guisan and Zimmermann, 2000). Through a logit link function, species presence/absence data are used as the response variable, whereas landscape data can be used as predictor variables for habitat characteristics. This is based on the established understanding that landscape factors of stream catchments can affect fishes through effects on habitats (Allan, 2004). However, regression-based models have several limitations, such as sensitivity to multicollinearity, influence of outliers, and difficulties representing interactions among predictor variables (Elith and others, 2008). Nonparametric machine learning models can overcome limitations inherent in regression-based models (Elith and others, 2008). Machine learning models can improve model performance automatically by experience; this occurs through building a model based on part of the sample data (training set) and using the remaining data points (testing set) to tune the model. This process is done iteratively to improve model predictions and maximize the proportion of model deviation that is explained (Hastie and others, 2009).

Among machine learning approaches, boosted regression trees (BRT) have been recognized as a powerful and robust method for SDM development (Elith and others, 2008).

An important aspect of SDM development is the evaluation of candidate model performance and final selection of the model with the best predictive ability. Reporting this predictive ability assures users of the validity of SDMs and their corresponding use in conservation planning and biodiversity assessments. Many evaluation metrics can be derived from a confusion matrix (Liu and others, 2011), which is a simple table that records numbers of correctly and incorrectly predicted presences and absences. Sensitivity and specificity, two metrics commonly derived from a confusion matrix, indicate the proportions of correctly predicted presences and correctly predicted absences.

Most evaluation metrics are calculated by identifying a threshold value associated with probabilities of SDM predictions. For example, 0.5 is often used as a threshold value for sampling data with similar numbers of presences and absences (Liu and others, 2011). However, the number of presences is often much smaller than the number of absences in an aquatic species survey, and 0.5 may not be suitable in these situations. Besides threshold-dependent evaluation metrics, threshold-independent metrics (for example, area under the receiver operating characteristic curve [AUC]) are frequently used in model evaluation (Liu and others, 2011). Due to inherent differences among evaluation metrics and associated strengths and weaknesses in measuring accuracy, no single metric provides a comprehensive measure of model predictive ability. Therefore, combining multiple evaluation metrics is crucial to appropriately assess model performance.

In addition to predicted habitat suitability, another important outcome of SDM development is the ability to characterize potential species responses to environmental factors that may be important drivers of species distributions. In the context of SDMs, this information can be derived from predictor variable contributions and evaluation of partial dependence plots. In SDMs, the contribution of each predictor variable to response variable prediction (species presence or absence) provides information on the major natural and anthropogenic factors influencing species distributions. Predictor variable contributions often vary among species; thus, SDMs can reveal critical patterns of predictor variable relative importance across multiple species. For instance, including both climate and anthropogenic predictor contributions in each model can help identify climate-sensitive species and species sensitive to anthropogenic stressors. Partial dependence plots investigate the influence of each individual predictor independently by holding all other predictors to their mean values, and they can be used to visualize complex, nonlinear species responses to predictors. Collectively, this information can assist managers in prioritizing conservation policies and management of habitats such as forest cover, dam density, and water withdrawals, based on the relative importance and fish responses to these landscape variables.

This report describes the development of SDMs for 271 fluvial fish species across the conterminous United States. Descriptions of SDM development include the following: (1) an overview for developing SDMs for the Aquatic GAP, (2) results from five diagnostic metrics evaluating overall model performance, (3) important model predictor variables that provide insights into the natural and anthropogenic factors limiting fluvial fish species distributions, (4) species presence/absence predictions for all stream reaches within their native ranges, and (5) habitat suitability assessment that offers valuable information for natural resources management.

Materials and Methods

This section includes detailed descriptions of response variables, predictor variables, statistical models, and model evaluation metrics. The following section titled “Spatial Framework and Landscape Predictors” describes the variables that were used to predict distributions of species.

Spatial Framework and Landscape Predictors

The 1:100,000 scale National Hydrography Dataset Plus Version 2.1, or NHDPlusV2.1, was used as the spatial framework for this project (McKay and others, 2012). This dataset includes ~2.3 million stream reaches in the conterminous United States. In this framework local catchments are defined as the land area draining directly to a given stream reach, while network catchments are defined as the entire upstream drainage area above a stream reach including a stream reach’s own local catchment. Similarly, local buffers include riparian land area within the local catchment that is 90 meters (m) on either side of stream reach, while network buffers include the 180-m riparian land area in the entire upstream network, including a stream reach’s own local buffer. Nine natural and 13 anthropogenic landscape factors were attributed to the spatial framework and used as predictor variables in species distribution modeling (table 1). These predictor variables have also been used in earlier Aquatic GAP fluvial fish distribution model development (Cooper and others, 2019; Yu and others, 2020) and were summarized within five spatial units, including the stream reach, catchments, or buffers (fig. 1).

Nine natural landscape variables were used as predictors in modeling. These included five at the network catchment scale, including catchment area, mean annual precipitation, percentage of overall wetland (combining forested and emergent wetlands) and open-water land-cover types, and base-flow index (percentage contribution of base flow to overall streamflow). The remaining four natural landscape predictor variables included mean annual air temperature and maximum elevation in local catchments, amount of forest land cover within network buffers, and stream reach slope (gradient).

Table 1. Predictor variables used in species distribution model development.

[km^2, square kilometer; EPA, U.S. Environmental Protection Agency; USGS, U.S. Geological Survey; %, percent; km, kilometer; mm, millimeter; OCS, Oregon Climate Service; °C, degrees Celsius; m/m, meter per meter; cm, centimeter; MRLC, Multi-Resolution Land Cover Characteristics Consortium; m, meter; NA, not applicable; no./km, number per square kilometer; PCS, permit compliance system; ICIS, Integrated Compliance Information System; SEMS, Superfund Enterprise Management System; NPDES, National Pollutant Discharge Elimination System; TRIS, toxic release inventory system; kg/km^2; kilogram per square kilometer; SPARROW, SPAtially Referenced Regression On Watershed attributes; HUC8, 8-digit hydrologic unit code; HUC12, 12-digit hydrologic unit code; TIGER, Topologically Integrated Geographic Encoding and Referencing]

Variable and description (units)	Source	Dataset	Scale or resolution
Predictor variable type: Natural			
N_areasqkm: network catchment area (km^2)	Ross and others, 2022	National Hydrography Dataset Plus version 2	1:100,000
N_bfi: network catchment base-flow index (% of base flow contribution to total flow)	Ross and others, 2022	Base-Flow Index Grid for the Conterminous United States (2003)	1 km
N_precip: network catchment mean annual precipitation (mm)	Ross and others, 2022	OCS PRISM 1990–2010	4 km
L_temp: local catchment mean annual air temperature (°C)	Ross and others, 2022	OCS PRISM 1990–2010	4 km
L_fl_slope: stream reach gradient (m/m)	Ross and others, 2022	National Hydrography Dataset Plus version 2	1:100,000
L_maxelev: local catchment maximum elevation (cm)	Ross and others, 2022	National Hydrography Dataset Plus version 2	30 m
NB_nlcd11_41_43: network buffer forest land cover (%)	Ross and others, 2022	2011 National Land Cover Database	30 m
N_nlcd11_11c: network catchment water land cover (%)	Ross and others, 2022	2011 National Land Cover Database	30 m
N_nlcd11_90_95: network catchment woody and emergent herbaceous wetland land cover (%)	Ross and others, 2022	2011 National Land Cover Database	30 m
Predictor variable type: Anthropogenic			
N_nlcd11_21_24: network catchment urban land use; developed open, low, medium, and high intensity (%)	Ross and others, 2022	2011 National Land Cover Database	30 m
N_nlcd81: network catchment cultivated crops (%)	Ross and others, 2022	2011 National Land Cover Database	30 m
N_nlcd82: network catchment pasture/hay (%)	Ross and others, 2022	2011 National Land Cover Database	30 m
N_pop11den: network catchment human population density (no./km^2)	Ross and others, 2022	U.S. Census 2010	1:100,000
N_allepa_den: network catchment density of EPA point-source pollution sites (PCS, ICIS, SEMS, NPDES, and TRIS sites) (no./km^2)	Ross and others, 2022	EPA Facility Registry Service	NA
N_allmine_den: network catchment mineral, coal, and uranium mine density (no./km^2)	Ross and others 2022	Locations of mines and mining activity in the United States	NA
N_total_p_yield: network catchment total phosphorus yield (kg/km^2)	Ross and others, 2022	SPARROW	HUC8
N_totww_mgalc: network catchment total water withdrawal (million gallons/year)	Ross and others, 2022	EnviroAtlas	HUC12
UDOR: degree of regulation: estimated annual discharge stored in upstream reservoirs (%)	Cooper and Infante (2022)	Dam fragmentation	1:100,000
UNDR: upstream network dam density (no./100 km)	Cooper and Infante (2022)	Dam fragmentation	1:100,000
DMD: downstream main stem dam density (no./100 km)	Cooper and Infante (2022)	Dam fragmentation	1:100,000
DM2D_fishtail: distance to downstream main stem dam if present; otherwise distance to network outlet if no downstream dam is present (km)	Cooper and Infante (2022)	Dam fragmentation	1:100,000
N_rx_stlen_dens: stream network road crossing density (no./km)	Ross and others, 2022	2006 TIGER Roads SE	1:100,000

The 13 anthropogenic variables used as model predictors included the following: total urban land use (combining open, low, medium, and high urban), row crop land use, pasture land use, human population density, total water withdrawal, total phosphorous yield, mine density, and point source pollution site density in network catchments. Dam influences were represented by downstream main stem dam density, downstream main stem availability, upstream network dam density, and upstream degree of regulation (percentage of predicted annual streamflow volume stored in all upstream reservoirs) (Cooper and others, 2017), while road influences were represented by upstream road-crossing density in network catchments for stream reaches.

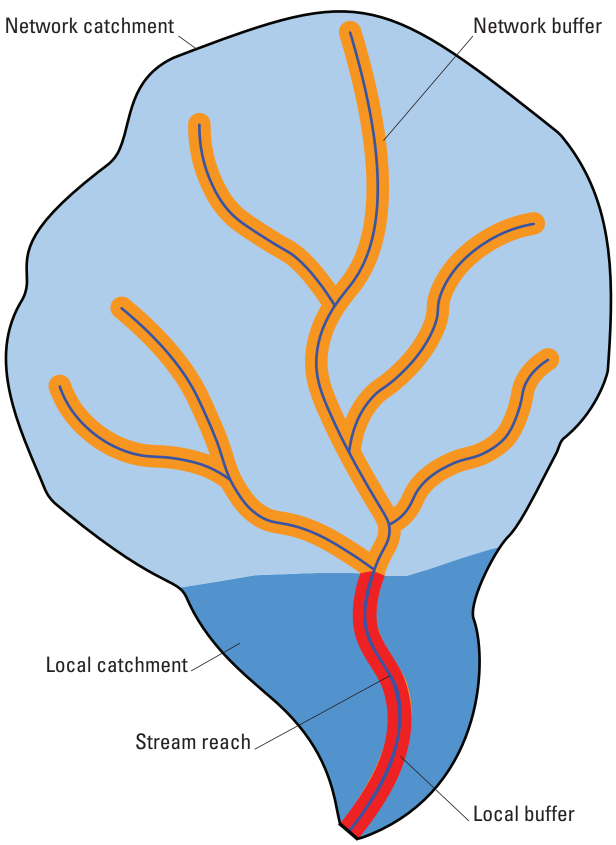

Figure 1. Simplified diagram representing the five spatial units used to summarize landscape variables.

Fish Community Data

The fish data used for presence/absence species distribution modeling were derived from an existing fish database developed for a national fish habitat assessment in support of the National Fish Habitat Partnership (NFHP; http://assessment.fishhabitat.org/) and additional fish data collected in support of this project. Goals for collecting additional fish data were threefold. First, because NFHP fish data span the period 1990–2013, more recent data from 2014 to 2019 were needed to evaluate current conditions. Second, few western States were represented with fish data in the NFHP fish database, thus collection of fish data from data-poor regions was given a high priority to fill in spatial gaps in data availability. Third, whereas NFHP analyses required data that characterized the abundances of all species comprising assemblages, methods for creating SDMs used in this study required presence/absence data. Presence/absence data enabled use of datasets that included targeted samples of specific species or that reported species presence/absence compared to relative abundances. In total, 51 State, academic, and nonprofit sources were contacted, with a total of 14 institutions providing new fish data that could be used for this effort in addition to many datasets previously provided for NFHP. The models also included data from a Federal source (USGS BioData, https://apps.usgs.gov/biodata/; March 15. 2019), an existing consortium of State agency fish databases called Multistate Aquatic Resources Information System (USGS, 2013), and publicly available online State databases. All samples were georeferenced to stream reaches in the National Hydrography Dataset Plus Version 2.1, and the Latin species, genus, and family names used here were validated against and referenced by Integrated Taxonomic Information System taxonomic serial numbers for all records (Integrated Taxonomic Information System, 2019).

A tiered site-selection process based on sample species richness and year sampled was used to create the final fish dataset used in presence/absence modeling. This process ensured that a single, most recent, and species-rich sample was selected for stream reaches that had multiple sampling events. First, samples were assigned to one of six periods (1990–94, 1995–1999, 2000–4, 2005–9, 2010–14, and 2015–19). For each stream reach, the sample with the highest species richness within the most recent period was selected. When the most recent period for a given reach had multiple samples with the same species richness, the sample with the most recent sampling date was selected. This process resulted in the final selection of 35,918 fish samples spanning 1990–2019 for the conterminous United States (fig. 2).

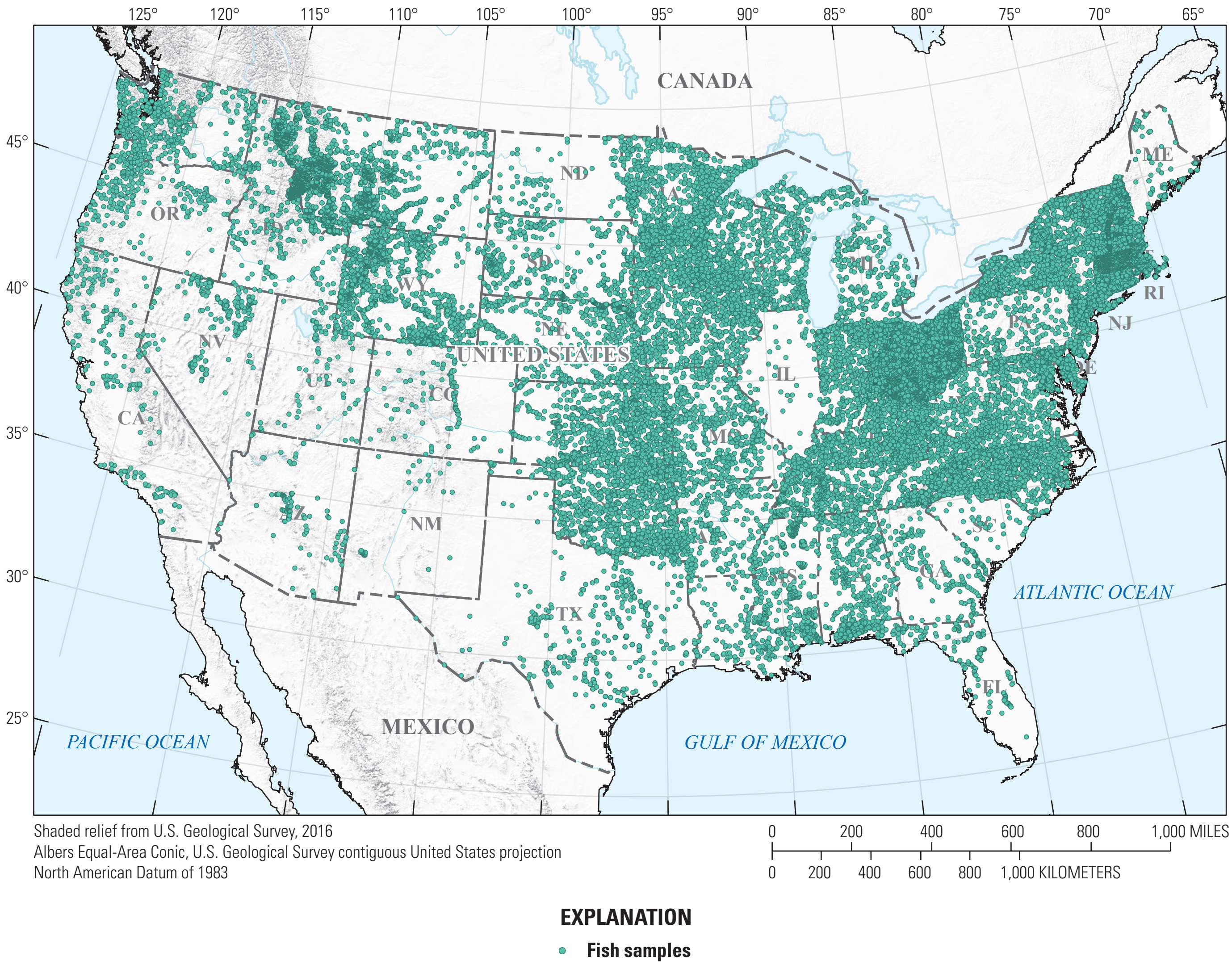

Figure 2. Locations of the 35,918 fish samples for the conterminous United States spanning 1990–2019.

Fish Species Native Ranges

Fish species native range maps were used to constrain model input (fish presence/absence data) and output (projected model presence/absence). The USGS Nonindigenous Aquatic Species Program assisted in acquisition of USGS eight-digit hydrologic unit code (HUC8) -level range maps of 149 species, delineating their range status as native or introduced (Daniel and Neilson, 2020) (fig. 3; table 2). In these cases, use of these range maps ensured that SDMs were built with presence/absence data occurring within a species' native range, excluding presence locations from introduced portions of the range. Species' introduced ranges could represent novel environmental conditions, and therefore, affect model development and potentially limit utility of results intended to support native species conservation. Additionally, range maps were used to limit model projections to stream reaches located within a given species' native range, ensuring that predicted presence/absence locations were not projected to areas where species are not known to be native.

For an additional 122 species lacking detailed native and introduced range maps, HUC8 range maps were developed using all known occurrences (noted as Michigan State University, or MSU, in table 2). These range maps were derived from four data sources: point occurrences from the Aquatic GAP fish database (previously described), point occurrences from the IchthyMaps dataset (Frimpong and others, 2015), point occurrences from Global Biodiversity Information Facility (2020), and HUC8 level range maps developed by NatureServe (NatureServe, 2020) (fig. 4). For Global Biodiversity Information Facility, or GBIF, data, the following data filters were applied to ensure accuracy of both species identification and observation locations: (1) observations were limited to the United States only, (2) observation coordinate uncertainty was less than or equal to 1,000 meters, and (3) observations were made by collectors from Federal, State, or academic institutions (observations based on citizen science were excluded). While these range maps do not include native compared to introduced range status, they do provide geographic boundaries from which to constrain model input/output data and are based on a large set of known occurrences and ranges.

Species Distribution Modeling with Boosted Regression Trees

Previous analyses by the USGS Aquatic GAP tested multiple species distribution modeling techniques for fluvial fishes, including logistic regression, BRT, classification and regression trees, and MaxEnt (A. Ostroff, U.S. Geological Survey, written commun., 2013). Based on results of these analyses and feedback from Aquatic GAP steering committee members, the BRT approach was selected for Aquatic GAP species distribution modeling efforts. BRT differs significantly from regression-based approaches by adaptively combining simple tree models using a boosting technique to improve predictive ability (Elith and others, 2008). Boosting is a sequentially stagewise procedure to link simple trees by emphasizing observations underrepresented in simpler models. Like other machine learning models, regularization is required for BRT to avoid overfitting in the training dataset. Three regularization parameters are commonly used in BRT: learning rate, tree complexity, and bag fraction. Learning rate is used to shrink the contribution of each individual tree in BRT. Tree complexity, ranging from 1 to 5, determines the number of nodes in each tree in the model. If the tree complexity equals 1, interaction effects are not analyzed in the BRT model. If the tree complexity equals 2, BRT models are fit with up to two-way interactions and so on (Elith and others, 2008). Finally, bag fraction is defined as the proportion of training data that are selected in each iteration, which introduces randomness into boosting. A preliminary study evaluated different value combinations of learning rate, tree complexity, and bag fraction (Cooper and others, 2019). Based on results of Cooper and others (2019), an initial learning rate of 0.05 for species with many occurrences (greater than 100) and a learning rate of 0.01 for species with few occurrences (less than or equal to 100) was used in this study. A tree complexity of 5 and bag fraction of 0.75 were used in each model. To ensure a minimum of 1,000 trees in the final model, the learning rate was divided by 2 in each iteration with the maximum number of trees capped at 10,000 to avoid overfitting. All the models were developed using the dismo package (Species Distribution Modeling Version 1.3-3R; Hijmans and others, 2020).

BRT models were evaluated using a tenfold cross-validation procedure in which the entire dataset was split into 10 nonoverlapping subsets and the BRT model was run 10 times. Each time, one of the 10 subsets was used as a test set while the remaining formed a training set for model fitting. The predicted values of all 10 test sets were then used to calculate diagnostic metrics for evaluating the BRT models.

Model Evaluation

Five diagnostic metrics were used to evaluate model performance in this study, including four fundamental measures often used in SDM evaluation: proportion of deviance explained (Elith and others, 2008), sensitivity, specificity, AUC, and True Skill Statistic (TSS) (Allouche and others, 2006). AUC is a threshold-independent metric that avoids the subjective selection of presence/absence cutoff values to develop a confusion matrix for model evaluation. AUC values range between 0 and 1, with larger values indicating better predictive ability. An AUC of 0.5 means that the prediction capability of the model is no better than random, and values greater than 0.7 are considered adequate for modeling species distributions (Swets, 1988). TSS is equal to the sum of sensitivity and specificity minus 1 (Fielding and Bell, 1997). In this study, predicted presences and absences for each fish species were separated by a threshold value that equals the observed prevalence of each sample species, where prevalence represents the proportion of sites in which the species was recorded present.

Predictor Relative Importance

The relative importance (or percent contribution) of each predictor variable was calculated for each species as follows:

$$RI_i = 100\% \times \frac{1}{M}\sum_{m=1}^{M} I_i^2(T_m)$$

where

RI_i stands for the relative importance of the i^{th} predictor variable,

M is the number of trees, and

$I_i^2(T_m)$ is the squared improvement of each predictor weighted by the number of times it was chosen as the splitting variable in tree m (Hastie and others, 2009).

The relative importance of each predictor variable was scaled so that the sum was equal to 100 percent (Elith and others, 2008). Relative importance of all predictor variables in the BRT model was calculated for each species, providing insights into the major natural and anthropogenic factors controlling species distributions.

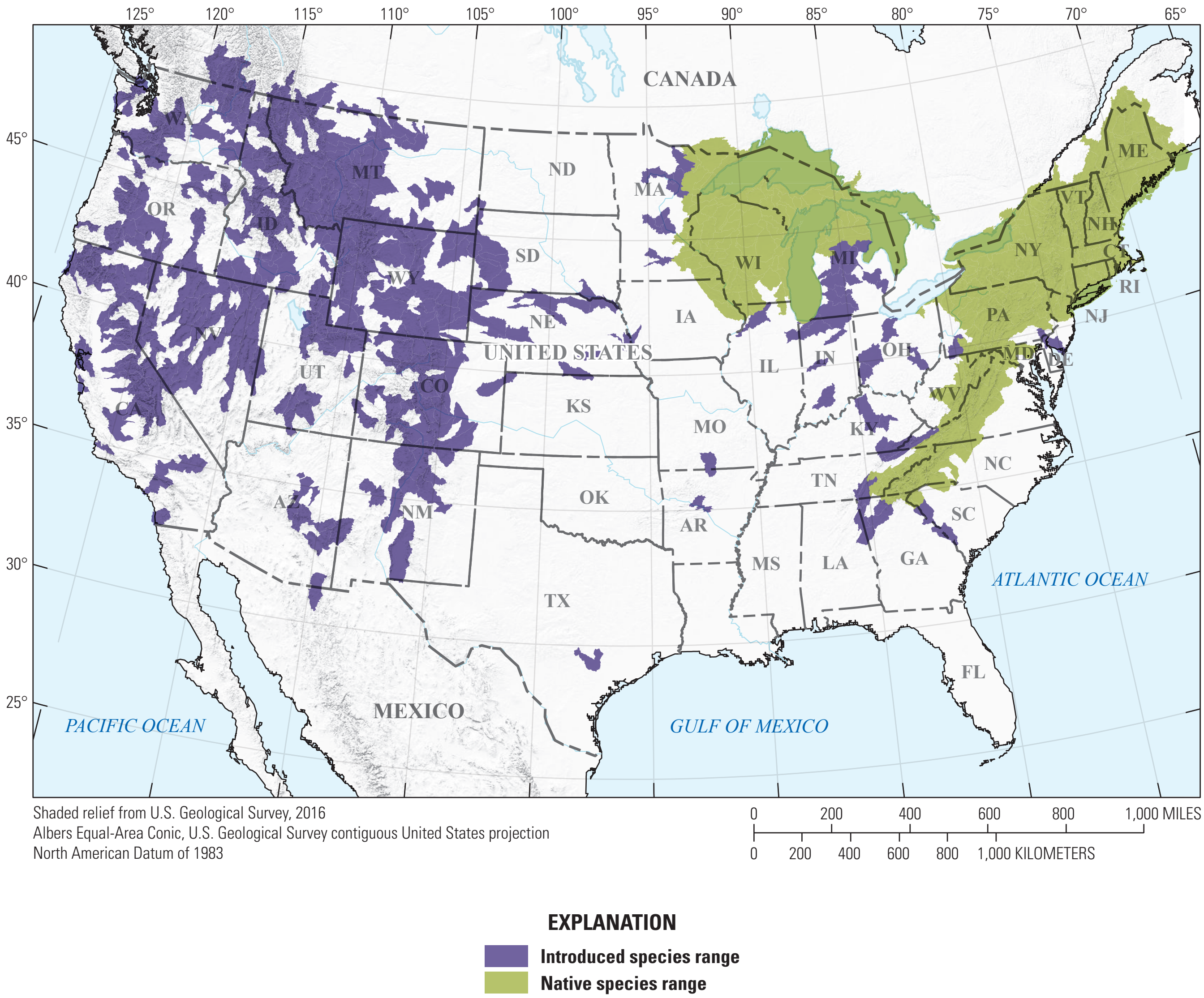

Figure 3. Example U.S. Geological Survey Nonindigenous Aquatic Species range map depicting native compared to introduced eight-digit hydrologic unit code (HUC8) origin status for *Salvelinus fontinalis* (Mitchill, 1814) (brook trout).

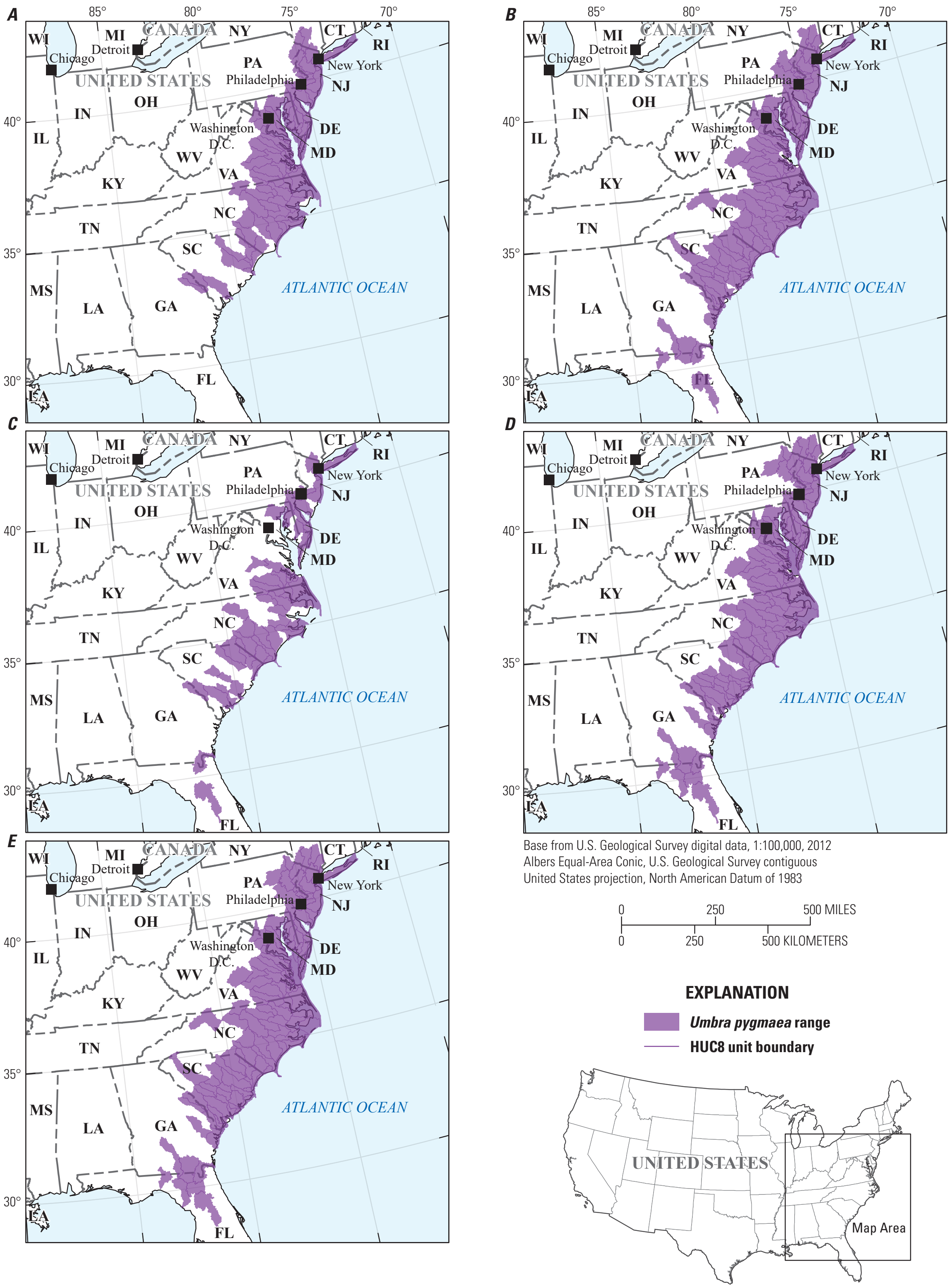

Figure 4. Example range map development for *Umbra pygmaea* (DeKay, 1842) (eastern mudminnow) using all known occurrences from *A*, the Aquatic Gap Analysis Project fish database, *B*, IchthyMaps, *C*, Global Biodiversity Information Facility, and *D*, NatureServe to produce *E*, a final range map used to constrain model input/output for this species.

Table 2. List of fluvial fish species analyzed.

[ITIS TSN, Integrated Taxonomic Information System taxonomic serial number; SGCN, species of greatest conservation need identification; NAS, native range developed by U.S. Geological Survey Nonindigenous Aquatic Species Program; Y, yes for at least one State; N, no for every State; MSU, coarse range developed by Michigan State University]

Scientific name	Common name	ITIS TSN	Presences	Absences	Prevalence	Range map source	Game fish	SGCN
Acantharchus pomotis	Mud sunfish	168095	124	1,509	0.0759	NAS	Y	Y
Acipenser fulvescens	Lake sturgeon	161071	16	7,477	0.0021	NAS	Y	Y
Alosa aestivalis	Blueback herring	161703	24	4,426	0.0054	NAS	Y	Y
Alosa chrysochloris	Skipjack herring	161707	130	5,260	0.0241	NAS	N	Y
Alosa pseudoharengus	Alewife	161706	39	4,942	0.0078	NAS	Y	Y
Alosa sapidissima	American shad	161702	49	7,256	0.0067	NAS	Y	Y
Ambloplites ariommus	Shadow bass	168099	252	1,236	0.1694	NAS	Y	N
Ambloplites cavifrons	Roanoke bass	168098	17	370	0.0439	NAS	Y	Y
Ambloplites constellatus	Ozark bass	168100	73	59	0.553	NAS	Y	N
Ambloplites rupestris	Rock bass	168097	4,853	9,237	0.3444	NAS	Y	Y
Ameiurus brunneus	Snail bullhead	164035	124	860	0.126	NAS	N	Y
Ameiurus catus	White catfish	164037	92	6,486	0.014	NAS	Y	Y
Ameiurus melas	Black bullhead	164039	2,310	17,055	0.1193	NAS	Y	Y
Ameiurus natalis	Yellow bullhead	164041	6,280	15,963	0.2823	NAS	Y	Y
Ameiurus nebulosus	Brown bullhead	164043	1,371	19,645	0.0652	NAS	Y	Y
Ameiurus platycephalus	Flat bullhead	164045	278	1,407	0.165	MSU	N	Y
Amia calva	Bowfin	161104	409	7,666	0.0507	NAS	Y	Y
Anguilla rostrata	American eel	161127	2,189	18,916	0.1037	NAS	Y	Y
Apeltes quadracus	Fourspine stickleback	166397	11	3,632	0.003	NAS	N	Y
Aphredoderus sayanus	Pirate perch	164405	1,363	4,571	0.2297	NAS	N	Y
Aplodinotus grunniens	Freshwater drum	169364	1,705	15,193	0.1009	NAS	Y	Y
Atractosteus spatula	Alligator gar	201897	34	1,607	0.0207	NAS	Y	Y
Campostoma anomalum	Central stoneroller	163508	9,931	8,224	0.547	NAS	N	Y
Campostoma oligolepis	Largescale stoneroller	163509	803	3,447	0.1889	MSU	N	Y
Carpiodes carpio	River carpsucker	163919	1,363	11,876	0.103	NAS	N	Y
Carpiodes cyprinus	Quillback	163917	1,341	16,314	0.076	NAS	N	Y
Carpiodes velifer	Highfin carpsucker	163920	224	10,368	0.0211	NAS	N	Y
Catostomus ardens	Utah sucker	163899	39	208	0.1579	NAS	N	Y
Catostomus catostomus	Longnose sucker	163894	558	7,142	0.0725	NAS	N	Y
Catostomus clarkii	Desert sucker	163901	85	63	0.5743	NAS	Y	Y
Catostomus commersonii	White sucker	553273	13,277	13,208	0.5013	NAS	Y	Y
Catostomus discobolus	Bluehead sucker	163902	45	620	0.0677	MSU	N	Y
Catostomus insignis	Sonora sucker	163905	72	73	0.4966	NAS	N	Y
Catostomus latipinnis	Flannelmouth sucker	163906	80	447	0.1518	NAS	N	Y
Catostomus macrocheilus	Largescale sucker	163896	217	2,226	0.0888	MSU	N	N
Catostomus occidentalis	Sacramento sucker	163908	50	69	0.4202	NAS	N	N
Catostomus platyrhynchus	Mountain sucker	163909	417	2,948	0.1239	MSU	N	Y
Catostomus tahoensis	Tahoe sucker	163914	66	249	0.2095	NAS	N	N
Centrarchus macropterus	Flier	168102	198	3,180	0.0586	NAS	Y	Y
Chrosomus eos	Northern redbelly dace	913993	707	5,087	0.122	MSU	N	Y
Chrosomus erythrogaster	Southern redbelly dace	913994	1,549	8,962	0.1474	MSU	N	Y
Chrosomus neogaeus	Finescale dace	913995	211	3,352	0.0592	MSU	N	Y

Table 2. List of fluvial fish species analyzed.—Continued

[ITIS TSN, Integrated Taxonomic Information System taxonomic serial number; SGCN, species of greatest conservation need identification; NAS, native range developed by U.S. Geological Survey Nonindigenous Aquatic Species Program; Y, yes for at least one State; N, no for every State; MSU, coarse range developed by Michigan State University]

Scientific name	Common name	ITIS TSN	Presences	Absences	Prevalence	Range map source	Game fish	SGCN
Chrosomus oreas	Mountain redbelly dace	913996	196	2,389	0.0758	MSU	N	Y
Clinostomus elongatus	Redside dace	163373	468	7,162	0.0613	NAS	N	Y
Clinostomus funduloides	Rosyside dace	163371	800	3,544	0.1842	MSU	N	Y
Cottus aleuticus	Coastrange sculpin	167230	81	395	0.1702	NAS	N	Y
Cottus bairdii	Mottled sculpin	167237	3,206	12,982	0.198	MSU	N	Y
Cottus beldingii	Paiute sculpin	167238	141	1,080	0.1155	NAS	N	Y
Cottus carolinae	Banded sculpin	167239	901	2,658	0.2532	MSU	N	Y
Cottus cognatus	Slimy sculpin	167232	1,067	9,945	0.0969	NAS	N	Y
Cottus confusus	Shorthead sculpin	167240	143	1,977	0.0675	MSU	N	N
Cottus hypselurus	Ozark sculpin	167263	13	129	0.0915	NAS	N	N
Cottus rhotheus	Torrent sculpin	167252	248	799	0.2369	MSU	N	Y
Couesius plumbeus	Lake chub	163535	185	4,548	0.0391	NAS	N	Y
Culaea inconstans	Brook stickleback	166399	1,622	8,675	0.1575	NAS	N	Y
Cycleptus elongatus	Blue sucker	163953	150	6,290	0.0233	NAS	N	Y
Cyprinella analostana	Satinfin shiner	163766	403	3,222	0.1112	MSU	N	N
Cyprinella camura	Bluntface shiner	163776	237	1,156	0.1701	MSU	N	Y
Cyprinella galactura	Whitetail shiner	163782	369	1,849	0.1664	MSU	N	Y
Cyprinella lutrensis	Red shiner	163792	2,819	2,904	0.4926	NAS	N	Y
Cyprinella spiloptera	Spotfin shiner	163803	3,941	12,708	0.2367	MSU	N	Y
Cyprinella venusta	Blacktail shiner	163809	780	1,944	0.2863	MSU	N	Y
Cyprinella whipplei	Steelcolor shiner	163811	452	7,548	0.0565	MSU	N	Y
Dorosoma cepedianum	Gizzard shad	161737	2,561	17,397	0.1283	NAS	Y	N
Dorosoma petenense	Threadfin shad	161738	136	716	0.1596	NAS	N	N
Elassoma zonatum	Banded pygmy sunfish	168171	154	2,610	0.0557	NAS	N	Y
Enneacanthus chaetodon	Blackbanded sunfish	168108	10	1,099	0.009	NAS	N	Y
Enneacanthus gloriosus	Bluespotted sunfish	168113	289	2,439	0.1059	NAS	N	Y
Enneacanthus obesus	Banded sunfish	168117	123	5,074	0.0237	NAS	N	Y
Entosphenus tridentatus	Pacific lamprey	159699	73	992	0.0685	NAS	Y	Y
Erimystax dissimilis	Streamline chub	163821	106	5,948	0.0175	NAS	N	Y
Erimystax x-punctatus	Gravel chub	163824	153	6,865	0.0218	NAS	N	Y
Erimyzon oblongus	Eastern creek chubsucker	163924	1,305	14,304	0.0836	MSU	N	Y
Erimyzon sucetta	Lake chubsucker	163922	132	7,845	0.0165	MSU	N	Y
Esox americanus	Redfin pickerel	162140	2,542	16,807	0.1314	NAS	Y	Y
Esox lucius	Northern pike	162139	1,282	6,473	0.1653	NAS	Y	Y
Esox niger	Chain pickerel	162143	1,126	13,551	0.0767	MSU	Y	Y
Etheostoma blennioides	Greenside darter	168375	4,373	6,875	0.3888	MSU	N	Y
Etheostoma caeruleum	Rainbow darter	168378	4,447	7,050	0.3868	MSU	N	Y
Etheostoma camurum	Bluebreast darter	168379	118	6,502	0.0178	NAS	N	Y
Etheostoma cragini	Arkansas darter	168386	150	715	0.1734	NAS	N	Y
Etheostoma exile	Iowa darter	168393	421	8,128	0.0492	MSU	N	Y
Etheostoma flabellare	Fantail darter	168394	5,415	9,525	0.3624	MSU	N	Y
Etheostoma fusiforme	Swamp darter	168358	118	4,975	0.0232	MSU	N	Y

Table 2. List of fluvial fish species analyzed.—Continued

[ITIS TSN, Integrated Taxonomic Information System taxonomic serial number; SGCN, species of greatest conservation need identification; NAS, native range developed by U.S. Geological Survey Nonindigenous Aquatic Species Program; Y, yes for at least one State; N, no for every State; MSU, coarse range developed by Michigan State University]

Scientific name	Common name	ITIS TSN	Presences	Absences	Prevalence	Range map source	Game fish	SGCN
Etheostoma gracile	Slough darter	168366	223	2,408	0.0848	MSU	N	Y
Etheostoma kennicotti	Stripetail darter	168405	117	896	0.1155	MSU	N	Y
Etheostoma lynceum	Brighteye darter	168456	76	615	0.11	NAS	N	Y
Etheostoma microperca	Least darter	168411	75	8,376	0.0089	NAS	N	Y
Etheostoma nigrum	Johnny darter	168369	7,774	10,643	0.4221	MSU	N	Y
Etheostoma olmstedi	Tessellated darter	168360	2,710	6,743	0.2867	MSU	N	Y
Etheostoma punctulatum	Stippled darter	168425	234	424	0.3556	MSU	N	Y
Etheostoma radiosum	Orangebelly darter	168426	249	357	0.4109	MSU	N	Y
Etheostoma rufilineatum	Redline darter	168428	191	1,091	0.149	MSU	N	Y
Etheostoma simoterum	Snubnose darter	168431	204	1,166	0.1489	MSU	N	N
Etheostoma spectabile	Orangethroat darter	168368	2,957	6,696	0.3063	MSU	N	Y
Etheostoma stigmaeum	Speckled darter	168437	238	2,461	0.0882	MSU	N	Y
Etheostoma swaini	Gulf darter	168439	202	1,166	0.1477	MSU	N	Y
Etheostoma variatum	Variegate darter	168446	254	4,871	0.0496	MSU	N	Y
Etheostoma whipplei	Redfin darter	168448	247	1,712	0.1261	MSU	N	Y
Etheostoma zonale	Banded darter	168449	1,849	7,541	0.1969	NAS	N	Y
Exoglossum maxillingua	Cutlip minnow	163356	852	2,773	0.235	NAS	N	Y
Fundulus catenatus	Northern studfish	165660	562	2,769	0.1687	MSU	N	Y
Fundulus diaphanus	Banded killifish	165646	282	11,938	0.0231	NAS	N	Y
Fundulus kansae	Northern plains killifish	165654	249	1,880	0.117	MSU	N	Y
Fundulus notatus	Blackstripe topminnow	165663	1,572	8,715	0.1528	MSU	N	Y
Fundulus olivaceus	Blackspotted topminnow	165655	1,321	2,545	0.3417	MSU	N	N
Fundulus seminolis	Seminole killifish	165667	34	101	0.2519	NAS	N	N
Fundulus zebrinus	Plains killifish	165658	258	2,234	0.1035	MSU	N	N
Gambusia affinis	Western mosquitofish	165878	3,068	10,477	0.2265	MSU	N	N
Gila robusta	Roundtail chub	163558	55	496	0.0998	NAS	Y	Y
Hesperoleucus symmetricus	California roach	163565	16	83	0.1616	NAS	N	N
Hiodon alosoides	Goldeye	161905	240	8,240	0.0283	NAS	N	Y
Hiodon tergisus	Mooneye	161906	134	8,789	0.015	NAS	N	Y
Hybognathus argyritis	Western silvery minnow	163362	79	2,024	0.0376	NAS	N	Y
Hybognathus hankinsoni	Brassy minnow	163363	921	5,673	0.1397	MSU	N	Y
Hybognathus nuchalis	Mississippi silvery minnow	163360	185	5,416	0.033	MSU	N	Y
Hybognathus placitus	Plains minnow	163361	225	4,337	0.0493	NAS	N	Y
Hybognathus regius	Eastern silvery minnow	163359	119	4,045	0.0286	MSU	N	Y
Hybopsis amblops	Bigeye chub	163476	567	10,796	0.0499	MSU	N	Y
Hybopsis amnis	Pallid shiner	201917	14	2,340	0.0059	NAS	N	Y
Hybopsis dorsalis	Bigmouth shiner	689231	109	814	0.1181	MSU	N	Y
Hybopsis winchelli	Clear chub	201918	1,601	4,040	0.2838	MSU	N	Y
Hypentelium etowanum	Alabama hog sucker	163950	147	792	0.1565	NAS	N	Y
Hypentelium nigricans	Northern hog sucker	163949	112	154	0.4211	NAS	N	Y
Hypentelium roanokense	Roanoke hog sucker	163951	6,064	9,113	0.3996	NAS	N	Y
Ichthyomyzon castaneus	Chestnut lamprey	159725	62	117	0.3464	MSU	N	Y

Table 2. List of fluvial fish species analyzed.—Continued

[ITIS TSN, Integrated Taxonomic Information System taxonomic serial number; SGCN, species of greatest conservation need identification; NAS, native range developed by U.S. Geological Survey Nonindigenous Aquatic Species Program; Y, yes for at least one State; N, no for every State; MSU, coarse range developed by Michigan State University]

Scientific name	Common name	ITIS TSN	Presences	Absences	Prevalence	Range map source	Game fish	SGCN
Ichthyomyzon fossor	Northern brook lamprey	159726	166	4,414	0.0362	NAS	N	Y
Ichthyomyzon gagei	Southern brook lamprey	159727	64	3,181	0.0197	NAS	N	Y
Ichthyomyzon greeleyi	Mountain brook lamprey	159728	169	2,168	0.0723	NAS	N	Y
Ictalurus furcatus	Blue catfish	163997	111	1,449	0.0712	NAS	Y	Y
Ictalurus punctatus	Channel catfish	163998	140	4,453	0.0305	NAS	Y	Y
Ictiobus bubalus	Smallmouth buffalo	163955	3,544	16,874	0.1736	MSU	Y	Y
Ictiobus cyprinellus	Bigmouth buffalo	163956	961	12,240	0.0728	MSU	Y	Y
Ictiobus niger	Black buffalo	163957	515	10,849	0.0453	MSU	N	Y
Labidesthes sicculus	Brook silverside	166016	273	7852	0.0336	NAS	N	Y
Lampetra aepyptera	Least brook lamprey	159705	1,467	13,856	0.0957	MSU	N	Y
Lampetra richardsoni	Western brook lamprey	159707	490	6,097	0.0744	NAS	N	Y
Lepisosteus oculatus	Spotted gar	161095	28	687	0.0392	NAS	N	Y
Lepisosteus osseus	Longnose gar	161094	432	5,417	0.0739	NAS	N	Y
Lepisosteus platostomus	Shortnose gar	161096	1,063	17,549	0.0571	NAS	N	Y
Lepisosteus platyrhincus	Florida gar	161098	359	6,157	0.0551	NAS	N	Y
Lepomis auritus	Redbreast sunfish	168131	76	218	0.2585	NAS	Y	Y
Lepomis cyanellus	Green sunfish	168132	1,970	6,304	0.2381	NAS	Y	N
Lepomis gibbosus	Pumpkinseed	168144	10,914	6,315	0.6335	NAS	Y	Y
Lepomis humilis	Orangespotted sunfish	168151	1,762	14,669	0.1072	NAS	Y	Y
Lepomis macrochirus	Bluegill	168141	1,821	10,191	0.1516	NAS	Y	N
Lepomis marginatus	Dollar sunfish	168152	9,943	8,599	0.5362	NAS	N	Y
Lepomis megalotis	Longear sunfish	168153	310	2,139	0.1266	NAS	Y	Y
Lepomis microlophus	Redear sunfish	168154	5,587	5,063	0.5246	NAS	Y	N
Lepomis miniatus	Redspotted sunfish	168157	589	2,914	0.1681	NAS	N	Y
Lepomis punctatus	Spotted sunfish	168155	306	2,275	0.1186	NAS	Y	Y
Lepomis symmetricus	Bantam sunfish	168156	374	458	0.4495	NAS	N	Y
Lethenteron appendix	American brook lamprey	914061	25	924	0.0263	MSU	N	Y
Lota lota	Burbot	164725	442	8,886	0.0474	NAS	Y	Y
Luxilus albeolus	White shiner	163826	518	8,031	0.0606	MSU	N	N
Luxilus cardinalis	Cardinal shiner	163828	254	825	0.2354	MSU	N	Y
Luxilus cerasinus	Crescent shiner	163830	190	386	0.3299	MSU	N	N
Luxilus chrysocephalus	Striped shiner	163832	150	903	0.1425	MSU	N	Y
Luxilus coccogenis	Warpaint shiner	163834	4,907	7,753	0.3876	NAS	N	Y
Luxilus cornutus	Common shiner	163836	269	474	0.362	NAS	N	Y
Luxilus zonatus	Bleeding shiner	163840	4,482	12,384	0.2657	MSU	N	N
Lythrurus ardens	Rosefin shiner	163847	215	410	0.344	MSU	N	Y
Lythrurus fasciolaris	Scarlet shiner	201928	166	4,968	0.0323	MSU	N	Y
Lythrurus fumeus	Ribbon shiner	163853	1,185	3,627	0.2463	MSU	N	Y
Lythrurus snelsoni	Ouachita shiner	163859	165	1,828	0.0828	NAS	N	Y
Lythrurus umbratilis	Redfin shiner	163861	45	111	0.2885	MSU	N	Y
Macrhybopsis storeriana	Silver chub	163870	1,732	10,098	0.1464	MSU	N	Y
Margariscus margarita	Allegheny Pearl Dace	163873	232	10,436	0.0217	NAS	N	Y

Table 2. List of fluvial fish species analyzed.—Continued

[ITIS TSN, Integrated Taxonomic Information System taxonomic serial number; SGCN, species of greatest conservation need identification; NAS, native range developed by U.S. Geological Survey Nonindigenous Aquatic Species Program; Y, yes for at least one State; N, no for every State; MSU, coarse range developed by Michigan State University]

Scientific name	Common name	ITIS TSN	Presences	Absences	Prevalence	Range map source	Game fish	SGCN
Menidia beryllina	Inland silverside	165993	76	1,882	0.0388	MSU	N	Y
Micropterus cataractae	Shoal bass	564610	203	4,350	0.0446	NAS	Y	Y
Micropterus coosae	Redeye bass	168163	17	64	0.2099	MSU	Y	Y
Micropterus dolomieu	Smallmouth bass	550562	91	833	0.0985	NAS	Y	Y
Micropterus punctulatus	Spotted bass	168161	4,035	10,099	0.2855	NAS	Y	Y
Micropterus salmoides	Largemouth bass	168160	1,908	6,343	0.2312	NAS	Y	Y
Minytrema melanops	Spotted sucker	163959	7,089	12,364	0.3644	MSU	N	Y
Morone americana	White perch	167678	1,167	11,600	0.0914	NAS	Y	N
Morone chrysops	White bass	167682	69	3,663	0.0185	NAS	Y	N
Morone mississippiensis	Yellow bass	167683	399	9,191	0.0416	NAS	Y	Y
Morone saxatilis	Striped bass	167680	28	1,657	0.0166	NAS	Y	Y
Moxostoma anisurum	Silver redhorse	163933	53	4,331	0.0121	MSU	N	Y
Moxostoma breviceps	Smallmouth redhorse	163929	1,157	12,953	0.082	NAS	N	N
Moxostoma carinatum	River redhorse	163936	404	5,257	0.0714	NAS	N	Y
Moxostoma collapsum	Notchlip redhorse	201946	245	8,392	0.0284	MSU	N	Y
Moxostoma congestum	Gray redhorse	163931	154	1,388	0.0999	NAS	N	Y
Moxostoma duquesnii	Black redhorse	553274	35	116	0.2318	MSU	N	Y
Moxostoma erythrurum	Golden redhorse	163939	1,498	9,783	0.1328	MSU	N	Y
Moxostoma macrolepidotum	Shorthead redhorse	163928	3,946	12,611	0.2383	NAS	N	Y
Moxostoma poecilurum	Blacktail redhorse	163932	1,517	9,614	0.1363	MSU	N	Y
Moxostoma rupiscartes	Striped jumprock	163946	252	1,176	0.1765	MSU	N	N
Moxostoma valenciennesi	Greater redhorse	163947	168	915	0.1551	NAS	N	Y
Mugil cephalus	Striped mullet	170335	168	4,099	0.0394	MSU	N	Y
Nocomis biguttatus	Hornyhead chub	163395	164	1,857	0.0811	NAS	N	Y
Nocomis leptocephalus	Bluehead chub	163393	1,410	9,375	0.1307	MSU	N	Y
Nocomis micropogon	River chub	163392	1,023	1,971	0.3417	MSU	N	Y
Notemigonus crysoleucas	Golden shiner	163368	1,027	9,186	0.1006	MSU	N	Y
Notropis amabilis	Texas shiner	163410	2,899	25,193	0.1032	NAS	N	N
Notropis atherinoides	Emerald shiner	163412	37	85	0.3033	MSU	N	Y
Notropis blennius	River shiner	163429	1,652	17,354	0.0869	MSU	N	Y
Notropis boops	Bigeye shiner	163430	154	7,945	0.019	MSU	N	Y
Notropis buccatus	Silverjaw minnow	163478	620	5,157	0.1073	MSU	N	Y
Notropis chiliticus	Redlip shiner	163435	2,774	6,427	0.3015	MSU	N	Y
Notropis cummingsae	Dusky shiner	163438	201	419	0.3242	MSU	N	Y
Notropis girardi	Arkansas River shiner	163442	17	1,036	0.0161	NAS	N	Y
Notropis heterolepis	Blacknose shiner	163446	263	8,626	0.0296	MSU	N	Y
Notropis hudsonius	Spottail shiner	163404	871	13,936	0.0588	MSU	N	Y
Notropis leuciodus	Tennessee shiner	163451	239	1,190	0.1672	MSU	N	Y
Notropis longirostris	Longnose shiner	163452	177	684	0.2056	MSU	N	N
Notropis lutipinnis	Yellowfin shiner	163453	88	599	0.1281	MSU	N	Y
Notropis nubilus	Ozark minnow	163456	378	1,116	0.253	NAS	N	Y
Notropis percobromus	Carmine shiner	689522	374	3,771	0.0902	MSU	N	N

Table 2. List of fluvial fish species analyzed.—Continued

[ITIS TSN, Integrated Taxonomic Information System taxonomic serial number; SGCN, species of greatest conservation need identification; NAS, native range developed by U.S. Geological Survey Nonindigenous Aquatic Species Program; Y, yes for at least one State; N, no for every State; MSU, coarse range developed by Michigan State University]

Scientific name	Common name	ITIS TSN	Presences	Absences	Prevalence	Range map source	Game fish	SGCN
Notropis petersoni	Coastal shiner	163460	125	997	0.1114	MSU	N	N
Notropis photogenis	Silver shiner	163461	1,020	7,449	0.1204	MSU	N	Y
Notropis procne	Swallowtail shiner	163407	351	2,873	0.1089	MSU	N	Y
Notropis rubellus	Rosyface shiner	163409	1,212	13,840	0.0805	MSU	N	Y
Notropis stramineus	Sand shiner	163419	4,843	12,834	0.274	MSU	N	Y
Notropis telescopus	Telescope shiner	163470	294	1,891	0.1346	MSU	N	N
Notropis texanus	Weed shiner	163420	359	2,670	0.1185	NAS	N	Y
Notropis topeka	Topeka shiner	163471	27	2,152	0.0124	NAS	N	Y
Notropis volucellus	Mimic shiner	163421	1,081	16,167	0.0627	MSU	N	Y
Noturus albater	Ozark madtom	164006	47	172	0.2146	NAS	N	N
Noturus exilis	Slender madtom	164010	654	1,913	0.2548	NAS	N	Y
Noturus flavus	Stonecat	164013	1,756	15,529	0.1016	NAS	N	Y
Noturus gyrinus	Tadpole madtom	164003	983	20,262	0.0463	NAS	N	Y
Noturus insignis	Margined madtom	164004	917	2,028	0.3114	NAS	N	Y
Noturus leptacanthus	Speckled madtom	164019	298	997	0.2301	MSU	N	N
Noturus miurus	Brindled madtom	164020	322	9,573	0.0325	NAS	N	Y
Noturus nocturnus	Freckled madtom	164005	337	3,546	0.0868	MSU	N	Y
Oncorhynchus clarkii	Cutthroat trout	161983	2,907	1,945	0.5991	NAS	Y	Y
Oncorhynchus kisutch	Coho salmon	161977	206	503	0.2906	NAS	Y	Y
Oncorhynchus mykiss	Rainbow trout	161989	659	510	0.5637	NAS	Y	Y
Oncorhynchus tshawytscha	Chinook salmon	161980	30	405	0.069	NAS	Y	Y
Opsopoeodus emiliae	Pugnose minnow	163876	192	6,576	0.0284	MSU	N	Y
Perca flavescens	Yellow perch	168469	1,807	18,000	0.0912	NAS	Y	Y
Percina caprodes	Logperch	168472	3,219	14,114	0.1857	MSU	N	Y
Percina evides	Gilt darter	168483	169	3,930	0.0412	NAS	N	Y
Percina maculata	Blackside darter	168488	2,659	11,986	0.1816	NAS	N	Y
Percina nigrofasciata	Blackbanded darter	168490	572	867	0.3975	MSU	N	Y
Percina peltata	Shield darter	168474	191	1,947	0.0893	MSU	N	Y
Percina phoxocephala	Slenderhead darter	168494	685	8,152	0.0775	MSU	N	Y
Percina roanoka	Roanoke darter	168496	179	698	0.2041	MSU	N	N
Percina sciera	Dusky darter	168475	547	5,712	0.0874	MSU	N	Y
Percopsis omiscomaycus	Trout-perch	164409	421	9,802	0.0412	MSU	N	Y
Petromyzon marinus	Sea lamprey	159722	197	5,203	0.0365	NAS	N	Y
Phenacobius mirabilis	Suckermouth minnow	163502	1,960	10,250	0.1605	NAS	N	Y
Pimephales notatus	Bluntnose minnow	163516	10,143	12,332	0.4513	MSU	N	Y
Pimephales promelas	Fathead minnow	163517	4,873	21,247	0.1866	MSU	N	N
Pimephales vigilax	Bullhead minnow	163518	1,334	10,333	0.1143	MSU	N	Y
Platygobio gracilis	Flathead chub	163882	245	2,867	0.0787	NAS	N	Y
Polyodon spathula	Paddlefish	161088	22	7,511	0.0029	NAS	Y	Y
Pomoxis annularis	White crappie	168166	1,332	14,622	0.0835	NAS	Y	N
Pomoxis nigromaculatus	Black crappie	168167	1,362	16,632	0.0757	NAS	Y	Y
Prosopium williamsoni	Mountain whitefish	162009	476	4,154	0.1028	NAS	Y	Y

Table 2. List of fluvial fish species analyzed.—Continued

[ITIS TSN, Integrated Taxonomic Information System taxonomic serial number; SGCN, species of greatest conservation need identification; NAS, native range developed by U.S. Geological Survey Nonindigenous Aquatic Species Program; Y, yes for at least one State; N, no for every State; MSU, coarse range developed by Michigan State University]

Scientific name	Common name	ITIS TSN	Presences	Absences	Prevalence	Range map source	Game fish	SGCN
Ptychocheilus grandis	Sacramento pikeminnow	163524	26	66	0.2826	NAS	N	N
Ptychocheilus oregonensis	Northern pikeminnow	163523	117	2,315	0.0481	NAS	Y	N
Pylodictis olivaris	Flathead catfish	164029	1,238	11,045	0.1008	NAS	Y	Y
Rhinichthys atratulus	Blacknose dace	163382	5,872	15,446	0.2754	MSU	N	Y
Rhinichthys cataractae	Longnose dace	163384	4,136	14,805	0.2184	MSU	N	Y
Rhinichthys obtusus	Western blacknose dace	689949	3,146	9,545	0.2479	MSU	N	Y
Rhinichthys osculus	Speckled dace	163387	654	1,627	0.2867	MSU	N	Y
Richardsonius balteatus	Redside shiner	163528	310	3,133	0.09	MSU	N	Y
Salmo salar	Atlantic salmon	161996	223	3,753	0.0561	NAS	N	Y
Salvelinus confluentus	Bull trout	162004	511	1,887	0.2131	NAS	Y	Y
Salvelinus fontinalis	Brook trout	162003	3,019	7,630	0.2835	NAS	Y	Y
Sander canadensis	Sauger	650171	451	9,942	0.0434	NAS	Y	Y
Sander vitreus	Walleye	650173	795	13,325	0.0563	NAS	Y	Y
Scaphirhynchus platorynchus	Shovelnose sturgeon	161082	50	4,286	0.0115	NAS	Y	Y
Semotilus atromaculatus	Creek chub	163376	13,586	13,198	0.5072	MSU	N	Y
Semotilus corporalis	Fallfish	163375	1,512	6,137	0.1977	NAS	N	Y
Thoburnia rhothoeca	Torrent sucker	553276	86	358	0.1937	MSU	N	Y
Umbra limi	Central mudminnow	162153	2,088	7,187	0.2251	NAS	N	Y
Umbra pygmaea	Eastern mudminnow	162148	412	1,639	0.2009	MSU	N	Y

Results

We modeled the distributions of 271 species out of a set of 298 total fluvial fish species (table 2). For 27 species, lack of occurrences resulted in either the inability to attempt model development due to low number of occurrences (less than 10) or an inability of the BRT approach to create a stable model due to lack of model convergence (see table 1.1 in app. 1). For modeled species, the range in number of presences, number of absences, and prevalence was large (figs. 5 and 6). In total, 263 species were considered to have low to moderate prevalence (less than 0.5), while 10 species had high prevalence (greater than 0.5). Species prevalence ranged from 0.0021 (*Acipenser fulvescens*, lake sturgeon) to 0.6335 (*Lepomis gibbosus*, pumpkinseed), with a mean of all the prevalence values plus or minus (±) standard error of 0.1566 ± 0.0082. The proportion of deviance explained by the BRT model also varied considerably across fish species (table 3; fig. 7), ranging from 0.0562 (*Moxostoma congestum*, gray redhorse) to 0.7198 (*Micropterus cataractae,* shoal bass) with a mean of 0.3442 ± 0.0065. The model predictive performance evaluation metrics calculated from tenfold cross validation varied across models (table 3; fig. 7). In total, 270 of 271 models were considered acceptable based on AUC values (greater than or equal to 0.7).

The contributions of the predictor variables varied across species; however, network catchment area was consistently the most influential predictor, and overall, natural variables tended to have the greatest influence across species (fig. 8). The top three natural predictors, listed here with their variable names in order of mean relative importance, were network catchment area (N_areasqkm, 14.88 percent), local catchment mean annual air temperature (L_temp, 9.12 percent), and local catchment maximum elevation (L_maxelev, 6.70 percent). The top three anthropogenic predictors based on mean relative importance were downstream main stem dam density (DMD, 5.59 percent), distance to downstream main stem dam (DM2D_fishtail, 4.85 percent), and network catchment pasture and hay (N_nlcd82, 4.03 percent). There were seven predictors whose mean relative importance was less than 3 percent. These included upstream network dam density, network catchment mine density, degree of regulation, network catchment point source pollution density, network catchment human population density, network catchment urban land use, and stream network road crossing density. While average relative importance for these variables was low, influence of these variables was occasionally very high (greater than 20 percent) for certain species. These predictors could be reassessed for future modeling efforts, potentially dropping these predictors for some species or regions. Among climate-based predictors, mean annual air temperature played an important role in BRT models (relative importance greater than 10 percent) for 87 of 271 species modeled, while precipitation was less important with 37 of 271 species modeled meeting this cutoff. The influences of these climate variables pointed to 61 climate-sensitive fish species (having a sum relative importance of mean annual air temperature and precipitation greater than 20 percent) (table 4). Further, 40 fish species were responsive to anthropogenic stressors (having a sum relative importance of all anthropogenic variables greater than 50 percent) (table 5). The summaries in this section are based on 2 climate predictor variables and 13 anthropogenic predictor variables used to develop SDMs for each species.

To provide examples of SDM output, three fish species with differing prevalence characteristics were selected as example species (table 6; figs. 9–11). The AUC plots of these species showed that the AUC scores were related to the number of presences (fig. 12). Partial dependence plots of predictor variables for these species showed the relative importance of the top 12 predictor variables (fig. 13). The predictor variables showed different levels of importance to species' distributions. For instance, network catchment area was the most important predictor variable for *Semotilus atromaculatus* (Mitchill, 1818) (creek chub), the sixth most important for *Cottus beldingii* (Eigenmann and Eigenmann, 1891) (Paiute sculpin), and not in the top 12 for *Enneacanthus chaetodon* (Baird, 1955) (blackbanded sunfish).

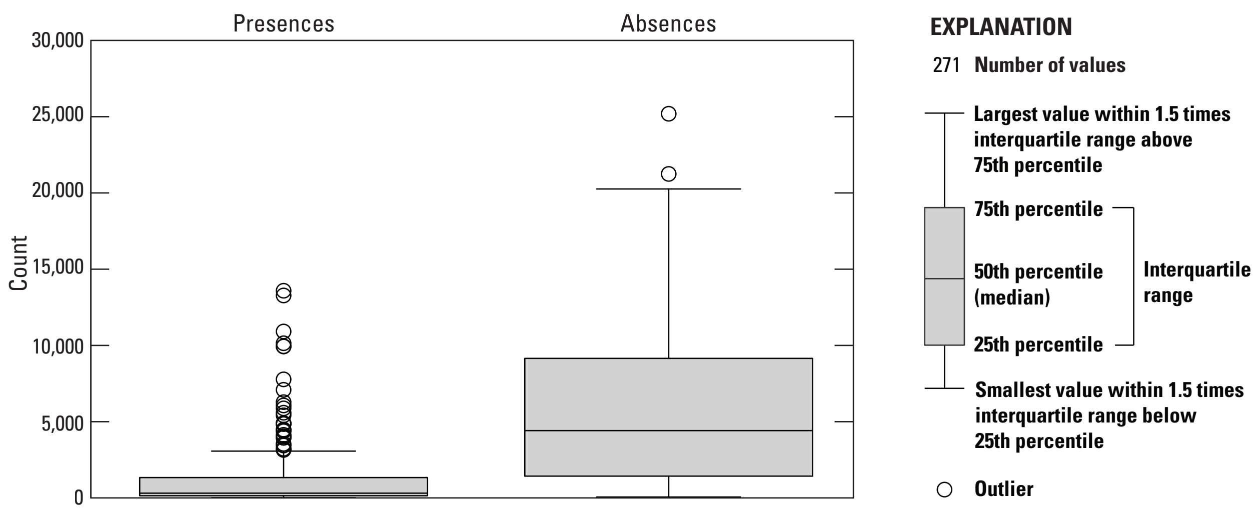

Figure 5. Boxplots of presences and absences for the 271 fish species modeled. The left boxplot represents the distribution of presences of the 271 fish species, and the right boxplot represents the distribution of absences of the 271 fish species.

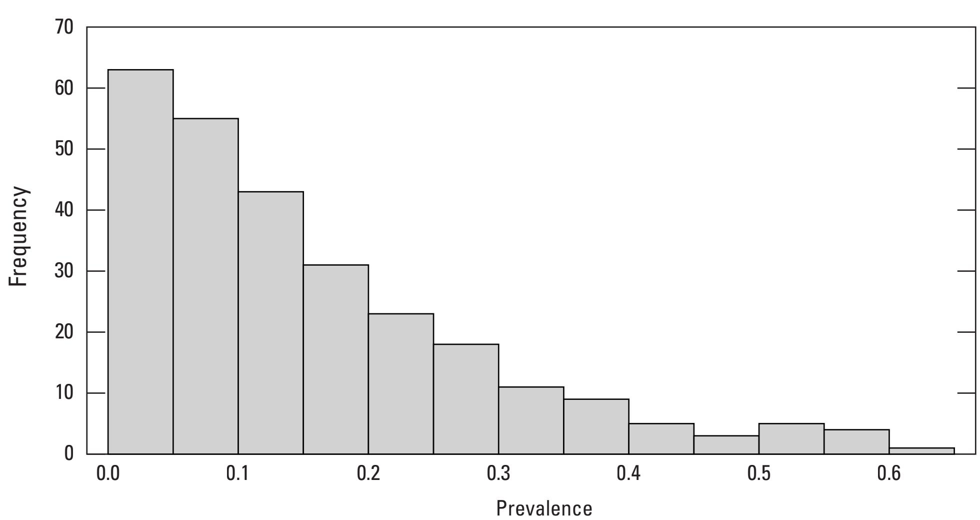

Figure 6. Histogram of prevalence for the 271 fish species modeled.

Table 3. Proportion of boosted regression tree model deviance and performance statistics for fluvial fish species distribution models.

[ITIS TSN, Integrated Taxonomic Information System taxonomic serial number; dev exp, deviance explained; AUC, area under the receiver operating characteristic curve; TSS, True Skill Statistic]

Scientific name	ITIS TSN	Dev exp	AUC	Sensitivity	Specificity	TSS
Acantharchus pomotis	168095	0.273455	0.882752	0.311644	0.975391	0.287035
Acipenser fulvescens	161071	0.380867	0.983065	0.083333	0.999321	0.082654
Alosa aestivalis	161703	0.242115	0.895504	0.059908	0.997401	0.057309
Alosa chrysochloris	161707	0.545646	0.969485	0.193929	0.996873	0.190802
Alosa pseudoharengus	161706	0.267295	0.942352	0.104167	0.998082	0.102249
Alosa sapidissima	161702	0.399099	0.962159	0.111413	0.998847	0.11026
Ambloplites ariommus	168099	0.206497	0.808224	0.395238	0.919476	0.314714
Ambloplites cavifrons	168098	0.270639	0.859777	0.243243	0.977143	0.220386
Ambloplites constellatus	168100	0.237221	0.820525	0.756757	0.706897	0.463653
Ambloplites rupestris	168097	0.378368	0.887134	0.68987	0.884575	0.574445
Ameiurus brunneus	164035	0.166562	0.767076	0.297297	0.935172	0.23247
Ameiurus catus	164037	0.321094	0.916765	0.100775	0.995449	0.096224
Ameiurus melas	164039	0.253209	0.84792	0.333595	0.957128	0.290723
Ameiurus natalis	164041	0.263703	0.834591	0.533159	0.894028	0.427187
Ameiurus nebulosus	164043	0.171372	0.813339	0.173433	0.974221	0.147654
Ameiurus platycephalus	164045	0.305727	0.878176	0.459746	0.949711	0.409457
Amia calva	161104	0.302421	0.893271	0.214386	0.984227	0.198614
Anguilla rostrata	161127	0.579605	0.966524	0.550324	0.986502	0.536826
Apeltes quadracus	166397	0.196631	0.759999	0.048276	0.998856	0.047132
Aphredoderus sayanus	164405	0.300667	0.863348	0.524964	0.927667	0.452631
Aplodinotus grunniens	169364	0.464044	0.934777	0.450032	0.978223	0.428255
Atractosteus spatula	201897	0.327291	0.879983	0.147651	0.991957	0.139608
Campostoma anomalum	163508	0.342746	0.866148	0.805279	0.764584	0.569863
Campostoma oligolepis	163509	0.403592	0.899819	0.573357	0.947117	0.520474
Carpiodes carpio	163919	0.459579	0.935917	0.424972	0.979507	0.404479
Carpiodes cyprinus	163917	0.409888	0.9248	0.345358	0.983411	0.328768
Carpiodes velifer	163920	0.43299	0.950289	0.178704	0.996741	0.175445
Catostomus ardens	163899	0.200186	0.815582	0.4	0.937853	0.337853
Catostomus catostomus	163894	0.364336	0.915454	0.328879	0.981898	0.310777
Catostomus clarkii	163901	0.120928	0.736508	0.725	0.602941	0.327941
Catostomus commersonii	553273	0.321196	0.856125	0.766613	0.780173	0.546786
Catostomus discobolus	163902	0.235354	0.855448	0.237288	0.968921	0.20621
Catostomus insignis	163905	0.342594	0.872527	0.794521	0.805556	0.600076
Catostomus latipinnis	163906	0.389564	0.918205	0.52381	0.965087	0.488897
Catostomus macrocheilus	163896	0.517104	0.947363	0.487936	0.983092	0.471027
Catostomus occidentalis	163908	0.371764	0.87913	0.769231	0.850746	0.619977
Catostomus platyrhynchus	163909	0.29378	0.878014	0.395503	0.954772	0.350275
Catostomus tahoensis	163914	0.231306	0.843252	0.534091	0.9163	0.45039
Centrarchus macropterus	168102	0.233491	0.843705	0.222772	0.977273	0.200045
Chrosomus eos	913993	0.275081	0.864422	0.35894	0.961485	0.320425
Chrosomus erythrogaster	913994	0.378826	0.903724	0.476281	0.962671	0.438952
Chrosomus neogaeus	913995	0.281555	0.882181	0.265472	0.983723	0.249196

Table 3. Proportion of boosted regression tree model deviance and performance statistics for fluvial fish species distribution models.—Continued

[ITIS TSN, Integrated Taxonomic Information System taxonomic serial number; dev exp, deviance explained; AUC, area under the receiver operating characteristic curve; TSS, True Skill Statistic]

Scientific name	ITIS TSN	Dev exp	AUC	Sensitivity	Specificity	TSS
Chrosomus oreas	913996	0.483888	0.95252	0.444444	0.987313	0.431758
Clinostomus elongatus	163373	0.378014	0.917223	0.332413	0.983649	0.316062
Clinostomus funduloides	163371	0.367426	0.890425	0.535004	0.943207	0.478211
Cottus aleuticus	167230	0.268905	0.867042	0.458333	0.926966	0.3853
Cottus bairdii	167237	0.341741	0.882146	0.530996	0.941915	0.47291
Cottus beldingii	167238	0.278707	0.864263	0.391473	0.958463	0.349936
Cottus carolinae	167239	0.461566	0.922751	0.685845	0.939123	0.624968
Cottus cognatus	167232	0.336628	0.888504	0.380466	0.968282	0.348749
Cottus confusus	167240	0.327649	0.913739	0.332344	0.982614	0.314958
Cottus hypselurus	167263	0.132454	0.771616	0.257143	0.962617	0.21976
Cottus rhotheus	167252	0.360038	0.885416	0.592262	0.931083	0.523345
Couesius plumbeus	163535	0.329156	0.911532	0.233333	0.987509	0.220842
Culaea inconstans	166399	0.367101	0.899351	0.467277	0.960563	0.42784
Cycleptus elongatus	163953	0.564445	0.978974	0.277344	0.99865	0.275994
Cyprinella analostana	163766	0.3326	0.891468	0.37851	0.966857	0.345367
Cyprinella camura	163776	0.396511	0.895493	0.545455	0.946378	0.491833
Cyprinella galactura	163782	0.355755	0.888902	0.504488	0.94702	0.451508
Cyprinella lutrensis	163792	0.474995	0.917167	0.824633	0.856017	0.68065
Cyprinella spiloptera	163803	0.441616	0.911478	0.630769	0.932669	0.563438
Cyprinella venusta	163809	0.300041	0.856279	0.603109	0.887436	0.490545
Cyprinella whipplei	163811	0.364136	0.914407	0.272167	0.98523	0.257397
Dorosoma cepedianum	161737	0.399987	0.908105	0.449346	0.96543	0.414776
Dorosoma petenense	161738	0.531808	0.951579	0.642458	0.968796	0.611255
Elassoma zonatum	168171	0.171764	0.834891	0.169935	0.976766	0.1467
Enneacanthus chaetodon	168108	0.333239	0.776934	0.175	0.997194	0.172194
Enneacanthus gloriosus	168113	0.35339	0.903757	0.395797	0.970793	0.36659
Enneacanthus obesus	168117	0.174368	0.858816	0.120357	0.990716	0.111073
Entosphenus tridentatus	159699	0.234519	0.871133	0.236111	0.974087	0.210198
Erimystax dissimilis	163821	0.4782	0.953587	0.206074	0.998033	0.204107
Erimystax x-punctatus	163824	0.552337	0.962424	0.319249	0.997421	0.31667
Erimyzon oblongus	163924	0.374934	0.910549	0.365091	0.978274	0.343365
Erimyzon sucetta	163922	0.436181	0.917359	0.140351	0.99613	0.136481
Esox americanus	162140	0.298854	0.871009	0.386146	0.957976	0.344122
Esox lucius	162139	0.348532	0.890135	0.463292	0.952175	0.415468
Esox niger	162143	0.269458	0.868757	0.251335	0.975325	0.22666
Etheostoma blennioides	168375	0.350333	0.872518	0.71538	0.857098	0.572478
Etheostoma caeruleum	168378	0.375941	0.884101	0.71934	0.871946	0.591286
Etheostoma camurum	168379	0.423671	0.944408	0.195021	0.99609	0.191111
Etheostoma cragini	168386	0.451713	0.925967	0.610577	0.964992	0.575569
Etheostoma exile	168393	0.218983	0.853443	0.145529	0.983955	0.129484
Etheostoma flabellare	168394	0.331533	0.865184	0.678233	0.86311	0.541343

Table 3. Proportion of boosted regression tree model deviance and performance statistics for fluvial fish species distribution models.—Continued

[ITIS TSN, Integrated Taxonomic Information System taxonomic serial number; dev exp, deviance explained; AUC, area under the receiver operating characteristic curve; TSS, True Skill Statistic]

Scientific name	ITIS TSN	Dev exp	AUC	Sensitivity	Specificity	TSS
Etheostoma fusiforme	168358	0.154425	0.838969	0.0887	0.989461	0.078161
Etheostoma gracile	168366	0.253753	0.873763	0.299639	0.972557	0.272196
Etheostoma kennicotti	168405	0.306356	0.868208	0.438776	0.962056	0.400832
Etheostoma lynceum	168456	0.228591	0.828477	0.322368	0.949907	0.272276
Etheostoma microperca	168411	0.284231	0.915993	0.093805	0.99721	0.091016
Etheostoma nigrum	168369	0.365097	0.877913	0.734528	0.858158	0.592686
Etheostoma olmstedi	168360	0.348808	0.878421	0.62068	0.907206	0.527886
Etheostoma punctulatum	168425	0.241885	0.819787	0.614286	0.835979	0.450265
Etheostoma radiosum	168426	0.442236	0.904717	0.775	0.90184	0.67684
Etheostoma rufilineatum	168428	0.308594	0.87293	0.463816	0.948875	0.412691
Etheostoma simoterum	168431	0.344841	0.888377	0.489933	0.945896	0.435828
Etheostoma spectabile	168368	0.354747	0.878778	0.65017	0.892431	0.542601
Etheostoma stigmaeum	168437	0.223362	0.831639	0.291429	0.960902	0.25233
Etheostoma swaini	168439	0.168828	0.797301	0.348168	0.93002	0.278188
Etheostoma variatum	168446	0.430444	0.948479	0.363636	0.989897	0.353533
Etheostoma whipplei	168448	0.278742	0.868433	0.423611	0.958088	0.381699
Etheostoma zonale	168449	0.369557	0.890777	0.544737	0.940565	0.485301
Exoglossum maxillingua	163356	0.350654	0.882804	0.584381	0.919952	0.504333
Fundulus catenatus	165660	0.395186	0.909642	0.554591	0.954455	0.509046
Fundulus diaphanus	165646	0.27785	0.883571	0.132161	0.992606	0.124767
Fundulus kansae	165654	0.433229	0.93005	0.484848	0.975882	0.460731
Fundulus notatus	165663	0.278731	0.857791	0.421846	0.943842	0.365689
Fundulus olivaceus	165655	0.325018	0.863321	0.646739	0.886878	0.533617
Fundulus seminolis	165667	0.236318	0.825568	0.555556	0.9	0.455556
Fundulus zebrinus	165658	0.281646	0.870353	0.322064	0.960104	0.282168
Gambusia affinis	165878	0.377076	0.893437	0.553506	0.942045	0.495551
Gila robusta	163558	0.466019	0.94813	0.511111	0.980477	0.491588
Hesperoleucus symmetricus	163565	0.14175	0.823042	0.4	0.918919	0.318919
Hiodon alosoides	161905	0.588814	0.974619	0.337079	0.996182	0.33326
Hiodon tergisus	161906	0.44664	0.965409	0.138826	0.998631	0.137458
Hybognathus argyritis	163362	0.238133	0.880773	0.188356	0.986748	0.175104
Hybognathus hankinsoni	163363	0.246744	0.848015	0.366331	0.951328	0.31766
Hybognathus nuchalis	163360	0.321884	0.914528	0.181481	0.992068	0.17355
Hybognathus placitus	163361	0.439618	0.92743	0.319343	0.987544	0.306887
Hybognathus regius	163359	0.220832	0.847169	0.109489	0.987353	0.096842
Hybopsis amblops	163476	0.371916	0.928704	0.31295	0.986764	0.299714
Hybopsis amnis	201917	0.08879	0.745147	0.038462	0.99636	0.034822
Hybopsis dorsalis	689231	0.383778	0.892119	0.63499	0.916969	0.551959
Hybopsis winchelli	201918	0.175267	0.800587	0.364964	0.929323	0.294287
Hypentelium etowanum	163950	0.290853	0.845199	0.72807	0.809211	0.537281
Hypentelium nigricans	163949	0.388501	0.888279	0.732438	0.863905	0.596343

Table 3. Proportion of boosted regression tree model deviance and performance statistics for fluvial fish species distribution models.—Continued

[ITIS TSN, Integrated Taxonomic Information System taxonomic serial number; dev exp, deviance explained; AUC, area under the receiver operating characteristic curve; TSS, True Skill Statistic]

Scientific name	ITIS TSN	Dev exp	AUC	Sensitivity	Specificity	TSS
Hypentelium roanokense	163951	0.248472	0.824097	0.630137	0.849057	0.479194
Ichthyomyzon castaneus	159725	0.303727	0.900949	0.204583	0.98967	0.194253
Ichthyomyzon fossor	159726	0.161049	0.839015	0.074919	0.993158	0.068077
Ichthyomyzon gagei	159727	0.234432	0.851544	0.251716	0.968947	0.220664
Ichthyomyzon greeleyi	159728	0.45405	0.947214	0.405172	0.987199	0.392371
Ictalurus furcatus	163997	0.552636	0.973566	0.351955	0.996694	0.34865
Ictalurus punctatus	163998	0.458865	0.927934	0.557011	0.964995	0.522006
Ictiobus bubalus	163955	0.475484	0.945477	0.388703	0.986591	0.375294
Ictiobus cyprinellus	163956	0.305864	0.899528	0.198404	0.985027	0.183432
Ictiobus niger	163957	0.401452	0.944709	0.20342	0.993299	0.196719
Labidesthes sicculus	166016	0.287671	0.869851	0.309819	0.967004	0.276823
Lampetra aepyptera	159705	0.278189	0.875435	0.29381	0.971875	0.265685
Lampetra richardsoni	159707	0.131325	0.829278	0.120301	0.979381	0.099682
Lepisosteus oculatus	161095	0.367696	0.914641	0.308318	0.980814	0.289132
Lepisosteus osseus	161094	0.370234	0.921974	0.266767	0.988237	0.255004
Lepisosteus platostomus	161096	0.344437	0.913221	0.269652	0.984032	0.253684
Lepisosteus platyrhincus	161098	0.415719	0.915319	0.693182	0.927184	0.620366
Lepomis auritus	168131	0.518801	0.938533	0.679472	0.9529	0.632372
Lepomis cyanellus	168132	0.229812	0.803475	0.806313	0.633304	0.439617
Lepomis gibbosus	168144	0.199398	0.80448	0.395651	0.915091	0.310742
Lepomis humilis	168151	0.310882	0.87573	0.425221	0.951312	0.376533
Lepomis macrochirus	168141	0.230955	0.804429	0.74223	0.72647	0.4687
Lepomis marginatus	168152	0.131248	0.765104	0.270833	0.927968	0.198802
Lepomis megalotis	168153	0.363044	0.872829	0.788885	0.79289	0.581775
Lepomis microlophus	168154	0.356235	0.877753	0.501742	0.940575	0.442317
Lepomis miniatus	168157	0.231595	0.841551	0.353448	0.949525	0.302974
Lepomis punctatus	168155	0.339714	0.866976	0.742268	0.806306	0.548574
Lepomis symmetricus	168156	0.079916	0.742381	0.085271	0.982927	0.068198
Lethenteron appendix	914061	0.264976	0.869286	0.221477	0.981845	0.203322
Lota lota	164725	0.386684	0.917286	0.310239	0.984138	0.294377
Luxilus albeolus	163826	0.357005	0.885674	0.605341	0.932615	0.537956
Luxilus cardinalis	163828	0.536089	0.941192	0.779343	0.933884	0.713227
Luxilus cerasinus	163830	0.391633	0.915836	0.531818	0.960384	0.492202
Luxilus chrysocephalus	163832	0.383179	0.88682	0.724564	0.871576	0.59614
Luxilus coccogenis	163834	0.324137	0.864438	0.66358	0.871122	0.534702
Luxilus cornutus	163836	0.334214	0.870555	0.580661	0.910411	0.491072
Luxilus zonatus	163840	0.630723	0.959716	0.834025	0.963542	0.797567
Lythrurus ardens	163847	0.381091	0.904683	0.238095	0.991104	0.2292
Lythrurus fasciolaris	201928	0.314831	0.865969	0.564229	0.91617	0.480398
Lythrurus fumeus	163853	0.178327	0.790839	0.22449	0.957474	0.181964
Lythrurus snelsoni	163859	0.357568	0.893694	0.62963	0.892157	0.521786

Table 3. Proportion of boosted regression tree model deviance and performance statistics for fluvial fish species distribution models.—Continued

[ITIS TSN, Integrated Taxonomic Information System taxonomic serial number; dev exp, deviance explained; AUC, area under the receiver operating characteristic curve; TSS, True Skill Statistic]

Scientific name	ITIS TSN	Dev exp	AUC	Sensitivity	Specificity	TSS
Lythrurus umbratilis	163861	0.364313	0.901866	0.475762	0.959964	0.435726
Macrhybopsis storeriana	163870	0.397959	0.944315	0.148618	0.996463	0.145081
Margariscus margarita	163873	0.11199	0.785377	0.127946	0.977122	0.105068
Menidia beryllina	165993	0.430607	0.930711	0.293286	0.99072	0.284006
Micropterus cataractae	564610	0.719817	0.976103	0.761905	0.983333	0.745238
Micropterus coosae	168163	0.419146	0.929699	0.474359	0.977865	0.452224
Micropterus dolomieu	550562	0.406161	0.896394	0.665835	0.91067	0.576505
Micropterus punctulatus	168161	0.284807	0.848611	0.524554	0.91048	0.435034
Micropterus salmoides	168160	0.183469	0.779072	0.572891	0.811039	0.383929
Minytrema melanops	163959	0.244046	0.847497	0.270626	0.966432	0.237058
Morone americana	167678	0.241679	0.881963	0.131443	0.994617	0.126061
Morone chrysops	167682	0.484609	0.950607	0.281488	0.992149	0.273637
Morone mississippiensis	167683	0.253758	0.908117	0.092308	0.993289	0.085596
Morone saxatilis	167680	0.306549	0.948822	0.105405	0.996512	0.101918
Moxostoma anisurum	163933	0.488505	0.941345	0.421702	0.985884	0.407586
Moxostoma breviceps	163929	0.629815	0.971759	0.529235	0.989788	0.519023
Moxostoma carinatum	163936	0.40987	0.940826	0.184188	0.995592	0.179779
Moxostoma collapsum	201946	0.218316	0.845831	0.292754	0.955723	0.248476
Moxostoma congestum	163931	0.056148	0.667734	0.404255	0.846154	0.250409
Moxostoma duquesnii	553274	0.349293	0.884616	0.432018	0.957324	0.389342
Moxostoma erythrurum	163939	0.375291	0.887747	0.580737	0.92266	0.503397
Moxostoma macrolepidotum	163928	0.507037	0.941914	0.548833	0.971986	0.520819
Moxostoma poecilurum	163932	0.244767	0.830327	0.432763	0.926398	0.359161
Moxostoma rupiscartes	163946	0.477555	0.944223	0.59919	0.976077	0.575267
Moxostoma valenciennesi	163947	0.333957	0.902576	0.207381	0.986479	0.193861
Mugil cephalus	170335	0.473984	0.946311	0.42284	0.98409	0.406929
Nocomis biguttatus	163395	0.428904	0.921717	0.507389	0.96761	0.474999
Nocomis leptocephalus	163393	0.52548	0.936905	0.76652	0.917698	0.684218
Nocomis micropogon	163392	0.410779	0.916047	0.44986	0.973422	0.423282
Notemigonus crysoleucas	163368	0.148159	0.77747	0.220037	0.952815	0.172852
Notropis amabilis	163410	0.349889	0.881717	0.690476	0.9	0.590476
Notropis atherinoides	163412	0.407726	0.91198	0.394856	0.977899	0.372756
Notropis blennius	163429	0.469311	0.940281	0.16771	0.99726	0.16497
Notropis boops	163430	0.446315	0.933524	0.456093	0.976185	0.432279
Notropis buccatus	163478	0.268151	0.837997	0.57448	0.878176	0.452656
Notropis chiliticus	163435	0.45368	0.913036	0.708155	0.906977	0.615131
Notropis cummingsae	163438	0.427514	0.921365	0.468421	0.972715	0.441136
Notropis girardi	689231	0.583401	0.98819	0.571429	0.999024	0.570453
Notropis heterolepis	163442	0.344755	0.921033	0.157703	0.993877	0.151581
Notropis hudsonius	163446	0.337654	0.903722	0.266983	0.984114	0.251097
Notropis leuciodus	163404	0.367799	0.894715	0.529586	0.945005	0.47459

Table 3. Proportion of boosted regression tree model deviance and performance statistics for fluvial fish species distribution models.—Continued

[ITIS TSN, Integrated Taxonomic Information System taxonomic serial number; dev exp, deviance explained; AUC, area under the receiver operating characteristic curve; TSS, True Skill Statistic]

Scientific name	ITIS TSN	Dev exp	AUC	Sensitivity	Specificity	TSS
Notropis longirostris	163451	0.223025	0.821563	0.465385	0.906822	0.372207
Notropis lutipinnis	163452	0.480297	0.938344	0.601695	0.970123	0.571818
Notropis nubilus	163453	0.453838	0.920279	0.66951	0.937561	0.607071
Notropis percobromus	163456	0.325086	0.890405	0.359335	0.972346	0.331681
Notropis petersoni	163460	0.206225	0.824851	0.334694	0.950969	0.285663
Notropis photogenis	163461	0.39533	0.910658	0.462457	0.969155	0.431612
Notropis procne	163407	0.372988	0.909449	0.414097	0.972867	0.386964
Notropis rubellus	163409	0.358354	0.899983	0.351249	0.977115	0.328364
Notropis stramineus	163419	0.395401	0.896293	0.628526	0.919472	0.547998
Notropis telescopus	163470	0.315882	0.881918	0.449064	0.954225	0.40329
Notropis texanus	163420	0.477143	0.93912	0.49359	0.978794	0.472384
Notropis topeka	163471	0.28941	0.894568	0.15	0.997057	0.147057
Notropis volucellus	163421	0.304347	0.887732	0.24728	0.982668	0.229948
Noturus albater	164006	0.170003	0.779193	0.414286	0.879195	0.29348
Noturus exilis	164010	0.399989	0.901661	0.650852	0.931805	0.582657
Noturus flavus	164013	0.31892	0.887501	0.353198	0.97155	0.324748
Noturus gyrinus	164003	0.298364	0.893476	0.203385	0.986463	0.189849
Noturus insignis	164004	0.297257	0.852559	0.617958	0.88115	0.499108
Noturus leptacanthus	164019	0.303465	0.864779	0.513453	0.918728	0.432181
Noturus miurus	164020	0.301551	0.895617	0.181458	0.989227	0.170685
Noturus nocturnus	164005	0.27884	0.873358	0.329562	0.967254	0.296816
Oncorhynchus clarkii	161983	0.378698	0.884544	0.847158	0.748886	0.596044
Oncorhynchus kisutch	161977	0.382641	0.88724	0.675676	0.931111	0.606787
Oncorhynchus mykiss	161989	0.257544	0.829257	0.809524	0.723562	0.533086
Oncorhynchus tshawytscha	161980	0.251126	0.855556	0.35	0.976	0.326
Opsopoeodus emiliae	163876	0.281799	0.899119	0.119841	0.992556	0.112398
Perca flavescens	168469	0.284847	0.863506	0.292748	0.968166	0.260914
Percina caprodes	168472	0.321559	0.874191	0.488583	0.939851	0.428434
Percina evides	168483	0.501329	0.961432	0.348894	0.992687	0.341581
Percina maculata	168488	0.347941	0.884012	0.508547	0.940377	0.448924
Percina nigrofasciata	168490	0.412658	0.897335	0.735385	0.880862	0.616246
Percina peltata	168474	0.363216	0.912057	0.39267	0.976651	0.369322
Percina phoxocephala	168494	0.470636	0.942591	0.399734	0.988413	0.388147
Percina roanoka	168496	0.361015	0.898049	0.580913	0.938679	0.519592
Percina sciera	168475	0.335449	0.898903	0.325653	0.975187	0.300839
Percopsis omiscomaycus	164409	0.338242	0.911788	0.238532	0.987773	0.226306
Petromyzon marinus	159722	0.288921	0.905082	0.203704	0.988506	0.192209
Phenacobius mirabilis	163502	0.342045	0.891119	0.476966	0.954545	0.431511
Pimephales notatus	163516	0.472587	0.917364	0.804217	0.873249	0.677466
Pimephales promelas	163517	0.379183	0.896154	0.520965	0.947253	0.468217
Pimephales vigilax	163518	0.496734	0.944375	0.503537	0.979266	0.482803

Table 3. Proportion of boosted regression tree model deviance and performance statistics for fluvial fish species distribution models.—Continued

[ITIS TSN, Integrated Taxonomic Information System taxonomic serial number; dev exp, deviance explained; AUC, area under the receiver operating characteristic curve; TSS, True Skill Statistic]

Scientific name	ITIS TSN	Dev exp	AUC	Sensitivity	Specificity	TSS
Platygobio gracilis	163882	0.498321	0.947296	0.45977	0.98319	0.44296
Polyodon spathula	161088	0.248879	0.917067	0.048544	0.999031	0.047575
Pomoxis annularis	168166	0.245571	0.855397	0.253546	0.97243	0.225977
Pomoxis nigromaculatus	168167	0.247369	0.857795	0.232049	0.974719	0.206768
Prosopium williamsoni	162009	0.345543	0.900391	0.380902	0.969268	0.350171
Ptychocheilus grandis	163524	0.291792	0.86014	0.548387	0.852459	0.400846
Ptychocheilus oregonensis	163523	0.44812	0.959454	0.362007	0.992569	0.354576
Pylodictis olivaris	164029	0.455277	0.936639	0.417552	0.981831	0.399383
Rhinichthys atratulus	163382	0.402116	0.899928	0.631613	0.92721	0.558823
Rhinichthys cataractae	163384	0.327922	0.874088	0.51829	0.931649	0.449939
Rhinichthys obtusus	689949	0.429729	0.912374	0.620787	0.939394	0.560181
Rhinichthys osculus	163387	0.37346	0.88653	0.644156	0.895433	0.539589
Richardsonius balteatus	163528	0.358293	0.912851	0.348675	0.97799	0.326665
Salmo salar	161996	0.265759	0.87731	0.25	0.977568	0.227568
Salvelinus confluentus	162004	0.448997	0.923728	0.645706	0.948454	0.594159
Salvelinus fontinalis	162003	0.31771	0.859587	0.580209	0.889417	0.469626
Sander canadensis	650171	0.540302	0.965833	0.396	0.994145	0.390145
Sander vitreus	650173	0.47434	0.945396	0.376731	0.990662	0.367393
Scaphirhynchus platorynchus	161082	0.46423	0.970359	0.149007	0.998761	0.147767
Semotilus atromaculatus	163376	0.40266	0.891953	0.807999	0.806592	0.614591
Semotilus corporalis	163375	0.281146	0.85258	0.49385	0.921525	0.415375
Thoburnia rhothoeca	553276	0.383496	0.912758	0.596491	0.945455	0.541946
Umbra limi	162153	0.447939	0.918557	0.631502	0.944385	0.575887
Umbra pygmaea	162148	0.407484	0.909618	0.587189	0.944929	0.532118

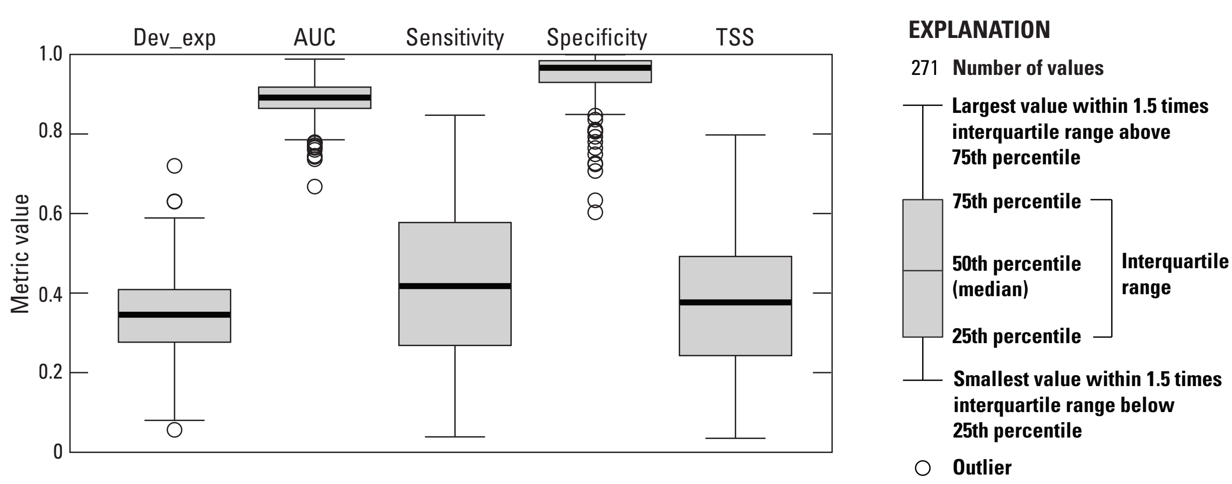

Figure 7. Boxplots of proportion of model deviance explained. Dev_exp, deviance explained; AUC, area under the receiver operating curve; TSS, True Skill Statistic.

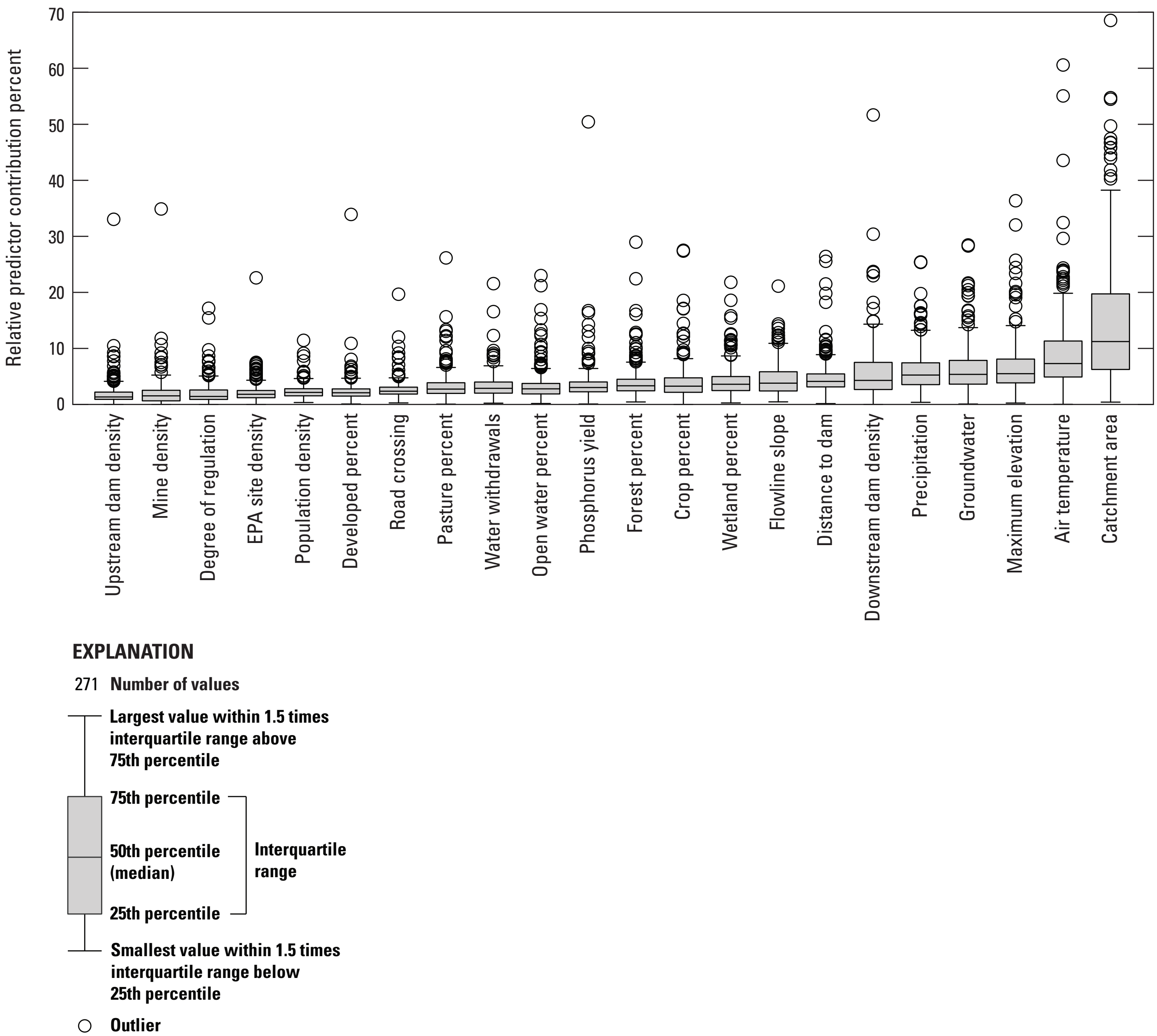

Figure 8. Boxplots of relative importance of the predictor variables for fluvial fish species presence, absence, and prevalence. See table 1 for predictor variable explanations. EPA, Environmental Protection Agency.

Table 4. Fluvial fish species considered sensitive to climate influences in the conterminous United States.

[L_temp represents the relative importance of mean annual air temperature in percent, and N_precip represents the relative importance of mean annual precipitation in percent. Species with the sum of temperature and precipitation (sum) with variable importance greater than 20 percent are considered sensitive. ITIS TSN, Integrated Taxonomic Information System taxonomic serial number]

Scientific name	ITIS TSN	L_temp	N_precip	Sum
Erimyzon sucetta	163922	60.70	4.26	64.96
Lepisosteus platyrhincus	161098	55.17	0.60	55.78
Lepomis auritus	168131	43.63	1.91	45.54
Gambusia affinis	165878	29.69	12.19	41.88
Culaea inconstans	166399	14.36	25.39	39.75
Lythrurus snelsoni	163859	32.51	2.56	35.08
Pimephales vigilax	163518	16.48	16.28	32.76
Campostoma oligolepis	163509	22.21	10.01	32.22
Cyprinella lutrensis	163792	11.71	19.82	31.52
Umbra limi	162153	5.59	25.47	31.07
Notropis texanus	163420	23.87	6.98	30.85
Erimyzon oblongus	163924	22.84	5.81	28.64
Rhinichthys obtusus	689949	11.76	16.38	28.15
Esox lucius	162139	11.78	16.28	28.06
Phenacobius mirabilis	163502	13.05	14.62	27.66
Notropis heterolepis	163446	18.05	9.53	27.59
Micropterus coosae	168163	15.60	11.97	27.57
Notropis boops	163430	18.00	9.36	27.36
Enneacanthus gloriosus	168113	24.02	3.13	27.15
Percina nigrofasciata	168490	12.70	14.42	27.12
Etheostoma cragini	168386	24.40	2.62	27.03
Lepisosteus oculatus	161095	23.65	3.24	26.89
Oncorhynchus clarkii	161983	16.84	10.05	26.88
Cottus hypselurus	167263	0.71	25.46	26.17
Lepomis macrochirus	168141	21.83	4.33	26.17
Sander vitreus	650173	22.51	3.54	26.05
Etheostoma nigrum	168369	16.35	9.32	25.67
Lepomis punctatus	168155	21.54	3.69	25.24
Ichthyomyzon fossor	159726	18.88	5.91	24.79
Etheostoma exile	168393	7.20	17.56	24.75
Salvelinus fontinalis	162003	19.88	4.76	24.64
Chrosomus eos	913993	14.87	9.73	24.61
Nocomis biguttatus	163395	14.31	10.14	24.45
Chrosomus oreas	913996	11.08	13.33	24.41
Etheostoma fusiforme	168358	21.15	3.11	24.26
Lythrurus ardens	163847	14.88	9.32	24.20
Chrosomus neogaeus	913995	13.22	10.46	23.68
Couesius plumbeus	163535	7.67	15.99	23.65
Salvelinus confluentus	162004	9.62	13.95	23.57
Catostomus commersonii	553273	15.36	7.97	23.34
Fundulus zebrinus	165658	14.37	8.68	23.05

Table 4. Fluvial fish species considered sensitive to climate influences in the conterminous United States.—Continued

[L_temp represents the relative importance of mean annual air temperature in percent, and N_precip represents the relative importance of mean annual precipitation in percent. Species with the sum of temperature and precipitation (sum) with variable importance greater than 20 percent are considered sensitive. ITIS TSN, Integrated Taxonomic Information System taxonomic serial number]

Scientific name	ITIS TSN	L_temp	N_precip	Sum
Etheostoma caeruleum	168378	5.39	17.62	23.01
Hesperoleucus symmetricus	163565	18.61	4.11	22.72
Ameiurus melas	164039	6.48	16.15	22.63
Mugil cephalus	170335	16.52	5.95	22.47
Cottus carolinae	167239	18.93	3.44	22.37
Lepomis megalotis	168153	18.59	3.67	22.26
Opsopoeodus emiliae	163876	17.11	4.94	22.05
Anguilla rostrata	161127	17.28	4.75	22.03
Etheostoma spectabile	168368	11.07	10.85	21.92
Cyprinella venusta	163809	15.71	6.17	21.88
Lepomis marginatus	168152	14.03	7.82	21.85
Cottus cognatus	167232	12.72	9.05	21.77
Enneacanthus obesus	168117	12.89	8.34	21.23
Semotilus atromaculatus	163376	10.07	10.99	21.05
Micropterus salmoides	168160	13.99	6.99	20.98
Etheostoma radiosum	168426	9.64	11.20	20.84
Etheostoma whipplei	168448	8.58	12.02	20.59
Centrarchus macropterus	168102	10.09	10.41	20.50
Notropis hudsonius	163404	15.70	4.59	20.29
Fundulus diaphanus	165646	10.78	9.50	20.28

Table 5. Fluvial fish species in the conterminous United States considered responsive to anthropogenic stressors.

[Species for which the sum of the anthropogenic variable importance (sum of relative importance) values is greater than 50 percent are considered sensitive. ITIS TSN, Integrated Taxonomic Information System taxonomic serial number]

Scientific name	ITIS TSN	Sum of relative importance
Acipenser fulvescens	161071	82.65
Alosa aestivalis	161703	55.63
Alosa sapidissima	161702	52.69
Apeltes quadracus	166397	63.19
Atractosteus spatula	201897	61.57
Catostomus discobolus	163902	50.31
Catostomus latipinnis	163906	55.60
Cottus confusus	167240	53.16
Cyprinella camura	163776	57.91
Enneacanthus chaetodon	168108	71.93
Erimystax dissimilis	163821	52.40
Erimystax x-punctatus	163824	56.54
Etheostoma camurum	168379	55.38
Etheostoma lynceum	168456	51.70
Etheostoma microperca	168411	57.86
Etheostoma spectabile	168368	53.08
Etheostoma variatum	168446	54.02
Fundulus catenatus	165660	53.71
Fundulus notatus	165663	51.11
Gila robusta	163558	68.38
Hybognathus regius	163359	53.79
Hybopsis amblops	163476	51.52
Hybopsis amnis	201917	64.91
Hybopsis dorsalis	689231	56.50
Lampetra aepyptera	159705	50.50
Lepomis symmetricus	168156	53.06
Luxilus zonatus	163840	73.31
Margariscus margarita	163873	54.86
Micropterus cataractae	564610	56.24
Morone americana	167678	53.84
Nocomis leptocephalus	163393	50.27
Notropis amabilis	163410	50.36
Notropis lutipinnis	163453	52.79
Notropis nubilus	163456	56.73
Notropis topeka	163471	67.81
Noturus exilis	164010	51.79
Polyodon spathula	161088	55.78
Richardsonius balteatus	163528	52.21
Scaphirhynchus platorynchus	161082	54.71
Thoburnia rhothoeca	553276	56.10

Table 6. Presences, absences, and prevalence for three fluvial fish species selected to provide examples of species distribution model output.

[ITIS, Integrated Taxonomic Information System]

Scientific name	ITIS	Presences	Absences	Prevalence	Characteristic
Semotilus atromaculatus	163376	13,586	13,198	0.5072	Highest presence
Cottus beldingii	167238	141	1,080	0.1155	Median prevalence
Enneacanthus chaetodon	168108	10	1,099	0.009	Lowest presence

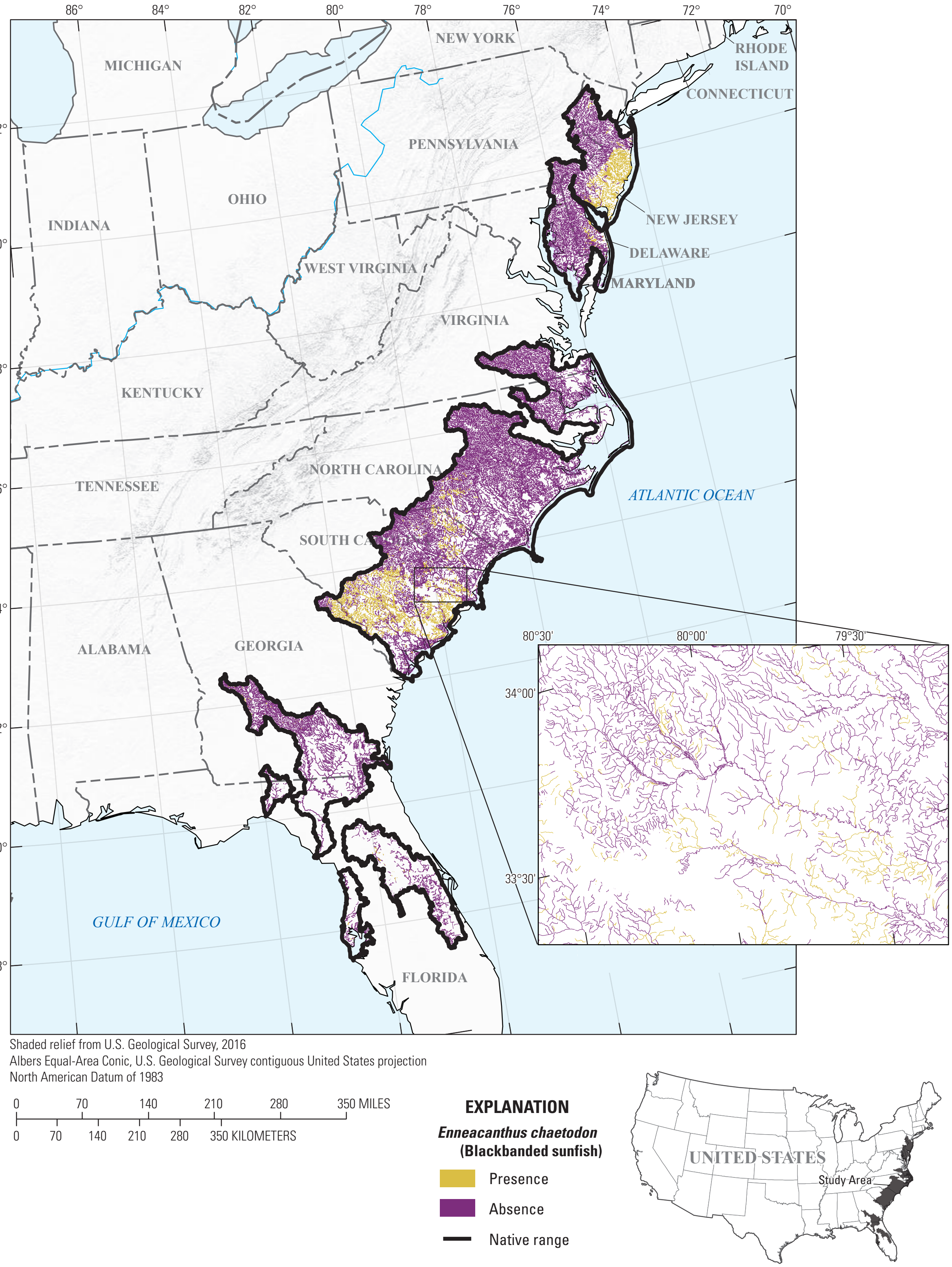

Figure 9. Map of species distribution model predictions for *Enneacanthus chaetodon* (Baird, 1855) (blackbanded sunfish).

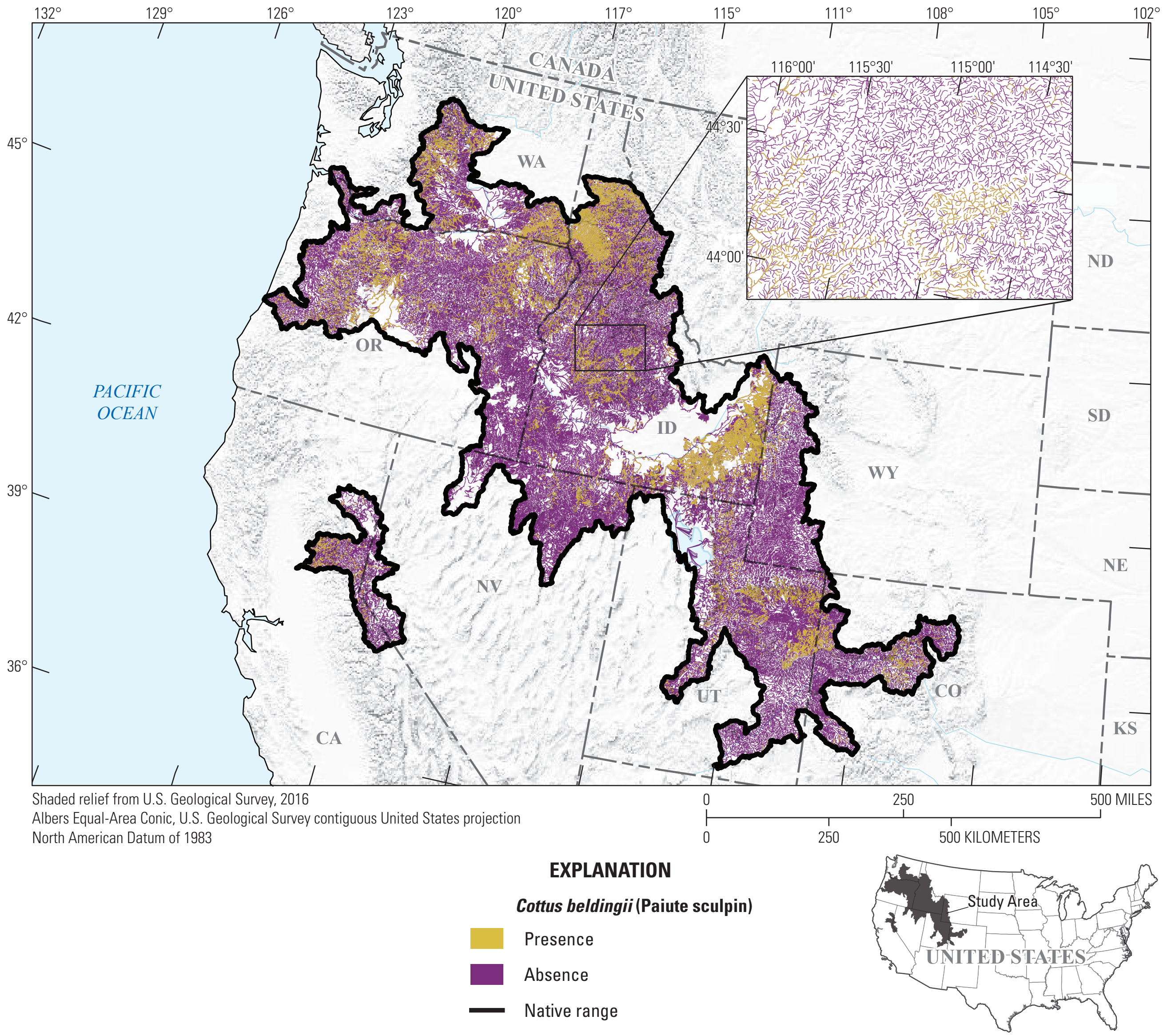

Figure 10. Map of species distribution model predictions for *Cottus beldingii* (Eigenmann and Eigenmann, 1891) (Paiute sculpin).

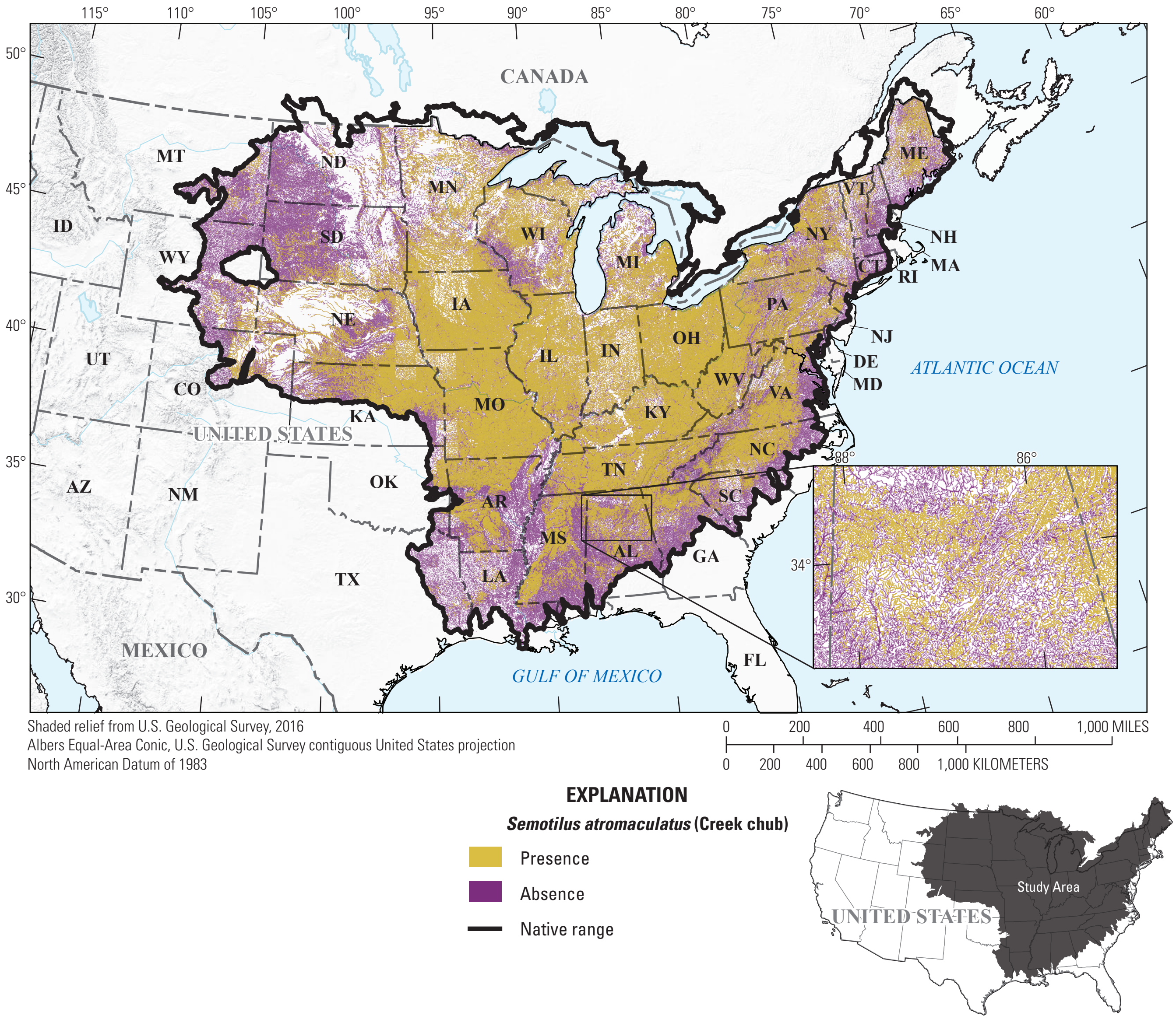

Figure 11. Map of species distribution model predictions for *Semotilus atromaculatus* (Mitchill, 1818) (creek chub).

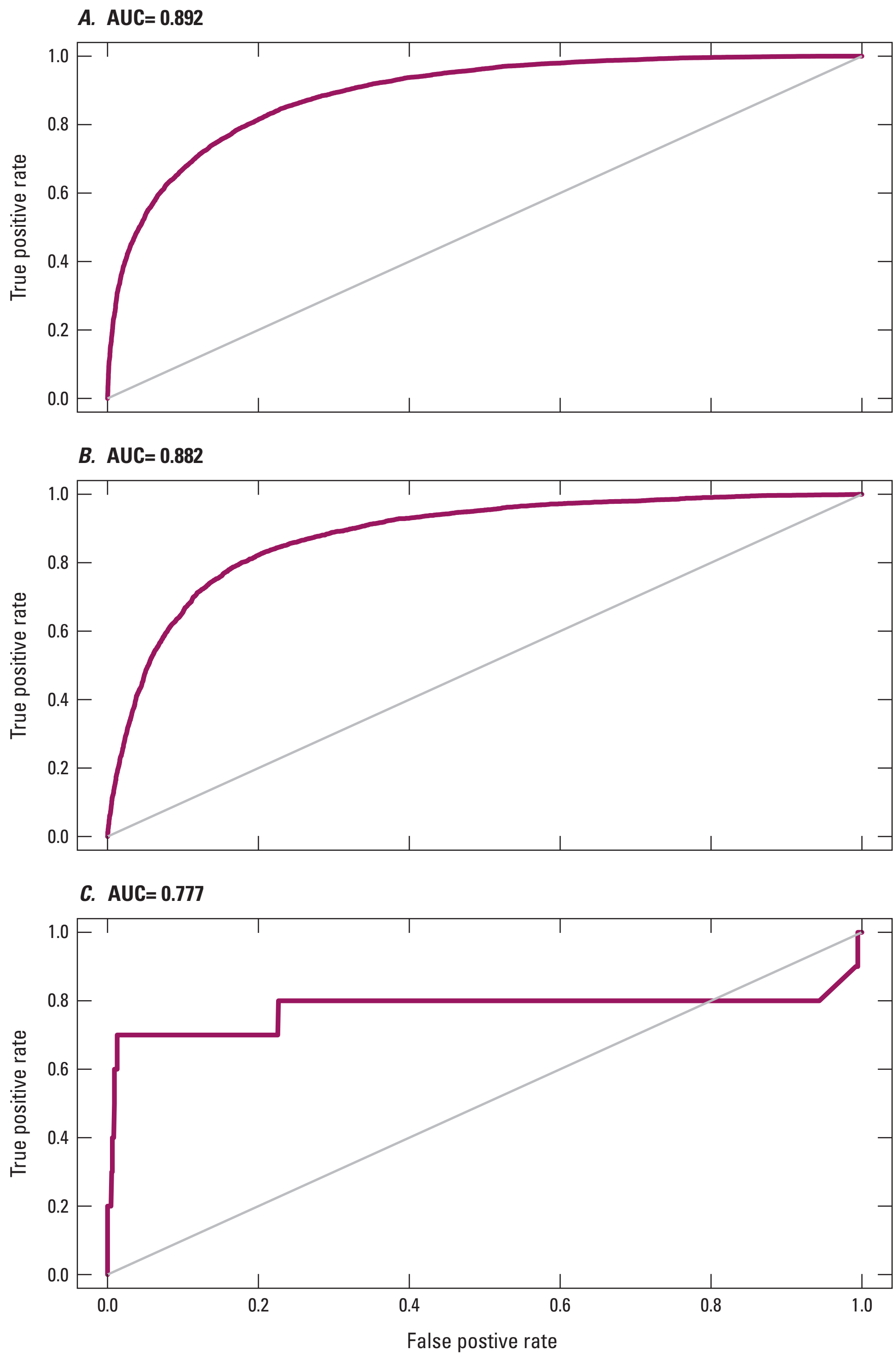

Figure 12. Area under the receiver operating characteristic curve (AUC) plots for three example fish species: *A*, *Semotilus atromaculatus* (Mitchill, 1818); *B*, *Cottus beldingii* (Eigenmann and Eigenmann, 1891); and *C*, *Enneacanthus chaetodon* (Baird, 1855).

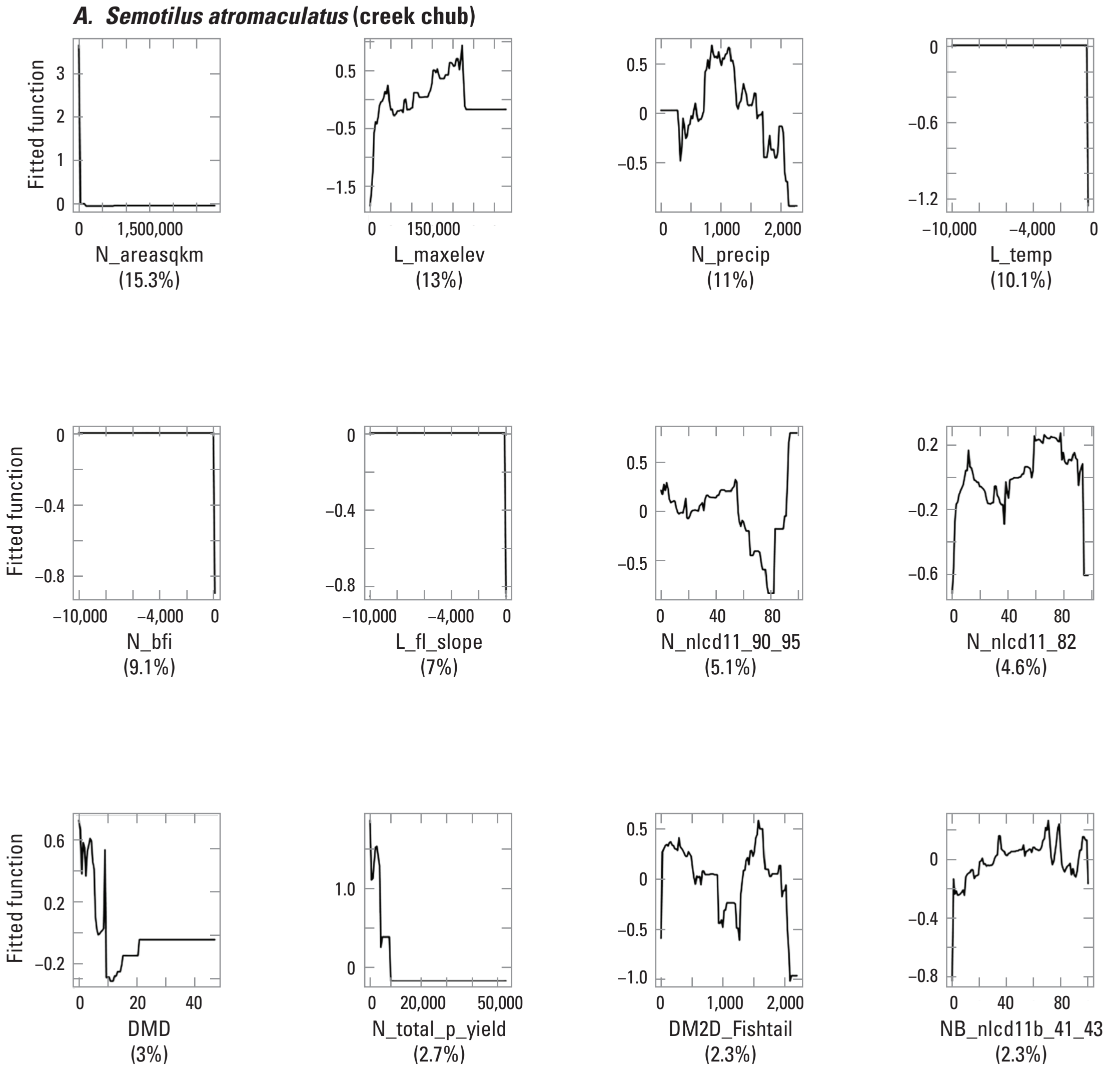

Figure 13. Partial dependence plots for three example fish species: *A*, *Semotilus atromaculatus* (Mitchill, 1818) (creek chub); *B*, *Cottus beldingii* (Eigenmann and Eigenmann, 1891) (Paiute sculpin); and *C*, *Enneacanthus chaetodon* (Baird, 1855) (blackbanded sunfish). The rug tiles in each figure represent the distribution of predictor variable values.

B. Cottus beldingii **(Paiute sculpin)**

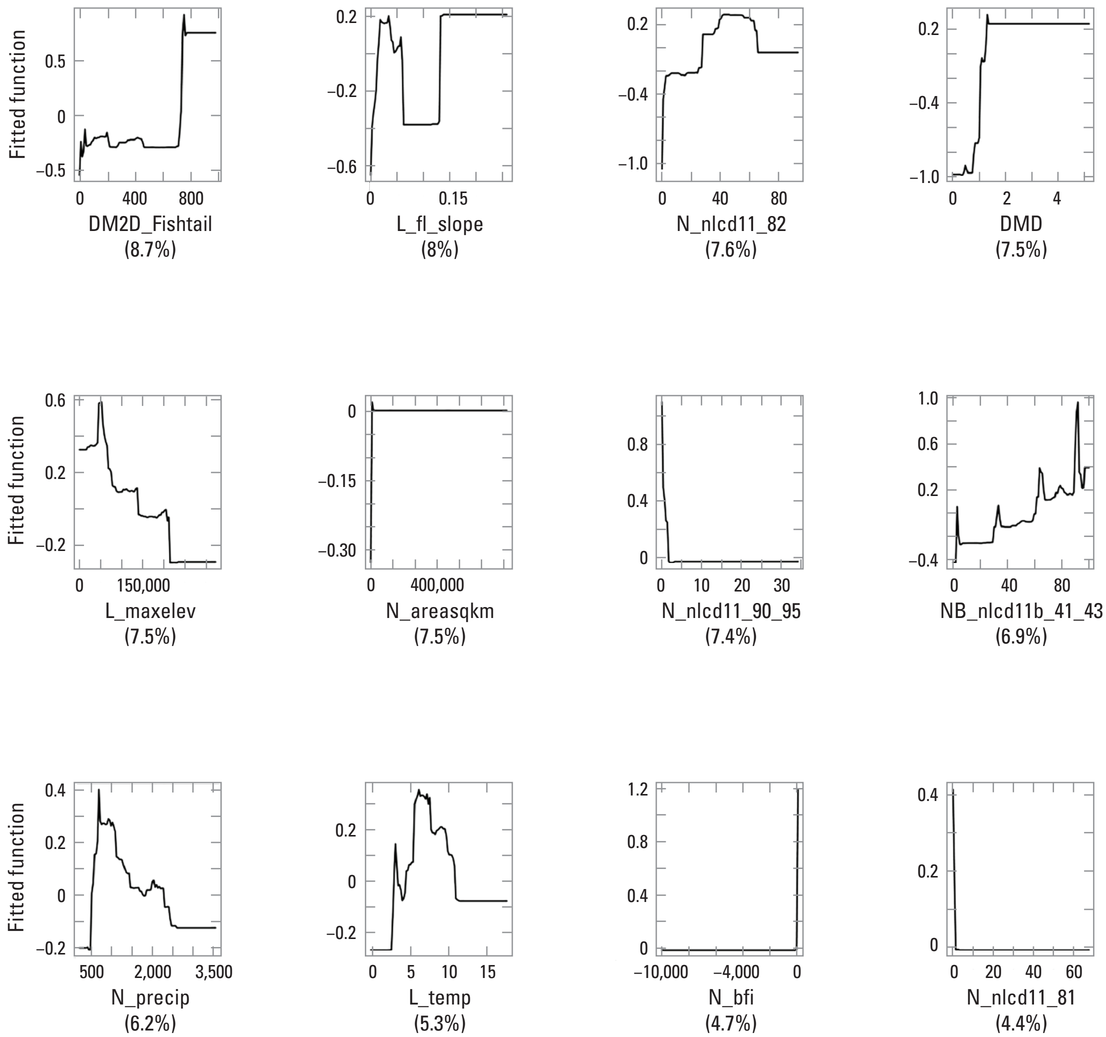

Figure 13.—Continued.

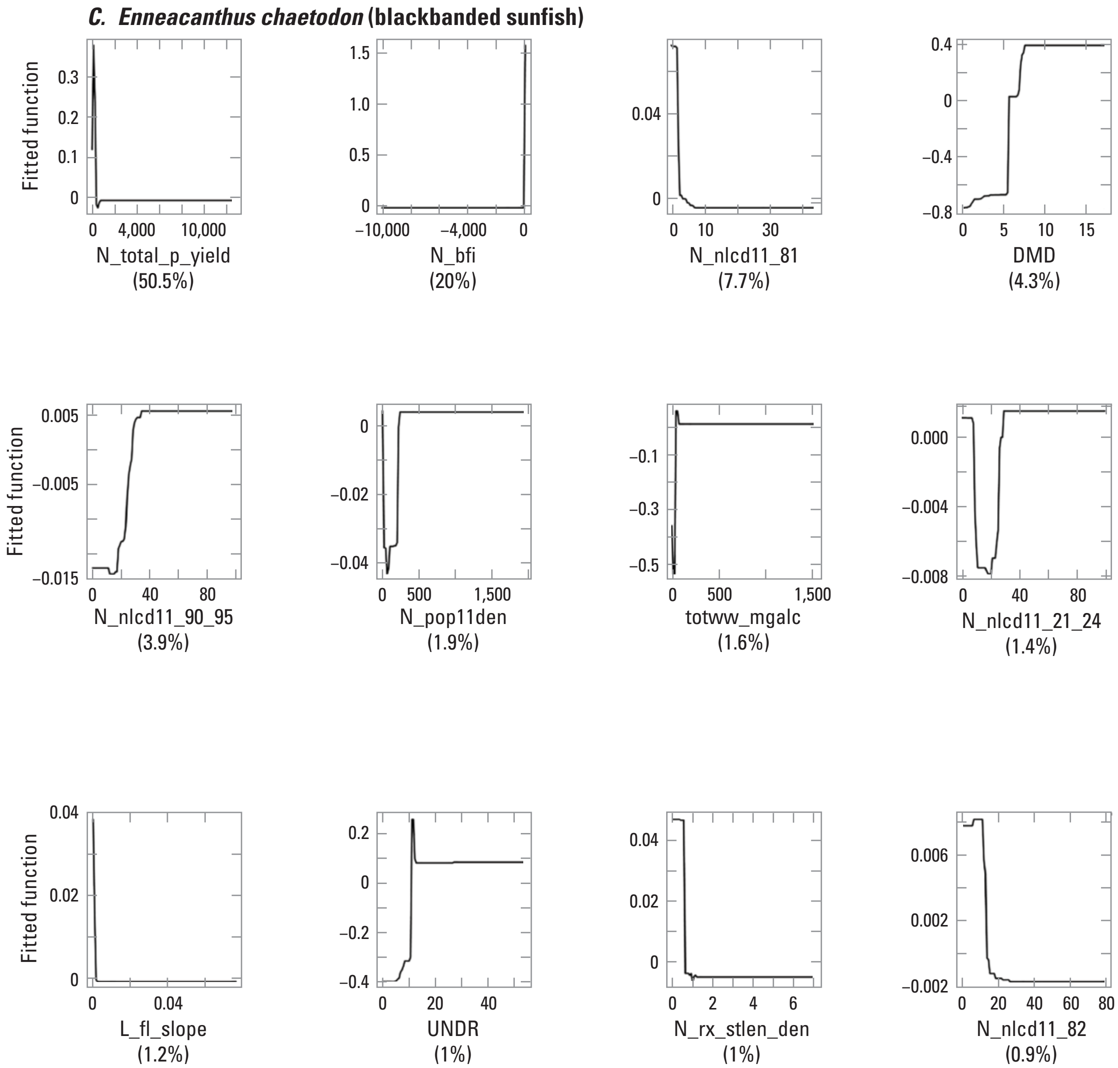

Figure 13.—Continued.

Discussion

The models in this study represent 271 (~34 percent) of approximately 800 known freshwater fish species in the United States (Warren and Burr, 1994), and to our knowledge, this study represents the largest effort of its type for freshwater fishes based on geographic and taxonomic scope in the conterminous United States. In addition, the unprecedented spatial scale of this modeling effort provides the ability to identify locations that support many species and locations that support individual species of conservation or recreational importance, including Species of Greatest Conservation Need or priority game fish.

With this scope in mind, the SDMs generated provide a critical framework to develop additional products that may be beneficial to management and conservation of fluvial fishes in the United States. In Cooper and others (2019), species characterized as common within nine large ecoregions in the conterminous United States were evaluated using range extent, abundance, and habitat usage, and their distributions were modeled within ecoregions. The amount of protected land area in catchments required to consider them protected (protection target levels) was established for all streams in the conterminous United States by using information on protected areas from the USGS Protected Areas Database of the United States (USGS, 2020) and the known responses of fish communities to two prominent landscape stressors (urban and agricultural land uses). An assessment of protection target levels in conjunction with predicted species distributions indicated that protected areas are severely lacking among fish habitats for most common species in the United States (Cooper and others, 2019). Based on these methods, predicted presences from SDMs developed in this study can be coupled with protected areas from the Protected Areas Database of the United States dataset to identify the percentage and location of habitats that meet protection target levels. This type of analysis can identify spatial gaps in species protection for both rare and common species and harkens to the foundational analyses that spurred the inception of the USGS Gap Analysis Project.

Future expansion of this modeling could provide additional insights and products in support of aquatic conservation initiatives. For instance, further evaluation of species responses based on model results can be used to gain understanding of the natural and anthropogenic factors limiting species distributions. Model results for more climate-sensitive fish species can be used to understand potential effects on habitat suitability from climate change, and they can also help map habitats potentially gained or lost with projected changes in climate for individual species. Further, this information could be coupled with known locations of dams to explore the role of fragmentation in constraining range expansions and population dynamics under climate change. For a subset of species in this study, both native and introduced ranges were available. This information could be used to test or project native range models into introduced portions of a species' range, providing an analytical framework for understanding potential invasiveness and ability of species to inhabit environments with novel conditions outside of a species' known native range.

Evaluating Habitat Condition

Data representing stream fragmentation by dams (Cooper and others, 2017) can be used to analyze species-level fragmentation, quantifying the amount of connected habitat for any given location. Such information can be used as the basis for analyzing fish passage mitigation opportunities and identifying potential project locations that maximize habitat reconnections for multiple species, including migratory or imperiled species. Projected species presence/absence can result in much-needed information for conservation because these projections provide results for numerous unsampled stream reaches through a given species' range. This information could inform field sampling efforts, with potential to identify previously unknown populations.

Identifying Sensitive Species

Climate change may dramatically affect fluvial fish distribution by altering air temperature and precipitation. The SDMs used in this study can be used to assess the effects of a changing climate by incorporating climate variables as predictor variables. The framework of building up SDMs, selecting model evaluation metrics, and ranking predictor relative importance will help classify and identify climate-sensitive species and sensitive stream reaches, information that can benefit natural resource managers.

Next Steps in Modeling

Extending modeling efforts to additional freshwater fish species could provide SDMs for species that have limited distributions or are underrepresented in the current Aquatic GAP fish database. Modeling of these species would likely require testing and application of novel analytical techniques (for example, weighted BRT, Maxent, random forest, deep-learning techniques, and community-based modeling approaches) to account for cases of limited presence/absence data. Further, adding measures of model uncertainty would improve model output by providing users with predictive uncertainty values that could be applied and analyzed for all predicted habitats for a given species. Yu and others (2020) used a novel approach that uses species abundances in model weighting to develop presence/absence SDMs. Results for 55 fluvial fish species native to the northeast United States indicated that this weighting approach outperforms a traditional, unweighted modeling approach for rarer fish species that have smaller range extents, lower abundance, and less diverse habitat usage (Yu and others, 2020). As a result, this new approach has the potential to improve SDMs for species of high conservation importance, with utility not only in aquatic studies but terrestrial realms as well. While updating the fish dataset used in SDM analysis was a focus during this project, acquiring new information on distributions of other types of aquatic organisms, including freshwater mussels, would set the stage for developing SDMs for other aquatic taxa.

Summary

This study offers insights into stream habitat suitability for 271 fluvial fish species (including Species of Greatest Conservation Need and game species) in the conterminous United States. Our results showed that network catchment area, mean annual air temperature of the local catchment, and maximum elevation of the local catchment were the three strongest natural predictors of fish distributions. Additionally, downstream main stem dam density, distance to downstream main stem dam, and the percentage of pasture/hay land use area within network catchment boundaries were the three strongest anthropogenic predictors of distributions. Additionally, by considering species-specific responses to individual environmental variables, we found that 40 fish species were sensitive to anthropogenic stressors, and 61 species were sensitive to climate variables. Such insights into the overall important predictors of fish distributions as well as important predictors for specific species can help natural resource managers better understand current habitat conditions and potential variations in the future. These and additional modeling efforts and potential applications using results from species distribution models, such as those described here, could contribute to efforts to conduct a national assessment in support of the Aquatic Gap Analysis Project, including integrating the effects of conservation actions into a landscape-scale context.

Data Access

Each of the datasets produced for this analysis are available to the public. The data are organized under a parent item with four child items. The parent item describes the modeling effort and includes a species list (species_model_list.csv), which provides a complete list of the species that have been modeled to date with the common name and the Integrated Taxonomic Information System taxonomic serial number allowing the user to know which species have been modeled. In addition, the species list includes the model's digital object identifier, the modeled habitat type, and geographic extent of that model. The citations for the data products include the following:

- Model Collection:
  - Wieferich and others (2022)
- Model Parameters:
  - Ross and others (2022)
  - Cooper and Infante (2022)
- Species Ranges and Occurrence Data:
  - Cooper and others (2022)
  - Yu, Ross, and others (2022)
- Species Distribution Model Predictions:
  - Yu, Cooper, and others (2022)

References Cited

Allan, J.D., 2004, Landscapes and riverscapes—The influence of land use on stream ecosystems: Annual Review of Ecology Evolution and Systematics, v. 35, p. 257–284, accessed December 1, 2019, at https://doi.org/10.1146/annurev.ecolsys.35.120202.110122.

Allouche, O., Tsoar, A., Kadmon, R., 2006, Assessing the accuracy of species distribution models—Prevalence, kappa and the true skill statistic (TSS): Journal of Applied Ecology, v. 43, no. 6, p. 1223–1232, accessed December 1, 2019, at https://doi.org/10.1111/j.1365-2664.2006.01214.x.

Bouska, K.L., Whitledge, G.W., and Lant, C., 2015, Development and evaluation of species distribution models for fourteen native central U.S. fish species: Hydrobiologia, v. 747, p. 159–176, accessed December 1, 2019, at https://doi.org/10.1007/s10750-014-2134-8.

Cooper, A.R., and Infante, D.M., 2022, Dam metrics representing stream fragmentation and flow alteration for the conterminous United States linked to the NHDPLUSV2.1: U.S. Geological Survey data release, accessed May 20, 2022, at https://doi.org/10.5066/P94JQOFU.

Cooper, A.R., Infante, D.M., Daniel, W.M., Wehrly, K.E., Wang, L., and Brenden, T.O., 2017, Assessment of dam effects on streams and fish assemblages of the conterminous USA: Science of the Total Environment, v. 586, p. 879–889, accessed December 1, 2019, at https://doi.org/10.1016/j.scitotenv.2017.02.067.

Cooper, A.R., Tsang, Y.P., Infante, D.M., Daniel, W.M., McKerrow, A.J., and Wieferich, D.J., 2019, Protected areas lacking for many common fluvial fishes of the conterminous USA: Diversity and Distributions, v. 25, no. 8, p. 1289–1303, accessed December 1, 2019, at https://doi.org/10.1111/ddi.12937.

Cooper, A.R., Yu, H., Infante, D.M., and Ross, J.A., 2022, Coarse range maps for fish species in the conterminous United States using HUC8s: U.S. Geological Survey data release, https://doi.org/10.5066/P9V390V2.

Crawford, S., Whelan, G., Infante, D.M., Blackhart, K., Daniel, W.M., Fuller, P.L., Birdsong, T., Wieferich, D.J., McClees-Funinan, R., Stedman, S.M., Herreman, K., and Ruhl, P., 2016, Through a fish's eye—The status of fish habitats in the United States 2015: National Fish Habitat Partnership website, accessed October 29, 2021, at http://assessment.fishhabitat.org.

Daniel, W.M., and Neilson, M.E., 2020, Native ranges of freshwater fishes of North America (ver. 1.0, May 2020): U.S. Geological Survey data release, accessed June 1, 2020, at https://doi.org/10.5066/P9C4N10N.

Elith, J., Leathwick, J.R., and Hastie, T., 2008, A working guide to boosted regression trees: Journal of Animal Ecology, v. 77, no. 4, p. 802–813, accessed December 1, 2019, at https://doi.org/10.1111/j.1365-2656.2008.01390.x.

Fielding, A.H., and Bell, J.F., 1997, A review of methods for the assessment of prediction errors in conservation presence/absence models: Environmental Conservation, v. 24, no. 1, p. 38–49, accessed December 1, 2019, at https://doi.org/10.1017/S0376892997000088.

Frimpong, E.A., Huang, J., and Liang, Y., 2015, Historical stream fish distribution database for the conterminous United States (1950–1990)—IchthyMaps: U.S. Geological Survey data release, accessed December 1, 2019, at http://doi.org/10.5066/F7M32ST8.

Global Biodiversity Information Facility, 2020, Global Biodiversity Information Facility Species Occurrence Downloads: Global Biodiversity Information Facility website, accessed July 18, 2020, at https://www.gbif.org/occurrence.

Guisan, A., and Zimmermann, N.E., 2000, Predictive habitat distribution models in ecology: Ecological Modelling, v. 135, nos. 2–3, p. 147–186, accessed December 1, 2019, at https://doi.org/10.1016/S0304-3800(00)00354-9.

Hastie, T., Tibshirani, R., and Friedman, J., 2009, The elements of statistical learning (2d ed.): New York, Springer, 745 p., accessed December 1, 2019, at https://doi.org/10.1007/978-0-387-84858-7.

Hijmans, R.J., Phillips, S., Leathwick, J., and Elith, J., 2020, dismo—Species distribution modeling, version 1.3-3: The Comprehensive R Archive Network website, accessed November 17, 2020, at http://cran.r-project.org/web/packages/dismo/index.html.

Integrated Taxonomic Information System, 2019, Integrated Taxonomic Information System online database, accessed November 20, 2019, at www.itis.gov. [Also available at https://doi.org/10.5066/F7KH0KBK.]

Liu, C., White, M., and Newell, G., 2011, Measuring and comparing the accuracy of species distribution models with presence-absence data: Ecography, v. 34, no. 2, p. 232–243. [Also available at https://doi.org/10.1111/j.1600-0587.2010.06354.x.]

McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., and Rea, A., 2012, NHDPlus version 2—User guide: U.S. Environmental Protection Agency, Horizon Systems NHDPlus website, 182 p., accessed December 1, 2019, at https://www.nhdplus.com/NHDPlus/NHDPlusV2_data.php.

NatureServe, 2020, NatureServe Explorer [web application]: Arlington, Va., NatureServe, accessed July 13, 2020, at https://explorer.natureserve.org/.

Ross, J.A., Infante, D.M., and Herreman, K., 2022, Anthropogenic disturbances and natural variables in the conterminous United States linked to catchments and buffers of the National Hydrography Dataset Plus version 2.1: U.S. Geological Survey data release, https://doi.org/10.5066/P9PM4HD0.

Swets, J.A., 1988, Measuring the accuracy of diagnostic systems: Science, v. 240, no. 4857, p. 1285–1293. [Also available at https://doi.org/10.1126/science.3287615.]

U.S. Geological Survey [USGS], 2013, Multistate Aquatic Resources Information System (MARIS): U.S. Geological Survey data release, accessed January 15, 2014, at https://doi.org/10.5066/F7BZ641R.

U.S. Geological Survey [USGS], 2020, Protected areas database of the United States (PAD-US) 2.1: U.S. Geological Survey data release, accessed September 15, 2020, at https://doi.org/10.5066/P92QM3NT.

Warren, M.L., Jr., and Burr, B.M., 1994, Status of freshwater fishes of the United States—Overview of an imperiled fauna: Fisheries, v. 19, no. 1, p. 6–18. [Also available at https://afspubs.onlinelibrary.wiley.com/doi/abs/10.1577/1548-8446%281994%29019%3C0006%3ASOFFOT%3E2.0.CO%3B2.]

Wieferich, D.J., McKerrow, A., Cooper, A.R., Yu, H., Ross, J., and Infante, D.M., 2022, Aquatic Gap Analysis Project (Aquatic GAP) aquatic species distribution modeling on the National Hydrography Dataset Plus version 2.1: U.S. Geological Survey data release, accessed December 1, 2019, at https://doi.org/10.5066/P94XM9XV.

Yu, H., Cooper, A.R., and Infante, D.M., 2020, Improving species distribution model predictive accuracy using species abundance—Application with boosted regression trees: Ecological Modelling, v. 432, article 109202, 11 p. [Also available at https://doi.org/10.1016/j.ecolmodel.2020.109202.]

Yu, H., Cooper, A.R., Infante, D.M., and Ross, J., 2022, Fluvial fish native distributions for the conterminous United States using the NHDPlusV2.1 and boosted regression tree models: U.S. Geological Survey data release, https://doi.org/10.5066/P9YX3EX6.

Yu, H., Ross, J., Cooper, A.R., and Infante, D.M., 2022, Presence absence database of fish in the conterminous United States: U.S. Geological Survey data release, https://doi.org/10.5066/P9FZ6J6R.

Appendix 1. Fluvial Fish for Which Insufficient Occurrence Data Were Available to Support Species Distribution Modeling

Table 1.1. Fluvial fish for which insufficient occurrence data were available to support species distribution modeling.

[ITIS TSN, Integrated Taxonomic Information System taxonomic serial number]

Scientific name	Common name	Family	Order	ITIS TSN	Presences	Absences	Prevalence
Acipenser oxyrinchus	Atlantic sturgeon	Acipenseridae	Acipenseriformes	553269	0	4,220	0.0000
Astyanax mexicanus	Mexican tetra	Characidae	Characiformes	162850	14	18	0.4375
Alosa alabamae	Alabama shad	Clupeidae	Clupeiformes	161705	7	1,997	0.0035
Alosa mediocris	Hickory shad	Clupeidae	Clupeiformes	161704	2	2,990	0.0007
Catostomus santaanae	Santa Ana sucker	Catostomidae	Cypriniformes	163912	2	19	0.0952
Cycleptus meridionalis	Southeastern blue sucker	Catostomidae	Cypriniformes	639711	2	328	0.0061
Moxostoma lachneri	Greater jumprock	Catostomidae	Cypriniformes	163942	5	47	0.0962
Xyrauchen texanus	Razorback sucker	Catostomidae	Cypriniformes	163968	5	160	0.0303
Cyprinella callitaenia	Bluestripe shiner	Cyprinidae	Cypriniformes	163774	3	78	0.0370
Cyprinella gibbsi	Tallapoosa shiner	Cyprinidae	Cypriniformes	163784	5	4	0.5556
Gila pandora	Rio Grande chub	Cyprinidae	Cypriniformes	163556	7	11	0.3889
Notropis candidus	Silverside shiner	Cyprinidae	Cypriniformes	163433	2	155	0.0127
Notropis perpallidus	Peppered shiner	Cyprinidae	Cypriniformes	163459	1	223	0.0045
Pogonichthys macrolepidotus	Splittail	Cyprinidae	Cypriniformes	163603	1	70	0.0141
Pteronotropis euryzonus	Broadstripe shiner	Cyprinidae	Cypriniformes	201939	1	9	0.1000
Novumbra hubbsi	Olympic mudminnow	Umbridae	Esociformes	162161	1	114	0.0087
Osmerus mordax	Rainbow smelt	Osmeridae	Osmeriformes	162041	0	3,650	0.0000
Archoplites interruptus	Sacramento perch	Centrarchidae	Perciformes	168175	0	66	0.0000
Micropterus notius	Suwannee bass	Centrarchidae	Perciformes	168164	2	50	0.0385
Micropterus treculii	Guadalupe bass	Centrarchidae	Perciformes	168162	9	91	0.0900
Herichthys cyanoguttatum	Rio Grande cichlid	Cichlidae	Perciformes	649487	5	2	0.7143
Elassoma okefenokee	Okefenokee pygmy sunfish	Elassomatidae	Perciformes	168170	1	208	0.0048
Etheostoma tallapoosae	Tallapoosa darter	Percidae	Perciformes	201996	5	4	0.5556
Oncorhynchus gorbuscha	Pink salmon	Salmonidae	Salmoniformes	161975	1	251	0.0040
Oncorhynchus nerka	Kokanee	Salmonidae	Salmoniformes	161979	1	672	0.0015
Salvelinus malma	Dolly Varden	Salmonidae	Salmoniformes	162000	1	168	0.0059
Salvelinus namaycush	Lake trout	Salmonidae	Salmoniformes	162002	0	3,142	0.0000

Publishing support provided by the Science Publishing Network,
Denver and Reston Publishing Service Centers

For more information concerning the research in this report, contact the

Center Director, USGS Science Analytics and Synthesis Program
P.O. Box 25046, Mail Stop 302
Denver, CO 80225

Or visit the Science Analytics and Synthesis Program website at
https://www.usgs.gov/programs/science-analytics-and-synthesis-sas

Yu and others—**Developing Fluvial Fish Species Distribution Models Across the Conterminous United States—A Framework for Management and Conservation**—SIR 2023–5088

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20235088