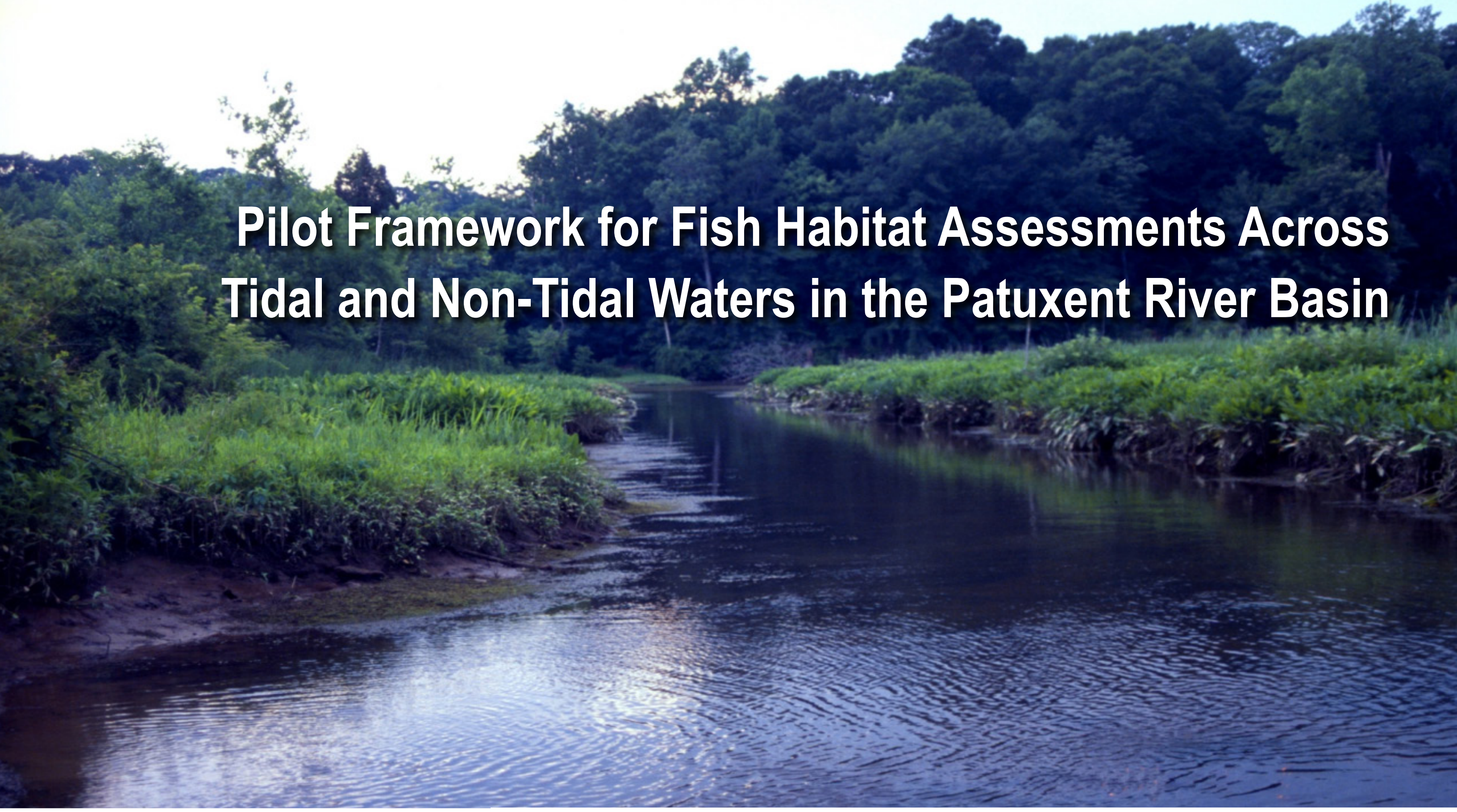

Pilot Framework for Fish Habitat Assessments Across Tidal and Non-Tidal Waters in the Patuxent River Basin

Authors

H. Nisonson,
A. Kiser,
B. Gressler,
A. Leight,
J. Young

July 2024

NOAA TECHNICAL MEMORANDUM NOS NCCOS 332

Citation: Nisonson, H., A. Kiser, B. Gressler, A. Leight, J. Young. 2024. Pilot Framework for Fish Habitat Assessments Across Tidal and Non-Tidal Waters in the Patuxent River Basin. NOAA NOS NCCOS Technical Memorandum 332. Silver Spring, MD. 47 pp.

.

https://doi.org/10.25923/4jqw-mw29

Pilot Framework for Fish Habitat Assessments Across Tidal and Non-Tidal Waters in the Patuxent River Basin

Authors:
H. Nisonson[1], A. Kiser[2], B. Gressler[2], A. Leight[3], J. Young[2]

[1]CSS-Inc,
2750 Prosperity Ave STE 220,
Fairfax, Virginia, USA.
Under NOAA contract.

[2]U.S. Geological Survey
Eastern Ecological Science Center
Kearneysville, WV 25430

[3]Marine Spatial Ecology Division
National Centers for Coastal Ocean Science
National Ocean Service
Cooperative Oxford Lab
Oxford, MD 21654

July 2024

NOAA Technical Memorandum NOS NCCOS 332

United States Department of Commerce — Gina M. Raimondo, Secretary
National Oceanic and Atmospheric Administration — Rick Spinrad, PhD, Under Secretary
National Ocean Service — Nicole LeBoeuf, Assistant Administrator

Table of Contents

List of Figures

List of Tables

List of Acronyms

ACFHP	Atlantic Coast Fish Habitat Partnership
AOT	average optimum threshold
AUC	area under the curve
B-IBI	Benthic Index of Biotic Integrity
BMP	best management practices
CBP	Chesapeake Bay Program
CESR	Comprehensive Evaluation of System Response
CMECS	Coastal and Marine Ecological Classification Standard
COL	Cooperative Oxford Laboratory
CoNED	Coastal National Elevation Database
CONUS	Contiguous United States
CSS	Consolidated Safety Services
CUSP	Continually Updated Shoreline Product
DEM	Digital Elevation Model
EESC	Eastern Ecological Science Center
EPSG	European Petroleum Survey Group
FHAT	Fish Habitat Action Team
GAM	generalized additive models
GIS	Geographic Information Systems
GIT	Goal Implementation Team (within the Chesapeake Bay Program)
HUC-8	8-digit hydrologic unit code
IDW	inverse distance weighted
km	kilometer
LULC	land use land cover
m	meter
MD	Maryland
MD DNR	Maryland Department of Natural Resources
NAD	North American Datum
NCCOS	National Centers for Coastal Ocean Science
NESDM	nested ensemble species distribution modeling
NFHP	National Fish Habitat Partnership
NHD	National Hydrography Dataset
NOAA	National Oceanic and Atmospheric Administration
PAD-US	Protected Areas Database of the United States
POC	probability of occurrence
PRISM	parameter-elevation regressions on independent slopes model
RAM	random access memory
SAV	submerged aquatic vegetation
sstru	shoreline structures
STAC	Science and Technical Advisory Committee (of the Chesapeake Bay Program)
TMDL	total maximum daily load
TSS	true skill statistics
US	United States [of America]
USDM	uncertainty analysis for species distribution modeling
USGS	U.S. Geological Survey
VIMS	Virginia Institute of Marine Science

Executive Summary

As part of the 2014 Chesapeake Bay Watershed Agreement, all Bay States and the District of Columbia have committed to improving the condition of the Bay, which includes a goal to achieve sustainable fisheries. One outcome under that broad goal is improved effectiveness of fish habitat conservation and preservation efforts. In support of that outcome, the U.S. Geological Survey Eastern Ecological Science Center (USGS-EESC) and the National Oceanic and Atmospheric Association's National Centers for Coastal Ocean Science (NOAA-NCCOS) are actively developing datasets, methods, and analyses to conduct fish habitat assessments in the Chesapeake Bay watershed, guided by recommendations from a regional stakeholder workshop held by the Chesapeake Bay Program's (CBP) Fish Habitat Action Team (FHAT) in 2018. The joint USGS and NOAA team has been collaborating on methods for conducting inland and estuarine assessments and exploring whether a seamless headwater to estuary assessment could be developed. The goals of this assessment are to benefit both State and Federal fisheries managers, help advance fisheries science, and provide beneficial information for the public. While past national and regional assessments (e.g. the National Fish Habitat Partnership National Assessment) treated inland and estuarine fish habitat conditions separately due to differences in environments, GIS data representation, and data availability, a seamless habitat assessment could be of value for a broad range of stakeholders as many fish species, several of which are invasive or under federal jurisdiction, use habitats across both inland and estuarine waters. This project developed a pilot framework, explored and tested methods necessary for a finer scale, seamless assessment across both inland and estuarine waters, and demonstrated its use.

Although there was interest by the CBP FHAT for the generation of a Baywide fish habitat assessment that spanned tidal salt, tidal fresh, warm non-tidal and cold non-tidal waters, there are a myriad of implementation details and considerations around conducting a Baywide assessment across all four of these general habitat areas. Therefore, the practical need to conduct a tributary-specific pilot assessment arose. At the beginning of this pilot process, members of the FHAT were presented with a decision matrix to choose a study basin using factors such as data availability and tributary size. FHAT members chose the Patuxent River basin, which has been relatively well sampled and studied. Several spatial frameworks were considered before selection of an inclusive gridded framework for summary and analysis that represented inland drainage networks and landscape influences as well as estuarine bathymetry. A suite of landscape and in-water stressor variables were summarized into the framework and were largely generalized over time. In order to assess the viability of the framework, we chose to use species distribution modeling for each of the species to test the framework's ability to predict habitat use of non-tidal resident, estuarine resident, and migratory species. Tessellated darter (*Etheostoma olmstedi*), American eel (*Anguilla rostrata*), and white perch (*Morone americana*) were chosen as illustrative fish species based on data availability, and differences in life history and habitat use. A nested modeling approach, which involved successive model runs at multiple scales (1000m, 100m, and 10m raster grids) was developed to examine differences in variable importance at different spatial scales and to enhance modeling efficiency. For white perch, a complementary modeling analysis was performed for variables available only in estuarine waters. For all testing, an ensemble modeling approach was conducted, using a suite of potential statistical techniques driven by model strength and variable predictive power. **The statistical testing that we conducted was intended only to test the framework and modeling approach, and not to definitively predict all habitats where specific fish species might be present.** The modeling we conducted to test the framework did have some limitations. For example, the spatial distribution of favorable habitat areas for white perch was likely influenced by the predominance of fish survey locations near the center channel of the river and the use of generalized in-water conditions. For all species, the use of juvenile and adult fish survey data limits the estimation of habitat use to those life stages. Despite such limitations of the data inputs and modeling approach, we found the framework could seamlessly predict fish habitat distribution across freshwater and tidal environments and integrate the influence of landscape stressors with local in-water factors. The developed framework presented to the Sustainable Fisheries Goal Implementation Team (GIT) and FHAT is informative and could potentially be used for other modeling applications in the Chesapeake Bay watershed and elsewhere. In particular the framework and modeling approach lend themselves to evaluating living resource distributions and underlying habitat conditions in shallow tidal waters and beyond, as recommended by the recent Comprehensive Evaluation of System Response (CESR) report from the Chesapeake Bay Program.

1.0 Introduction

Fish habitat assessments attempt to relate past, current, or future landscape conditions to the state of fish species occurrence, distribution, abundance, or community and habitat condition in streams, rivers, or estuaries. Landscape conditions in catchments draining to streams, rivers, and estuaries have long been known to influence water quality and physical habitat conditions in receiving waters (Hynes 1975, Allan 2004). Previous national efforts, such as the National Fish Habitat Partnership (NFHP) National Assessment (Crawford et al., 2016) evaluated landscape conditions that correlated with or were associated with impaired community condition in inland rivers and streams, and separately assessed factors influencing habitat quality in estuaries. Due to the national scale of the effort, Crawford et al. (2016) focused their analysis on landscape factors nationally available datasets only. This potentially limited the scope of predictor data used in statistical models, particularly for the Chesapeake Bay watershed which has a rich collection of datasets developed for the Chesapeake Bay restoration effort.

Scientists from the U.S. Geological Survey (USGS) have been actively developing datasets, methods, and techniques to conduct fish habitat assessments in non-tidal waters of the Chesapeake Bay watershed. This larger research portfolio investigates influences of scale on assessment methods, summary methods for assessing landscape influences on streams and rivers throughout the watershed, and the ability to model species presence/absence or community condition and other biological metrics from landscape-catchment predictors. Similarly, researchers from NOAA's National Centers for Coastal Ocean Science (NOAA-NCCOS) have conducted various types of species, community and habitat assessments, primarily in tidal and estuarine systems. Recently the NOAA team produced a set of recommendations for conducting fish habitat assessments in tidal waters of the Chesapeake Bay (Leight et al., 2021).

The overall Chesapeake Bay fish habitat assessment effort being performed by the NOAA and USGS joint team was initiated for the Chesapeake Bay Program's (CBP) Habitat Goal Implementation Team (GIT) with a Science and Technical Advisory Committee (STAC) scoping meeting, held with stakeholders in 2018. The primary goal of the workshop was to identify factors important for influencing tidal and non-tidal fish habitat assessments including landscape stressors, analysis scale, analysis frameworks, and the best ways of providing information relevant to managers (Hunt et al., 2018). As an outcome of the workshop, USGS and NOAA-NCCOS collaborated to test whether methods for conducting inland and estuarine assessments (respectively) could be developed into a jointly applied and seamless process that would support State and Federal fisheries managers, help advance fisheries science, and provide beneficial information for the public. This study was informed and built upon by the separate non-tidal and tidal assessments described above.

1.1 Joint Pilot Assessment Objectives

The USGS and NOAA collaborated to support the CBP's Sustainable Fisheries GIT by exploring and developing methods for a pilot assessment that would inform managers about the condition of fish habitats in both the estuarine and inland portions of Chesapeake Bay rivers (Fish Habitat Action Team, 2022). The objectives were as follows:

1. Research and develop methodologies and frameworks for conducting integrated headwater to estuary fish habitat assessments using species occurrence data, landscape and hydrologic data, and statistical modeling.
2. Demonstrate testing and application of the developed methodology for a complete river basin within the Chesapeake Bay watershed.
3. Examine potential for meeting stakeholder needs (CBP Sustainable Fisheries GIT and others) for joint inland-estuarine fish habitat assessments through application of coordinated USGS/NOAA science investigations.
4. Develop an assessment of fish habitat for targeted species of interest within the selected pilot basin.

1.2 Course of Action for Joint Pilot Assessment

The following steps were taken to initiate and conduct the pilot assessment:

1. Conduct a review of previous fish habitat assessment methodologies and results.
2. Coordinate with FHAT efforts on stakeholder engagement with interest groups to determine species and issues of interest as it relates to fish habitat assessment.
3. Collate and organize fish (response) data and organize, summarize, and/or generate appropriate landscape and hydrologic predictors, including time-sequenced data.
4. Develop criteria for selecting a tributary of focus, use criteria to propose a tributary (Patuxent, Rappahannock, James, Potomac, Susquehanna, York, Chester, Choptank, Sassafras, or Severn Rivers) and seek concurrence from the FHAT.
5. From (2 and 3) above, develop fish datasets for selected species of interest for our tributary of focus, including headwater resident fish, estuarine resident fish, and diadromous fish that move between headwaters and the estuary.
6. Explore the relationships of various summary frameworks including: 2-D river reach and catchment framework for inland rivers and streams; hexagonal, square, or voxel grids for estuarine habitat areas; and/or hybrid, multi-scale grids for common application for headwaters to estuary summaries.
7. Develop species and habitat distribution modeling applications using maximum entropy (e.g. MaxEnt), machine learning (e.g. Random Forest, boosted regression tree), or ensembles of multiple model methods to predict probability of suitable habitat and/or fish abundance (i.e. conduct assessment).
8. Develop communication products (presentations and reports) to convey results to the FHAT, managers, other scientists, and the public.
9. Publish journal article(s) describing the project objectives, methods, and results.

1.3 Existing Regional Fish Habitat Assessments and the Need for an Integrated Model

In preparation for the Chesapeake Bay Program 2018 STAC workshop (described above), a steering committee reviewed the methods and data used in previous fish habitat assessments, including the 2010 and 2015 NFHP national assessments, the Gulf of Mexico and Pacific Coast regional assessments (Crawford et al., 2016), and the 2015 Chesapeake Bay Habitat Prioritization Tool (Martin, 2015). Several of the preceding fish habitat assessment efforts included the Chesapeake Bay within their geographic extent. For example, the NFHP habitat assessments included (separately) both estuarine and non-tidal waters of the Chesapeake Bay (Esselman, 2011; Greene et al., 2015). These assessments were conducted to better understand how human activities and stressors were impacting fish habitat. The non-tidal assessment developed statistical models of vulnerability for stream and river catchments based on local and upstream catchment summaries of 26 landscape predictors using nationally available datasets at medium spatial resolution (1:100,000 scale) linked to stream lines and catchment polygons of the USGS National Hydrography Dataset (NHD). Habitat vulnerability was evaluated using empirically derived thresholds of fish abundance-stressor relationships (Crawford et al., 2016) and categorized into vulnerability classes (organized by ecoregion). Although limited to nationally available stressor data, every catchment of inland U.S. waters surrounding a stream or river segment (as mapped by the NHD) was assessed using this methodology.

However, the separate estuarine-specific assessment conducted for the NFHP could only broadly generalize habitat conditions in the Chesapeake Bay, since it was limited to national datasets and employed simple, semi-quantitative variable scoring methods based on expert opinion. The 2015 NFHP estuarine assessment used 18 metrics of anthropogenic disturbance, chosen by an expert panel, to evaluate habitat condition (Greene et al., 2015). These variables were combined to form a cumulative disturbance index for each estuarine segment. The assessment described poor habitat in large segments of the tidal Chesapeake Bay, principally due to nutrient concentrations in rural areas (e.g. the Choptank River) and impervious surfaces in urban watersheds (e.g. the Patapsco River). Importantly, the estuarine assessments classified some tributaries as being in very poor condition despite their support of productive spawning grounds.

The recent Atlantic Coast Fish Habitat Mapping and Prioritization Project (ACFHP; Martin et al., 2020) improved on the NFHP estuarine assessment by using a mesh-grid of 1 km hexagons to categorize data for the northeast U.S. estuaries and coastal waters. For estuarine waters, the ACFHP Prioritization incorporated eight variables. These variables were selected by a panel of experts which convened in 2019 and considered known datasets and parameters used in previous assessments. For each of the variables, the expert panel identified a measurement (e.g. percent of land use) that was then evaluated against a cutpoint (e.g. 10% within a given watershed), scored, and scores were added across variables for final rankings (ACFHP report[1]). ACFHP developed this assessment primarily to identify and prioritize focus areas for ACFHP funding opportunities.

Scientists from the USGS have been actively developing datasets, methods, and analysis techniques to conduct fish habitat assessments in non-tidal waters of the Chesapeake Bay watershed. This research investigates influences of scale on assessment methods, summary methods for assessing landscape influences on streams and rivers throughout the watershed, and the ability to model community condition and other biological metrics from landscape-catchment predictors (Maloney et al., 2022). Similarly, researchers from NOAA- NCCOS have conducted assessments of habitat, primarily in tidal and estuarine systems. Recently the NOAA team produced a set of recommendations for conducting fish habitat assessments in tidal waters of the Chesapeake Bay (Leight et al., 2021). This set of recommendations considered the types of data relevant to fish habitat in the Chesapeake Bay, different ways to organize data and different methods for assessing habitat. The NOAA team ultimately recommended the summarization of numerous data types into uniformly sized grid cells. One of the recommended assessment approaches involved the development and projection of a habitat suitability model.

1.4 Selection of the Patuxent

Considerable thought was invested in selecting the Patuxent River Watershed for the pilot assessment. Key factors in the selection included 1) the presence of all four general habitat types (tidal salt, tidal fresh, warm non-tidal, and cold non-tidal waters), 2) the relatively ample amount of accessible fish survey data, 3) a limited number of governmental jurisdictions within the watershed, and 4) a tractable spatial size for a pilot assessment. The USGS and NOAA team developed a decision matrix for all major tributaries of the Chesapeake Bay, including a scoring criteria for each of the factors above. The selection criteria identified the Patuxent, Rappahanock, and James Rivers as the top three candidate tributaries. Members of the FHAT, after being presented with a decision matrix of alternatives, agreed with the assessment team to use the Patuxent River basin as a pilot area for beginning the process of testing and assessing methodologies in conducting a "headwaters to estuary" joint habitat assessment.

1.5 Gathering and Addressing Stakeholder Input

Leading up to the Patuxent River basin pilot, the team received input from stakeholders about conducting a regional Chesapeake Bay fish habitat assessment. We gathered information from members of the Patuxent River Basin Commission, the Maryland Department of Planning, and local residents (via outreach events with the University of Maryland Integration and Application Network (ian.umces.edu)). We also leveraged feedback from stakeholders gathered during an extensive set of targeted interviews with managers and planners (Leight et al., 2019). Although extremely varied in focus and scope, this feedback shared the following elements:

- Use a finer spatial scale than previous assessments;
- Use the rich and diverse fisheries and environmental data available for the Chesapeake Bay;
- Provide a snapshot of current fish habitat status and threats, but then provide insights into conducting a long-term trend analysis;
- Develop a resource to inform local planning and land use decision makers, project designers and implementers; fishery managers and state agencies; and, federal agency project planners and those conducting fish habitat consultations

Originally, there were discussions within the CBP FHAT for the generation of a Baywide fish habitat assessment that spanned tidal salt, tidal fresh, warm non-tidal, and cold non-tidal waters. However, since there are a myriad of implementation details and considerations around conducting a Baywide assessment across all four of those general habitat areas, the practical need to conduct a pilot assessment arose. The Patuxent River pilot was designed and conducted with the intention of testing the implementation viability of these elements for eventual Baywide application.

Breakout Box: Patuxent River Commission Annual Action Plan Items

The Patuxent River Commission has previously identified priorities in their Annual Action Plans (Patuxent River Commission, 2019) that included the following requested activities that may overlap this assessment:

- "Request DNR [Department of Natural Resources] or appropriate academic group review the status and trends of commercially, recreationally and ecologically important finfish and shellfish species in the Patuxent River"
- "Preserve and restore movement of water, fish and wildlife through identifying and removing barriers."
- "Establish a workgroup to review the potential for creating a Patuxent River Watershed geodatabase that would be used for educational and research efforts."
- "Develop a Patuxent River report card that can be used to assess water and living resources conditions in each major part of the river and its tributaries, both tidal and nontidal."
- "Ensure and encourage public access to the river, its tributaries, and recreational opportunities within the watershed."

2.0 Patuxent Pilot Framework Approach

2.1 Framework Dilemma

As mentioned above, previous habitat assessments for the Chesapeake Bay have been conducted by USGS and NOAA. For both tidal and non-tidal assessments, land-based environmental factors that may impact in-water fish habitat and fish populations have typically been summarized within different watershed sizes or based on proximity to the water. However, up to this point, both in-water variables and fish survey information have been summarized differently between tidal and non-tidal assessments. For non-tidal assessments, such as that conducted by USGS (Maloney et al., 2022), in-water habitat condition was inferred from models relating field-sampled fish communities or derived biotic condition metrics to landscape variables summarized around linear stretches of river defined by various versions of the National Hydrography Dataset (NHD; About National Hydrography Products | U.S. Geological Survey[2]). In the NHD system, catchments are associated with individual stream/river reaches, allowing direct attribution of land-based factors to stream reaches. Improvements in the spatial resolution of the NHD (i.e. from 1:100,000 to 1:24,000 scale) have allowed for improved analysis of the relationship between catchment and stream reach conditions. In contrast, most tidal assessments employ the use of two- or three-dimensional areas of the estuarine waterbody for representing in-water factors. For the 2015 NFHP Estuarine Assessment (Greene et al., 2015), the tidal portions of the Chesapeake Bay were divided into approximately 20 polygons which represented large portions of tidal rivers and the mainstem of the Bay. The ACFHP Assessment used a hexagon framework, with uniform 1 km diameter hexagons (Martin et al., 2020). Additionally, in 2021, the NOAA team developed recommendations for conducting fish habitat assessments in tidal waters of the Bay, including the use of a contiguous grid-pattern of hexagons, similar to the ACFHP assessment (Leight et al., 2021).

2.2 Framework Selection and Design

The Joint Pilot team considered all of the previous analytical frameworks described above for summarizing and analyzing data over the entire Patuxent River watershed. The linear nature of the NHD presented challenges for summarizing data in the broad extent of the river's estuarine mainstem. Conversely, hexagonal, gridded frameworks tended to be at scales too large for summarizing data in small headwater streams. Another potential solution, commonly employed for hydrodynamic models, consists of a gridded framework with variable cell shape and/or size. As highlighted in Leight et al. (2021), this design presents some unique challenges and potential biases in summarizing spatial data to variable sized polygons.

The team decided to use a 2-dimensional raster framework with uniformly sized square grid cells. However, since ideal sizing of the raster grids was not known and likely differs by species, we tested the framework under a multi-scale approach that captured coarse, medium, and fine resolution influences (represented in this study by 1000m, 100m, and 10m raster cells, respectively). The use of a continuous raster grid layer over the entire HUC-8 watershed allowed the application of several peer-reviewed data summary approaches, such as Flow-Condition Parameter Grids[3] (Barnhart et al., 2021) and Inverse Distance Weighted Accumulation (Peterson and Pearse, 2017). The spatial resolutions used in this analysis were chosen for several reasons, chiefly among them, they match previous studies and data sources allowing for greater comparison across habitat studies. Secondly, the ideal spatial size is not known for these species, and likely varies between the species. As such, this methodology allows for an initial starting point in assessment of species relevant spatial frameworks. Further studies building off of these may continue to test additional spatial resolutions to find optimum fits for each species.

2.3 Species Tested

Previous efforts have synthesized or characterized fisheries independent fish survey data for non-tidal waters (Krause and Maloney, 2021) and tidal waters (Tetra Tech, 2020) in the Chesapeake Bay. Using this information, the Joint Pilot Team decided to select a few species, representing different habitat associations in the Bay, to test the gridded-framework and associated analytical methods**. The statistical testing that we conducted was intended only to test the framework and modeling approach, and not to definitively predict all habitats where specific fish species might be presen**t. In order to test the full extent of the framework, stretching from headwaters to river mouth, the team sought out representative species from non-tidal resident, diadromous, and estuarine assemblages. Within those groups, the team selected species with relatively robust observational data from monitoring surveys, ultimately deciding to focus on tessellated darter (*Etheostoma olmstedi*; freshwater species), American eel (*Anguilla rostrata*; diadromous species), and white perch (*Morone americana*; estuarine species). The spatial distribution of sites marked as species presence for white perch, American eel, and tessellated darter in **Figure 1** below uniquely summarizes the differential distributions for each species. Additional information on fish collection data is provided in section 3.2.10 and metadata for the surveys used in this study can be found in **Appendix Table A1**. Due to temporal, spatial, and methodological differences between sampling studies used to obtain occurrence data, it was difficult to identify true absences within the entire basin extent. Therefore we calculated pseudo-absences to use during the modeling process using a statistical approach (see Section 3.3).

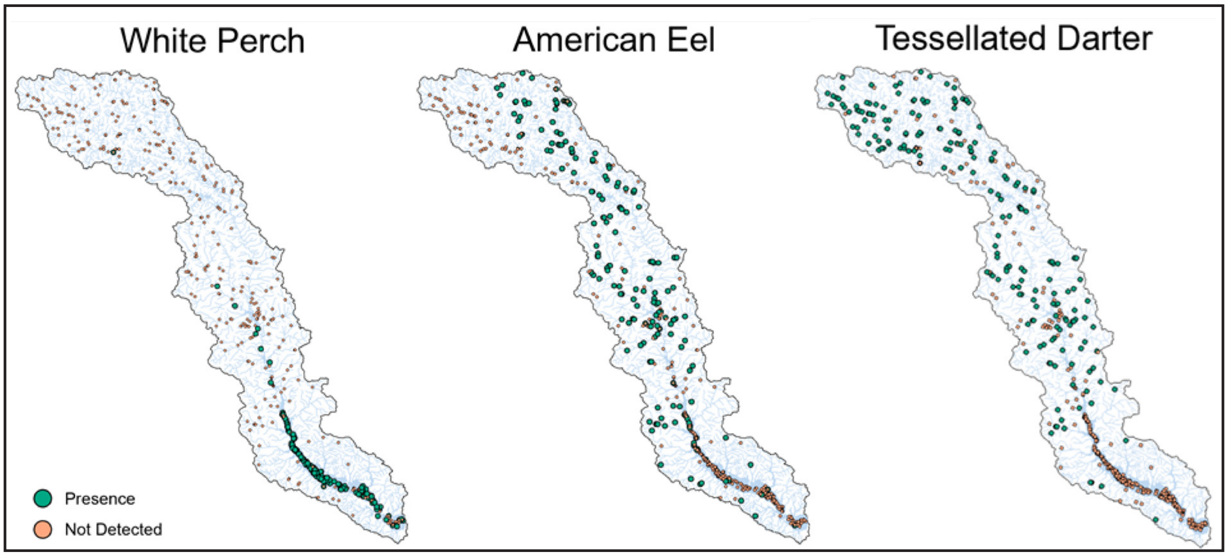

Figure 1. Spatial distribution of presence and not detected sites from the compiled dataset provided in **Appendix Table A1** for white perch, American eel, and tessellated darter within the Patuxent River HUC-8 watershed (Buto and Anderson, 2020).

2.4 Habitat Considerations across Species

A gridded rasterized framework is a flexible method for representing aspects of the environment that play a role in structuring and/or influencing fish habitat. In contrast to vector stream reach and catchment representations (typically used in NHD-based frameworks), additional factors can be extracted from gridded raster representations including distance and proximity metrics, overland flow and runoff characteristics, and surface characteristics (roughness, slope, shape) among other influences at a variety of spatial resolutions. These characteristics are particularly well-suited for habitat assessment and modeling because they represent environmental variables in a spatially explicit manner. In the context of fish habitat assessments, various factors such as water temperature, substrate type, vegetation cover, and water depth may play crucial roles in determining the suitability of an area for different fish species. By dividing the aquatic environment into a grid of cells of different scales, each with assigned values for these environmental variables, a gridded rasterized framework allows for a detailed characterization of in-water habitat, especially for the wider estuarine waterbody, beyond the vector accumulation approach commonly used for the linear NHD streams. This raster approach can facilitate the identification and characterization of preferred habitats for different fish species across the entire waterbody. Furthermore, spatial analyses within this framework enables the assessment of habitat connectivity, the impact of anthropogenic activities, and the identification of priority areas for conservation efforts. The use of such a framework congruent with GIS-based tools, coupled with species distribution modeling in the R statistical package[4], can provide a holistic and spatially explicit approach to fish habitat assessments, eventually aiding resource managers and conservationists in making informed decisions for the sustainable management of fisheries for the entirety of the Chesapeake Bay.

3.0 Framework in Action

3.1 Gridded Framework Construction

The collaborative research efforts between USGS and NOAA teams involved considerable thought and exploration of diverse habitat summary frameworks: including a 2-D river reach and catchment model tailored for inland hydrological systems and hexagonal, square, or voxel grids optimized for estuarine habitats. Following comprehensive consideration of these options, the teams landed on a hybrid, multi-scale gridded rasterization approach. This method, deemed suitable for summarizing data from headwaters to estuary across various spatial scales, was driven by considerations such as data availability, the efficacy of geoprocessing tools, and interpolation methods capable of accurately representing land-based variables within the watershed. The selected methodology not only met the criteria of robustness but also facilitated the comparison of disparate output results based on varying spatial scales.

3.1.1 Framework Design

The team developed the raster framework using the USGS Coastal National Elevation Database (CoNED; Coastal National Elevation Database (CoNED) Applications Project | U.S. Geological Survey[5]). This dataset is an integration of the best available bathymetric data (sonar, soundings) and high resolution topography (from lidar), and is obtainable for selected coastal regions in the United States, including the Chesapeake Bay and Delaware Bay. We chose this dataset as the basis for our framework as it seamlessly integrated upland topography with bathymetry, allowing flexible data summary techniques incorporating upstream flow and runoff influences as well as bottom characterization. It should be noted that accurate bathymetric sources were only included in CoNED for tidal portions of the study area. Non-tidal river environments are represented as river channels, or “hydro-flattened” features. While the original resolution of the CoNED data is a 1m raster, we ultimately decided to assess habitat relationships at three spatial scales - square raster cells with sides measuring 1000m, 100m, and 10m - requiring resampling of the original dataset (**Figure 2**).

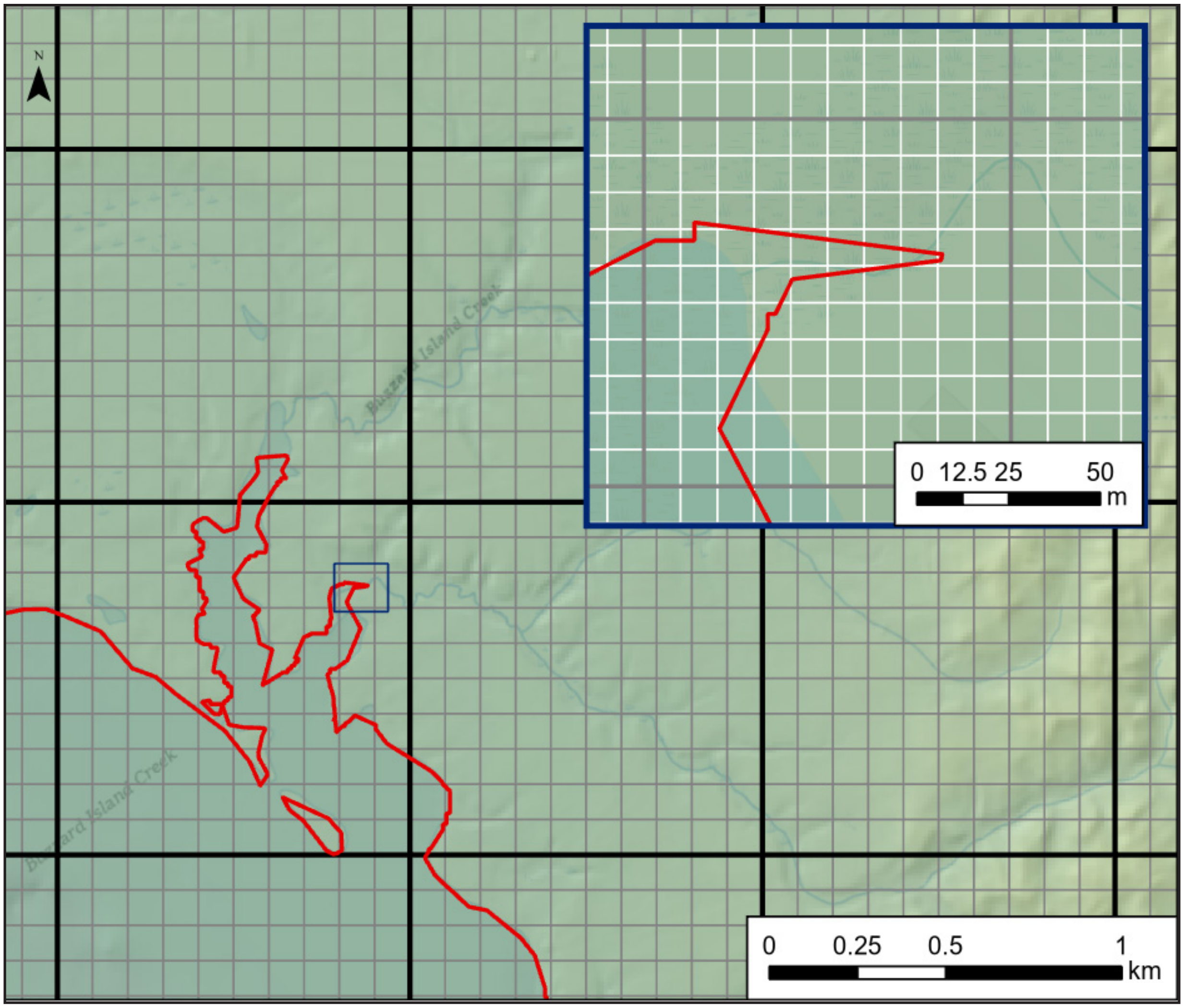

Figure 2. Multiscale gridded network generated for this project. Black lines indicate 1000m grid and dark gray lines indicate 100m grid, white lines in the inset indicate the 10m grids. Light blue lines represent the NHD flowlines. Red lines indicate the boundary used in the tidal-bound analysis. World Topo Basemap was used to underlie the gridded network[6].

3.1.2 Framework, Watershed, and Waterbody Extents

As indicated earlier, the gridded framework extended throughout the USGS Patuxent River watershed boundary area. However, a separate analysis was conducted exclusively for white perch throughout the tidal waters. The extent of the estuarine/riverine waterbody, in contrast to the HUC-8 watershed, was more challenging to define. Ultimately we developed a waterbody polygon based on a combination of data from the NHD flow-network ("swnet.tif", a 10m raster representation of the NHD surface water flowline network), the NOAA Continually Updated Shoreline Product (CUSP; Continually Updated Shoreline Product[7]), and satellite imagery. For determining the extent of tidal waters, we used the upper extent of the Chesapeake Bay Interpolator cells (Bahner et al., 2001), that were used for water quality interpolation. This shoreline extent was used for calculations of water quality and distance to shoreline types and submerged aquatic vegetation (SAV) throughout the waterbody. However, in order to include benthic habitat structure, which was available at a smaller extent, we then had to adjust the waterbody extent to areas for which benthic substrate data were available.

3.2 Data Assembly

All data were standardized to the same spatial coordinate reference system (CONUS Albers NAD83, EPSG 6350). Environmental data were compiled from a wide range of sources, leveraging a metadata catalog (USGS Chesapeake Bay Studies - Data Catalog | U.S. Geological Survey[8]) assembled by the team specifically for Chesapeake Bay fish habitat assessments. The team attempted to collect the most recent data available, though various data sources extended over the last 30 years. Because many variables were not collected consistently over that time period, we choose to "flatten" the data over time and conduct the analysis as a snapshot in time rather than an assessment of change in habitat over time.

In order to test the framework and the modeling approach, a broad suite of predictor data for the watershed was included in the assessment to include variables of different spatial formats and to allow for a data-driven selection of factors based on data mining and statistical analysis, rather than a predetermined selection by the team members[9]. With this in mind, our team took into consideration many environmentally driven factors previously used in fish habitat assessments (illustrated in **Table A2** within the Appendix) as well as a variety of topographic and DEM based variables used to analyze finer upstream hydrodynamic processes. In order to estimate the potential influence of land-based factors into the waterbody, estimates of factor "accumulation" moving downstream from headwaters were employed. These techniques include flow accumulation and flow direction approaches using a DEM (Jenson and Domingue, 1988) and distance-based approaches using inverse distance weighting (Peterson and Pearse, 2017).

3.2.1 DEM Based Variables

Digital Elevation Model (DEM) data came from the USGS Coastal National Elevation Data product (CoNED; Danielson et al., 2018). DEM based metrics were used to identify heterogeneity and homogeneity of the landscape by computing terrain indexes, roughness, and slope (Hijmans et al., 2022). In addition, DEM metrics also generate "geomorphons" (Jasiewicz & Stepinski, 2013) that identify general landform characteristics into 10 defined classes (peak, ridge, shoulder, spur, slope, hollow, footslope, valley, pit, and flat, see **Figure 3** for reference). Percentage of each geomorphon type was calculated for the gridded cells at all resolutions (i.e. % area flat within 100m cell). These DEM metrics were used to create flow accumulation metrics helping identify watersheds and flow paths based on topography.

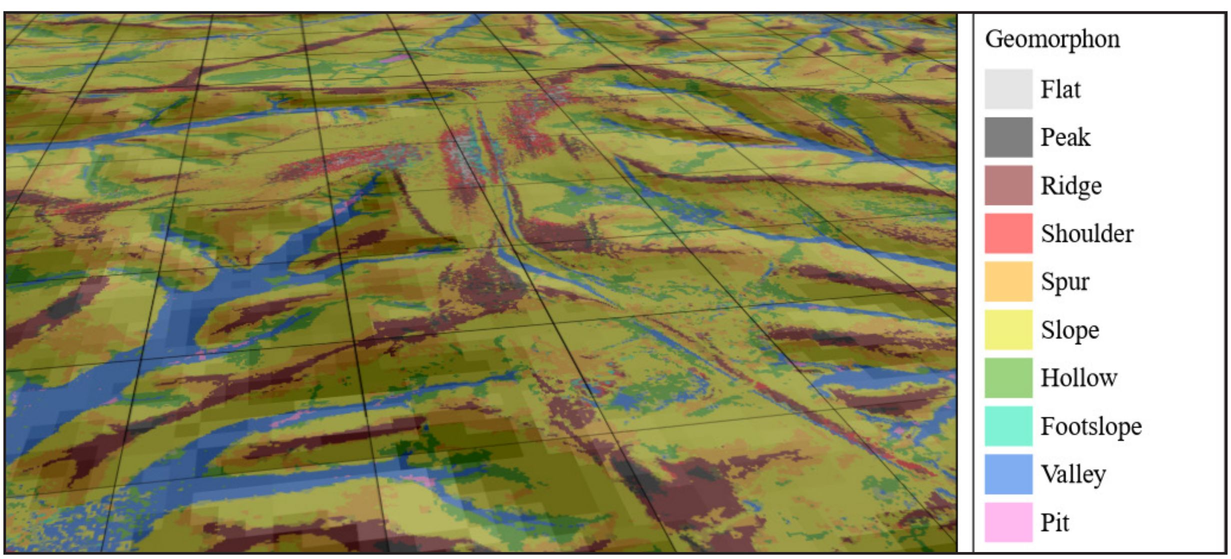

Figure 3. Geomorphon landscape representations created from 1m DEM (Danielson et al., 2018). Black boxes represent the 100m gridded nets used to extract land use/land cover (LULC) and geomorphon data. Values were calculated as the percent of each geomorphon (or LULC) within the grid cell (e.g., 30% spur, 12% cropland, etc.).

3.2.2 Land Use/Land Cover

The team used land use and land cover (LULC) data from the 2013-2014 Chesapeake Conservancy (Chesapeake Bay Program Land Use/Land Cover Data Project[10]). This data set represents land classes at a 1m scale resolution that were, temporally overlapped with the majority of our species occurrence data. The original data contain 49 classes of land cover/land use (see **Appendix Table A3**) that were combined into 22 LULC classifications for the purpose of this study. For example, the original classes of “impervious structure” and “other impervious” were combined into the new general class of “impervious”. LULC metrics were computed as the percentage of the LULC type within the cell (i.e. 40% of a grid cell was cropland). In addition to percent land cover maps, using the derived DEM metrics, we also included several metrics that described the aggregate LULC influence of the watershed respective to each grid cell. This was done by combining the LULC with the flow accumulation metrics to estimate total LULC influences to an area and the percent of the total watershed potentially influenced by a LULC type. This was done by running an additional flow accumulation calculation for each LULC type. The flow accumulation tool then sums the LULC type rather than counting cells.

3.2.3 Climate Data

Climate data were obtained from the PRISM Climate Group (Oregon State University, PRISM Climate Data[11]) as monthly 30-year normalized datasets for air temperature minimum and maximum, and precipitation at an 800m resolution. Climate data were resampled to match our three study resolutions (1000m, 100m, and 10m) using the bilinear sampling method in ArcGIS Pro version 3.2.1 (Esri Inc., Redlands, CA). Using these datasets, 19 bioclimatic variables were created utilizing the biovars function within the R Dismo package (Hijman et al., 2017). The bioclimatic variables describe seasonal events on both monthly and quarterly scales to identify general climatic trends over the region and have been shown to be useful in predicting species distributions (Kiser et al., 2022). It should be noted that resampling the PRISM data to work within our raster framework scales does not create new data and that there may be implied precision of the climate data at the finest resolution that is not necessarily warranted. However, because climate data are not generally available at very fine resolution, we used the data best available to us.

3.2.4 Surface Water Quality

Water quality conditions were estimated exclusively for the tidal portion of the watershed using data from the Maryland Department of Natural Resources’ “Eyes on the Bay” monitoring program[12] and the Chesapeake Bay Interpolator (Bahner, 2001). We selected three years of monitoring data (2004, 2012, and 2018), to represent a range of climatic conditions, based on U.S. Climate Division data (Vose et al., 2014). Compared to the average conditions over 2000-2022, precipitation in 2004 was near the 23-year average, and air temperature was just slightly below the average. In contrast, 2012 was a warm and dry year. Air temperatures in 2018 were fairly normal, but precipitation was the highest on record for the Chesapeake Bay region (see **Figure 4**).

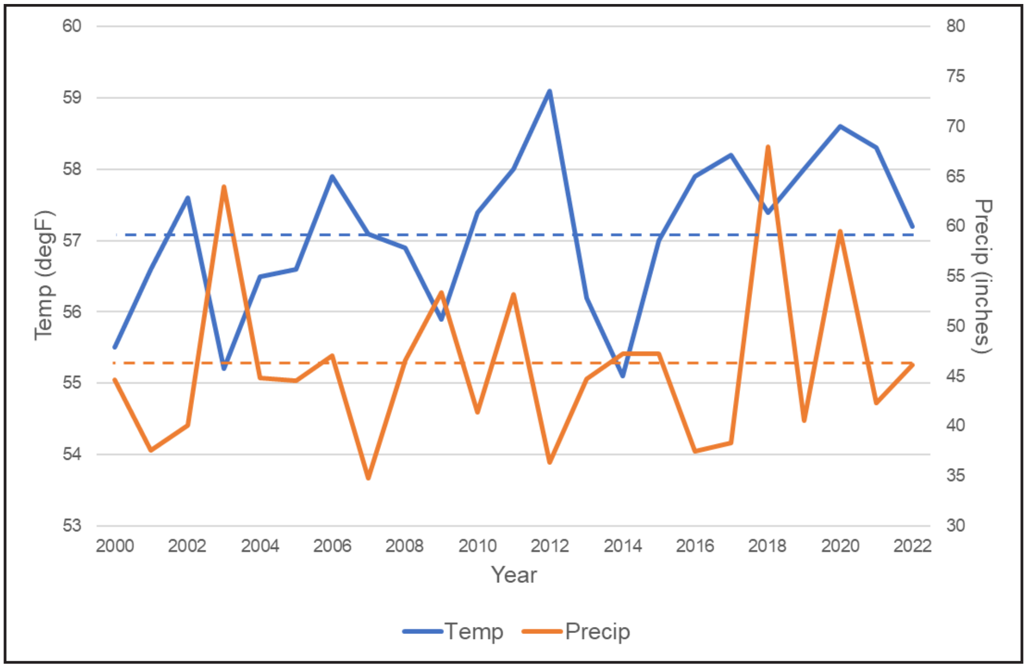

Figure 4. Annual average air temperature (solid blue line) and precipitation levels (solid orange line) for Climate Division 'Maryland - Upper Southern (Vose et al., 2014)', compared to the 23 average (dotted lines) for years 2000-2022.

Water temperature, dissolved oxygen, and salinity data from the monitoring stations were interpolated for the Patuxent tidal river for each sampling date (Richard Tian, Chesapeake Bay Program, unpublished data, 2023)[13]. We then averaged the already interpolated water quality metrics between the three years (2004, 2012, and 2018) for each interpolator cell (see **Figure 5** for a visual reference of the CBP interpolator grid cell framework). Each cell contains multiple depth values for each water quality metric within that given water column. We consolidated these to portray three different depth values: surface, mid-water, and deep with the surface level containing the top 3m of water from the interpolator data set. The surface values for seasonal water quality layers were spatially joined to the nearest 10m raster grid cell and resampled bilinearly to the 100m and 1000m scales using the resample[14] tool in ArcGIS Pro version 3.2.1. Because the biological response data used in this study consisted mainly of data taken from surveys conducted in late spring to early fall, and because mid-water to deep depth layers did not extend throughout the entire tidal waterbody, only summer (June-August) surface estimates were used. As explained in Section 3.4 below, the water quality data were only used for testing our tidal species (white perch) in the non-nested ensemble modeling approach.

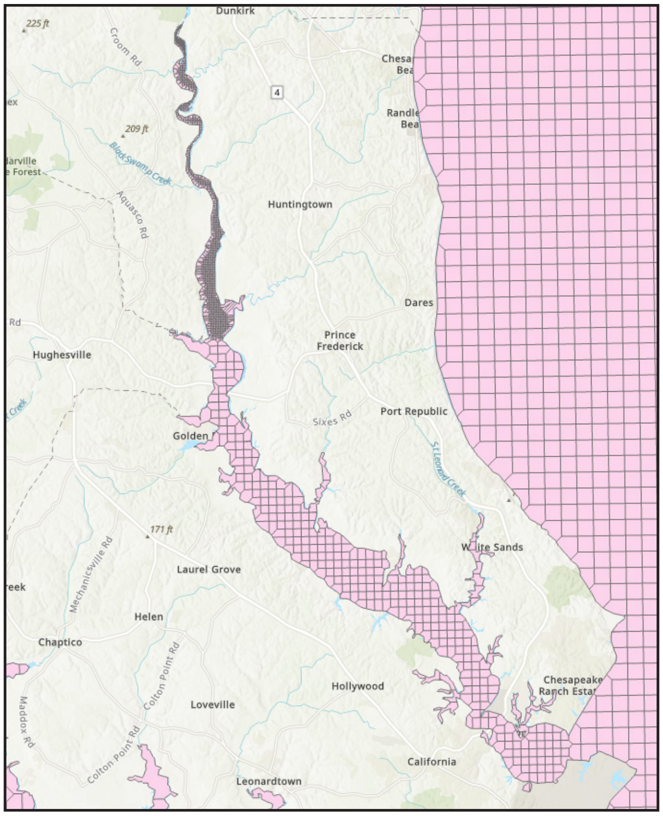

Figure 5. Chesapeake Bay Interpolator grid (Bahner, 2001) overlaying the Patuxent estuary with World Topo Map as the basemap[15]. Each cell represents a water column with 1m depth layers. Due to the small cell size in the upper river, only the gray cell outlines can be seen.

3.2.5 Substrate Bottom Type

For the tidal-bound analysis and model, benthic habitat data were included. This data layer was previously aggregated and curated by NOAA's Chesapeake Bay Office and standardized to the Coastal and Marine Ecological Classification Standard (CMECS; Coastal and Marine Ecological Classification Standard (CMECS)[16]. Polygons of the various habitat types were transformed into a 10m gridded raster layer, and separated into layers of individual habitat types using a ArcPy script in ArcGIS Pro version 3.2.1 intended to binarize a multi-faceted categorical layer. Nine habitat types were represented within the tidal portion of the Patuxent watershed and subsequently used in the non-nested ensemble model: mud, sand, biogenic oyster rubble, biogenic oyster reef, sandy mud, muddy sand, anthropogenic shell rubble, gravel mixes, and unclassified. At the 10m scale, each substrate bottom type was represented as a binary raster layer with values of 1 where the habitat type occurs and values of 0 where that habitat type is not present. For scales of 100m and 1000m, the binary rasters were resampled to become percentage cover of that given habitat type within the 100m^2 and 1000m^2 raster cells. This was done using the tabulate area[17] and calculate field[18] tools in ArcGIS Pro version 3.2.1 to find the percentage of a given habitat type within a given grid cell.

3.2.6 Submerged Aquatic Vegetation

Distance to SAV beds was assessed using a distance accumulation[19] geoprocessing tool in ArcGIS Pro version 3.2.1. The data collected for SAV spanned the years 1999 to 2020 and data collection efforts were conducted at the Virginia Institute of Marine Science (VIMS; SAV Reports & Data | William & Mary[20]). In this analysis, the DEM served as the surface raster input, with the identified waterbody (described above in section 3.1.2) serving as a barrier input into the ArcGIS Pro version 3.2.1 tool. Geodesic distance to the combined masses of SAV beds was calculated at a 10m scale from 1999-2020 and subsequently resampled using a bilinear technique to achieve distance metrics at resolutions of 100m and 1000m with the resample tool in ArcGIS Pro version 3.2.1. Distance to SAV beds was calculated within the bounds of the tidal waterbody and therefore only used in the non-nested tidal model runs for white perch.

3.2.7 Hardened Shoreline

The assessment of distance to hardened shoreline also involved the application of the distance accumulation geoprocessing tool in ArcGIS Pro version 3.2.1. A hardened shoreline polyline layer was constructed using the most recent GIS shoreline inventory data release for each county within the tidal portion of the Patuxent watershed. These data are hosted and collected by VIMS (Shoreline & Tidal Marsh Inventory | William & Mary[21]). Counties included along with the year of most up to date GIS data are as follows: Anne Arundel (2020), Calvert (2020), Charles (2022), Prince Georges (2023), and St. Marys (2022). The *sstru* (shoreline structures) shapefile within each county's GIS package is a linear shapefile delineating hard structures at the shoreline (bulkhead, breakwater, dilapidated bulkhead, debris, jetty, marina, marsh toe revetment, unconventional, riprap, and groin fields). In this analysis, the DEM was utilized as the surface raster, while the waterbody boundary (described in section 3.1.2) served as a barrier within the distance accumulation geotool environment. Geodesic distance was calculated from the nearest merged hardened *sstru* polyline layer at the 10m scale and grid values were resampled bilinearly to generate distances to hardened shoreline at both 100m and 1000m scales. This layer was only utilized for the non-nested tidal model runs for our largely tidal species white perch.

3.2.8 Protected Areas

The evaluation of distance to protected areas employed the distance accumulation geoprocessing tool, where no specific barrier was considered. Proximity to protected areas was also exclusively used for the tidal-bounded white perch model runs. The Protected Area Database of the United States (PAD-US) data was utilized for this assessment, and the relevant information for the state of Maryland was downloaded from the U.S. Geological Survey website (USGS; PAD-US Data Download | U.S. Geological Survey[22]). This database encompasses mostly public lands owned in fee (the owner of the property has full and irrevocable ownership of the land); however, permanent and long-term easements, leases, agreements, Congressional (e.g. 'Wilderness Area'), Executive (e.g. 'National Monument'), and administrative designations (e.g. 'Area of Critical Environmental Concern') documented in agency management plans are also included (USGS GAP, 2022). In this analysis, the DEM served as the surface raster, and geodesic distance calculations were applied at a 10m resolution scale. The data were resampled using a bilinear technique, resulting in distance values at both 100m and 1000m scales.

3.2.9 Benthic IBI

The Chesapeake Bay Benthic Monitoring Program, conducted by Versar, Inc. (Dulles, VA), has been a part of Maryland's Water Quality Monitoring Program for the Chesapeake Bay since 1984 (Versar; Chesapeake Bay Benthic Monitoring Program[23]). Data for benthic monitoring sampling events, sample collection, and Benthic Index of Biotic Integrity (B-IBI) were downloaded as comma-separated text files (csv) for years 1999 to 2021 and systematically organized in R by sampling events and the fixed station replicate-averaged B-IBI score at each sample collection site. The average B-IBI score served as a key metric at each site, providing a comprehensive assessment of benthic health in that location. To facilitate spatial analysis for the estuarine white perch model, the average B-IBI scores for each survey site were linked to the nearest 10m raster grid cell and were bilinearly resampled, resulting in representations at both 100m and 1000m scales. This comprehensive approach allows for a nuanced exploration of benthic conditions in the Chesapeake Bay region over the specified timeframe. We attempted to combine the tidal waters B-IBI data to the nontidal Chesapeake basin-wide index of biotic integrity for stream macroinvertebrates (e.g. " Chessie B-IBI") but the scoring metrics between the two indices were not compatible. Therefore, B-IBI was utilized exclusively for the tidal-bound non-nested ensemble modeling approach for white perch.

3.2.10 Fish Data

Because survey data were collected using different sampling gear and different sampling techniques, we decided to restrict our analysis to survey-determined presence data and pseudo-absences (see **Appendix Table A1** for a summarization of biological sampling surveys used for our area of focus). For each of the three focal species, if the species was ever captured by the sampling method and reported for the survey, then we included data from that survey as part of our study. The surveys included in this study focused on juvenile and adult fish. As such, this study did not test the modeling approach for early life stages (eggs and larvae) of these species. Tessellated darters and white perch were found primarily in non-tidal and tidal waters, respectively. Therefore, the surveys used for presence data for those two species were specific to those parts of the river. However, American eel were present in some samples from both tidal and non-tidal waters. Because American eels are migratory and this study did not include larval data, the observations of eels included in this study were likely both residential and migrating juveniles or adults. To account for this instance, survey data from both tidal and non-tidal waters were merged based on standardized species names while also maintaining presence occurrence within a given survey.

3.3 Methods and Initial Results for a Basin-wide Nested Ensemble Modeling Approach

The USGS-NOAA team then tested the framework as a platform for statistical modeling. Recent fish habitat assessments have employed a range of different statistical and non-statistical approaches for exploring relationships between habitat and fish occurrence and/or fish abundance. Statistical tests range from various forms of generalized linear models (e.g. linear and logistic regressions), general additive models (a.k.a. GAMs), and tree-based models, such as boosted regression trees and Random Forest analysis. For context, the non-tidal assessment conducted by USGS (Maloney et al., 2022) employed Random Forest modeling, while the habitat suitability model recommended by Leight et al. (2021) for tidal waters relied on linear regressions. Instead of focusing on one statistical method, we employed an ensemble approach that utilized the strength of multiple predictive models and reduced individual model bias for predicting fish habitat suitability based on environmental characteristics. This approach has been used by Kiser et al. (2022) and others. All analyses were conducted in R using the ‘usdm’ and ‘sdm’ packages (Naimi et al., 2014; Naimi and Araújo, 2016).

Due to the large number of environmental parameters being considered, we first tested predictor variables for collinearity (Pearson’s Correlation ≥ 0.8). Within each collinear pair, the variable with the greater variance inflation factor was excluded from the species distribution modeling. While this could potentially remove mechanistically important variables, this method was better suited to this study due to the number of species considered, their diverse interactions with their environments between varied life histories, and incomplete/varied knowledge of the species interactions with all variables presented in this study. As such, the unbiased mathematical solution presented here is more appropriate than potentially excluding variables based on incomplete knowledge. Provided more information on specific species interactions, variable selection could be refined.

For each of our test species, an ensemble model of variable importance was developed using fish presence and pseudo-absence data derived from all the variables that passed the test for collinearity. Variables included in the nested ensemble models for each species at a given scale can be seen in **Appendix Table A4**. Due to temporal, spatial, and methodological differences between sampling studies used to obtain occurrence data, it is difficult to identify true absences within the entire basin extent. Thus, we created random pseudo-absence points at two times the number of presence points (Barbet-Massin et al., 2012; Liu et al., 2013), distributed throughout the basin for each species.

3.3.1 Nested Ensemble Species Distribution Modeling Approach

In order to explore spatial relationships of environmental variables at specific scales, and the different impacts our variable may have on a species distribution across spatial scales, the team created a Nested Ensemble Species Distribution Modeling (NESDM) approach utilizing all three raster cell resolutions: 1000m, 100m, and 10m. The NESDM is based on hierarchical niche theory (Frissell et al., 1986; Poff, 1997) that states a species’ distribution is defined by multi-resolution forces from larger scale climate and geological factors to meso- and micro-habitat influences at the local scale; each layer acting as a sieve or filter excluding a species from an area based on life history traits and adaptive ability. Under this general framework, we used our multi-resolution study to first identify species distribution at the lowest resolution (1000m). Using this identified habitat as the environmental extent for the next lowest resolution (100m), we repeated our correlation and modeling analysis at the new extent to identify suitable habitat at a finer resolution. This continued to our highest resolution (10m), identifying progressively smaller regions of suitability based on the environmental variables at each scale before it. This study design allows us to both decrease the processing needs and time by limiting the spatial extent at which the models are applied at higher resolutions, and more importantly allows us to identify the influence of environmental characteristics at individual scales. We acknowledge that we are potentially limiting areas that would be predicted as suitable at finer scales. However, with the exception of tessellated darter at the 10m resolution (84%), greater than 90% of all presence records were included at each scale for all species. The benefit of this nested iteration allows for a more informed decision-making process regarding spatial scales for future environmental studies and identifying potential management priorities based on varied spatial resolutions. This method is also beneficial in the scope of expanding our framework to the full Chesapeake Bay and the need for an efficient and targeted Baywide model that is able to quickly identify areas of suitable habitat of resident fish species.

For each individual fish species, a multitude of species distribution model methods (21; see **Appendix Table A5**) were evaluated in a preliminary step and tested to ensure accuracy based on thresholds (area under the curve (AUC) > 0.7 and true skill statistics (TSS) > 0.6, Kiser et al., 2022). Selected models were then bootstrapped and combined into an ensemble model based on the weighted average (maximized [sensitivity + specificity]). This was then used to create probability of occurrence (POC) based on habitat suitability (see **Figure 6** for full layout of the NESDM workflow). The POC maps were then used to identify areas of predicted presence based on the average optimum threshold (AOT; the POC above which would be predicted as suitable). The identified presence areas were then used as the spatial extent of the next finer resolution data analysis and modeling steps until the highest resolution was reached (in this framework study, until we obtained outputs for a 10m resolution).

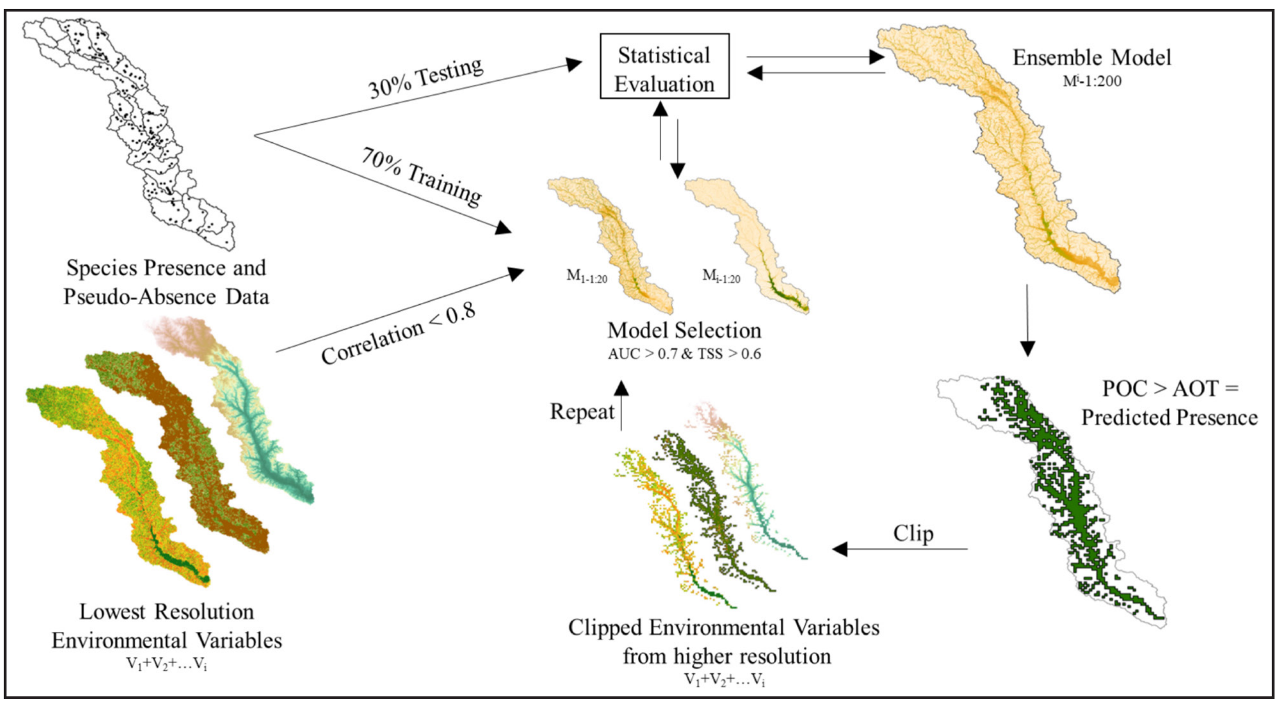

Figure 6. Conceptual model for the Nested Ensemble Species Distribution Model (NESDM) for the basin-wide modeling analysis.

Species occurrence and pseudo-absence data are split into 70% training and 30% testing datasets. Preliminary tests using all model methods (21) are run using 20 bootstrap repetitions and 20 replicates of random resamples. Models that met our inclusion criteria (AUC > 0.7 and/or TSS > 0.6) were included in the final ensemble model run where bootstrap and replicates were increased to 200. Individual model results were then assembled based on a weighted average that maximized sensitivity and specificity. The area of predicted presence was then determined where POC exceeded the AOT at that scale. These selected areas were then used as the spatial extent for the next higher resolution of predictor variables. This method was repeated until the highest spatial resolution was reached.

3.3.2 Preliminary Results for the Basin-wide Nested Ensemble Model

Using data for the entire watershed, nested ensemble model outputs were generated for each species (American eel, tessellated darters, and white perch) at the 1000m, 100m, and 10m raster scales to assess spatial distribution patterns. The variables of importance were derived from a data driven automated process rather than predetermined habitat associations. It is important to emphasize that these results are based on a limited portion of the species' geographic range, and caution must be exercised in drawing explicit conclusions or making connections between the identified variables of importance at each scale and the corresponding species.

Variables considered in the nested models for American eel, tessellated darter, and white perch were systematically ranked based on the AUC, with the top 10 variables of importance presented in **Tables 1, 2, & 3** below. Furthermore, to provide a comprehensive overview, we have also included outputs for model performance, incorporating species distribution methods that exhibited sufficient accuracy to be included in the final ensemble model. Model performances for each test species were quantified through categorical ratings, reflected by AUC and TSS metrics, as detailed in **Tables 4, 5, & 6** below. It is imperative to note that the rating scale spans from poor to excellent for each spatial distribution method as well as the final ensemble model, emphasizing the restricted use of these preliminary findings. As further analyses and validations are undertaken, the robustness and reliability of these results will become more thoroughly established.

Table 1. Top 10 Variables of Importance for American eel (NESDM Approach)
Derived from the Nested Modeling Efforts at each raster grid scale. Rank is shown on the left side ranging from 1 (more influential) to 10 (less influential).

	American Eel		
VIP Rank	1000m	100m	10m
1	Mean temp of wettest month	Flow accumulation	Solar radiation (winter)
2	% Pit (geomorphon)	% Pit (geomorphon)	% Pit (geomorphon)
3	Temperature seasonality	Total flow accumulation of tree canopy over turf	Topographic roughness index
4	% Spur (geomorphon)	% Forest	Solar radiation (summer)
5	Flow accumulation	Solar radiation range	% Hollow (geomorphon)
6	Isothermality	Flow direction	Flow accumulation
7	% Watershed that is impervious surface with exponential fall off based on distance	% Forested river wetland	Flow direction
8	Precipitation during driest month	Bottom Elevation	Total annual precipitation
9	Number of major dams	Solar radiation (winter)	Bottom Elevation
10	Topographic position index	Solar radiation (summer)	Aspect

Table 2. Top 10 Variables of Importance for tessellated darter (NESDM Approach)
Derived from the Nested Modeling Efforts at each raster grid scale. Rank is shown on the left side ranging from 1 (more influential) to 10 (less influential).

Tessellated Darter			
VIP Rank	**1000m**	**100m**	**10m**
1	% Pit (geomorphon)	Solar radiation (range)	% Watershed that is forest
2	% Roads	Stream power index	% Hollow (geomorphon)
3	Temperature seasonality	Solar radiation (summer)	Flow direction
4	% Forest	% Watershed that is orchard	% Cropland
5	% Scrub	% Pit (geomorphon)	Aspect
6	Flow direction	% Watershed that is terrine wetland	Solar radiation (range)
7	Precipitation during the warmest month	% Foot slope (geomorphon)	Bottom Elevation
8	% Tidal forested wet-lands	% Forest	Stream power index
9	Mean temp during the warmest quarter	Aspect	Minor dams
10	% Watershed that is cropland	% Watershed that is forested river wetland with exponential fall off based on distance	% Foot slope (geomor-phon)

Table 3. Top 10 Variables of Importance for white perch (NESDM Approach)
Derived from the Nested Modeling Efforts at each raster grid scale. Rank is shown on the left side ranging from 1 (more influential) to 10 (less influential).

White Perch			
VIP Rank	**1000m**	**100m**	**10m**
1	% Spur (geomorphon)	Bottom Elevation	% Flat (geomorphon)
2	% Hollow (geomorphon)	% Estuary marine	Aspect
3	Total flow accumulation of estuary marine	Total flow accumulation of forest	Topographic wetness index
4	% Watershed that is orchard with linear fall off based on distance	% Watershed that is tidal wetland with exponential fall off based on distance	% Roads
5	Total flow accumulation of extractive land use	Stream power index	Topographic position index
6	Flow accumulation of orchard with exponential fall off based on distance	Flow direction	% Watershed that is cropland
7	% Flat (geomorphon)	Flow accumulation of forest with linear fall off based on distance	% Forested river wetlands
8	% Watershed that is orchard	Flow accumulation of cropland with exponential fall off based on distance	Total flow accumulation of scrub
9	Mean temp of the wettest month	Flow accumulation	Solar radiation (mean)
10	% Slope (geomorphon)	Total annual precipitation	% Watershed that is natural succession with exponential fall off based on distance

From the 21 types of species distribution modeling methods listed in **Appendix Table A5**, only select distribution models were found suitable at each given ensemble run. **Tables 4, 5, and 6** show the models included in the final ensemble for American eel, tessellated darter, and white perch at each resolution along with AUC and TSS performance ratings based on prior literature. AUC and TSS ratings are indicators of model accuracy for individual SDM models and our final ensemble model used at each given scale. TSS values range from 0 to 1; values from 0.2 to 0.5 indicate poor model fit, values from 0.6 to 0.8 denote adequate/fair model fit, and values greater than 0.8 are considered excellent model fit (Coetzee et al., 2009). Models with AUC values <0.5 are considered worse than random (poor), values from 0.5 to 0.7 are considered fair, 0.7–0.9 are considered good, and values > 0.9 are considered an excellent fit (Swets, 1988).

Table 4. American eel NESDM Performance

Model	American Eel 1000m AUC	1000m TSS	100m AUC	100m TSS	10m AUC	10m TSS
BRT	Good	Fair	Good	Fair	Good	Poor
CART	Good	Poor				
Domain	Fair	Poor				
GLM	Good	Poor				
MARS	Good	Poor	Good	Fair		
MAXENT	Good	Fair	Good	Fair	Fair	Poor
MAXLike	Good	Poor				
MLP	Good	Fair	Good	Fair	Fair	Poor
RBF	Fair	Poor			Fair	Poor
RF	Good	Fair	Good	Fair	Good	Poor
RPart	Good	Poor				
SVM	Good	Poor	Good	Fair	Fair	Poor
Ensemble	Good	Fair	Good	Fair	Good	Fair

Table 5. Tessellated darter NESDM Performance

Model	Tessellated Darter 1000m AUC	1000m TSS	100m AUC	100m TSS	10m AUC	10m TSS
BRT	Good	Poor	Good	Fair	Good	Poor
CART	Good	Poor	Good	Fair	Good	Poor
Domain	Fair	Poor				
GLM	Fair	Poor				
MARS	Good	Poor	Good	Fair		
MAXENT	Good	Fair	Good	Fair	Good	Poor
MAXLike	Good	Poor				
MLP	Good	Poor	Good	Poor	Good	Poor
RBF	Fair	Poor			Fair	Poor
RF	Good	Fair	Good	Fair	Good	Fair
RPart	Good	Poor				
SVM	Good	Poor				
Ensemble	Good	Fair	Good	Fair	Fair	Poor

Table 6. White perch NESDM Performance

Model	White Perch 1000m AUC	1000m TSS	100m AUC	100m TSS	10m AUC	10m TSS
BRT	Excellent	Excellent	Good	Fair	Good	Poor
CART	Excellent	Excellent	Good	Fair		
Domain	Excellent	Excellent				
GLM	Excellent	Excellent				
MARS	Excellent	Excellent	Good	Fair		
MAXENT	Excellent	Excellent	Good	Fair	Good	Poor
MAXLike	Excellent	Excellent				
MLP	Excellent	Excellent	Good	Fair	Good	Poor
RBF	Excellent	Excellent				
RF	Excellent	Excellent	Good	Fair	Good	Poor
RPart	Excellent	Excellent	Good	Fair		
SVM	Excellent	Excellent	Fair	Poor		
Ensemble	Excellent	Excellent	Excellent	Excellent	Good	Fair

For layers representing probability of occurrence for each species (American eel, tessellated darter, and white perch), values ranged from 0 to 1 with 0 being low/no probability of presence and 1 being high probability of presence. We categorized the probability of occurrence into either low (<= 0.33), medium (>0.33 & <=0.66), or high (>0.66) bins based on the numerical value in a given grid cell (**Figure 7; a- American eel, b- tessellated darter, c- white perch**).

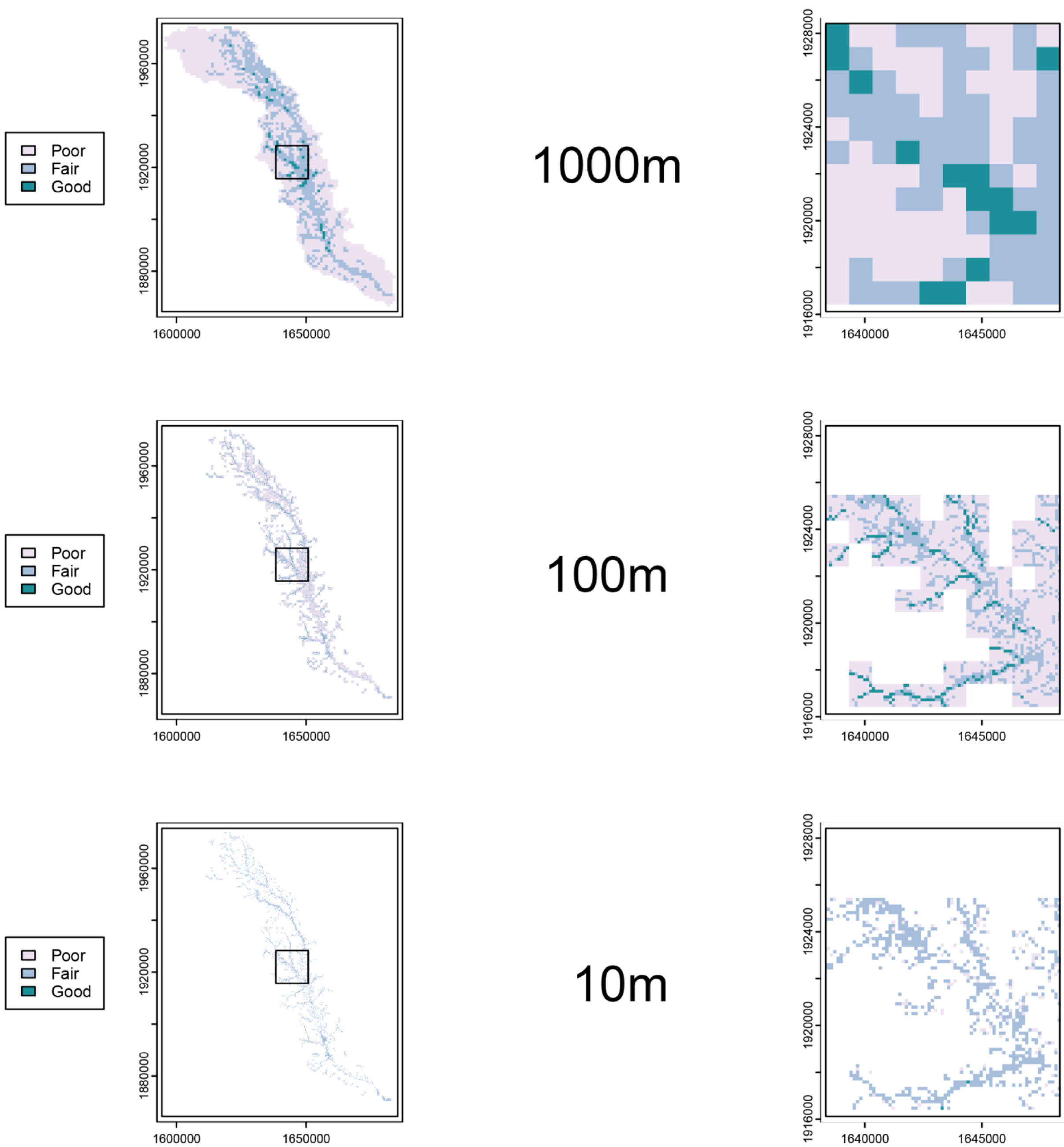

Figure 7a. NESDM outputs of probability of presence layers for American eel at three spatial resolutions (1000m, 100m, and 10m). Higher resolutions are subset to areas predicted as suitable by the previous scale model. Plots on the right show a zoomed in area defined by the boundary boxes in the left-hand plots at each resolution.

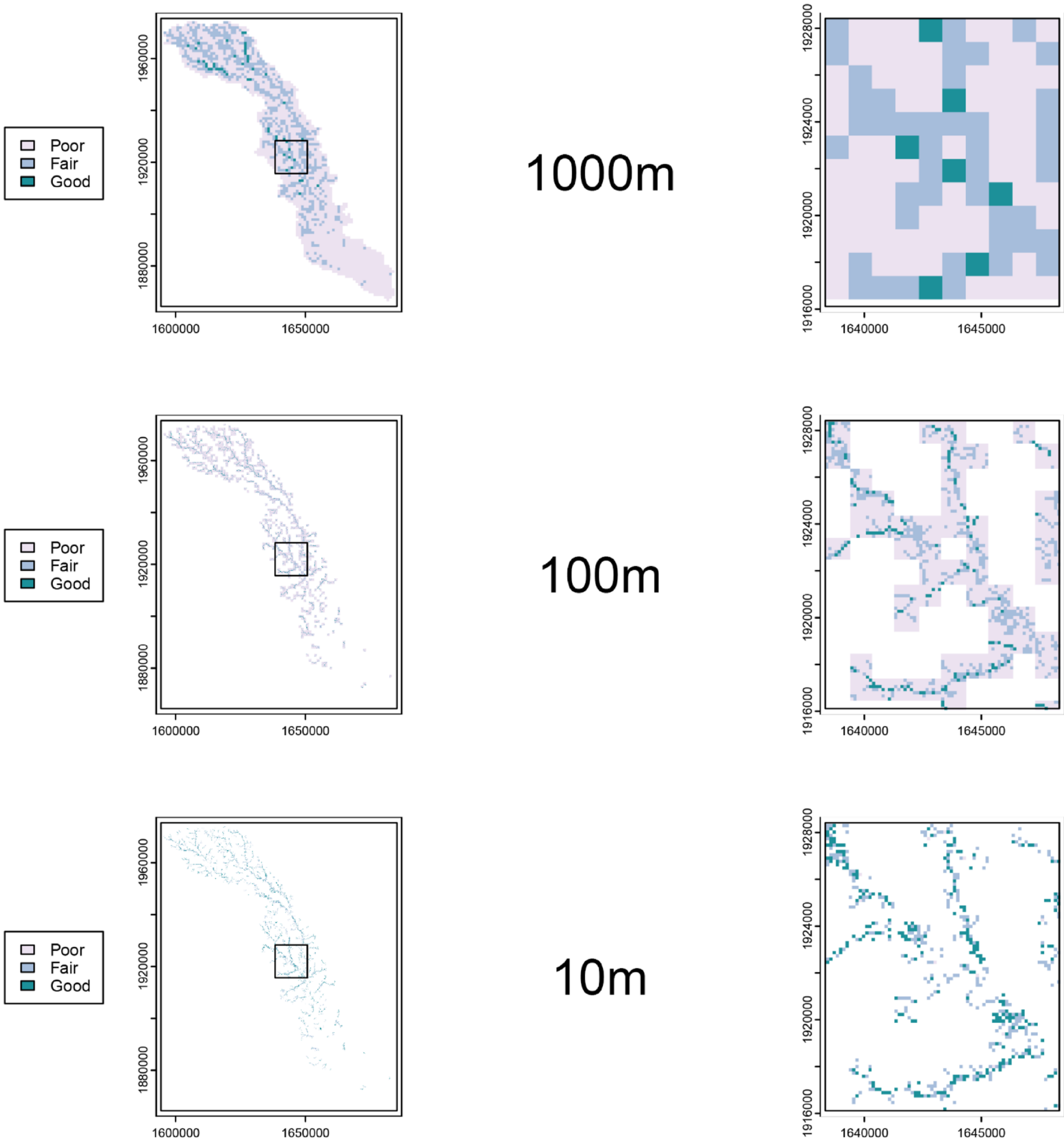

Figure 7b. NESDM outputs of probability of presence layers for tessellated darter at three spatial resolutions (1000m, 100m, and 10m). Higher resolutions are subset to areas predicted as suitable by the previous scale model. Plots on the right show a zoomed in area defined by the boundary boxes in the left-hand plots at each resolution.

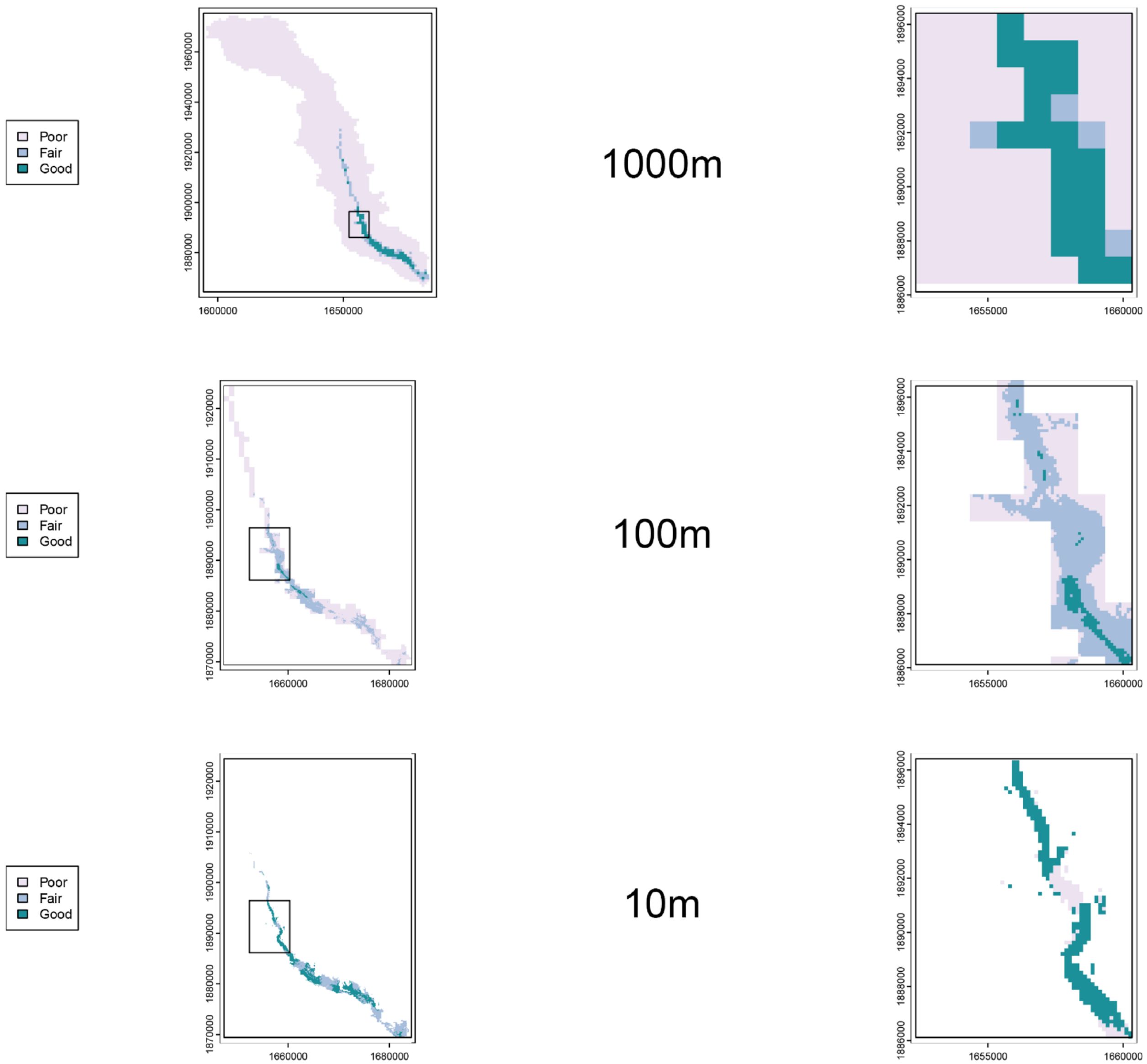

Figure 7c. NESDM outputs of probability of presence layers for white perch at three spatial resolutions (1000m, 100m, and 10m). Higher resolutions are subset to areas predicted as suitable by the previous scale model. Plots on the right show a zoomed in area defined by the boundary boxes in the left-hand plots at each resolution.

Between the multiple scales of reference, there are noticeable differences in predicted area of presence (in m^2). Looking at the Patuxent basin-wide model where a species is indicated as having possible presence according to the trained ensemble output, the percentage of predicted area was calculated from the 100m to the 1000m and 10m to the 1000m, illustrating the differences in predicted areas across finer scales (see **Table 7**).

Table 7. Difference in Predicted Area between Scales (Basin-wide NESDM Approach)
Area (in m^2) of the Patuxent basin where a species is indicated as having possible presence according to the trained ensemble model output at the 1000m scale. Percentage of predicted area respectively from the 100m to the 1000m and 10m to the 1000m.

Resolution	American Eel (Nested)		Tessellated Darter (Nested)		White Perch (Nested)	
	m^2	% of lowest resolution	m^2	% of lowest resolution	m^2	% of lowest resolution
1000m	810,000,000		716,000,000		140,000,000	
100m	188,140,000	23.23%	121,710,000	17.00%	60,770,000	43.41%
10m	77,575,900	9.58%	76,336,600	10.66%	21,266,300	15.19%

The area predicted as suitable for all three species were substantially reduced with each increased level of resolution. As unsuitable spatial area was removed, the predicted area of the nested models could not be greater than the previous resolution. However, subsequent models did not predict to the entire spatial extent of the previous level, suggesting sufficient tuning of habitat model predictions at finer resolutions (see **Figure 8**; example exclusive to American eel).

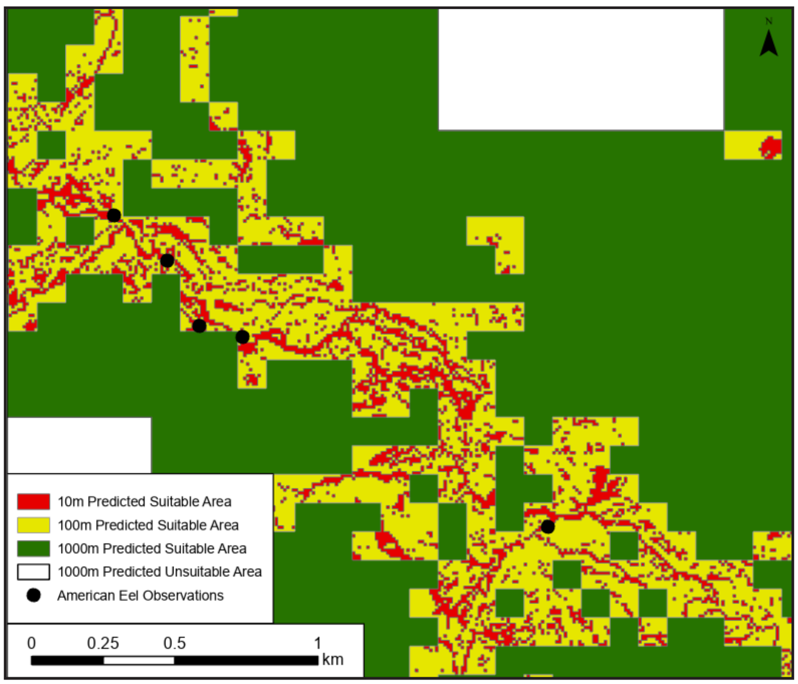

Figure 8. Nested predictions of suitable habitat for American eel. Areas predicted as suitable from multi-resolution models (1000m, 100m, 10m) resolution models. Subsequent models were confined within the extent of predicted area from the previous models (nested).

3.4 Methods and Initial Results for a Tidal-bound Non-nested Modeling Approach

Most environmental variables extended throughout the entire Patuxent basin. However, some of the in-water variables, such as water temperature and dissolved oxygen, lacked sufficient spatial resolution in non-tidal waters to be included in the full watershed assessment. This will likely be true of other factors that may be included in the future and represents a good test of the framework with different spatial extents. Therefore, we conducted two sets of analyses, one that included the entire watershed and the other that only included tidal waters. Distribution modeling for tessellated darter and American eel were exclusively conducted for the entire watershed, while white perch distribution was modeled across the entire watershed as well as using a tidal-bound restricted model incorporating the following environmental predictor layers: surface water quality metrics, distance to hardened shoreline, distance to SAV beds, distance to protected areas, tidal B-IBI, and bottom substrate type.

For white perch, an ensemble model of variable importance was developed using fish presence and pseudo-absence data derived from all the variables that passed the test for collinearity. Variables included in the tidal-bound non-nested ensemble model can also be seen in **Appendix Table A4**. Similarly to the NESDM approach, we created random pseudo-absence points at two times the number of presence points (Barbet-Massin et al., 2012; Liu et al., 2013) for white perch distributed throughout the tidal waterbody.

3.4.1 Tidal-bound Non-nested Ensemble Modeling Approach

The non-nested ensemble species distribution model for our largely estuarine resident species, white perch, took on a very similar structure to the NESDM approach. The variables initially used in the NESDM approach were also utilized in the non-nested approach for white perch, with the addition of environmental variables that lacked sufficient spatial resolution in non-tidal waters. These tidal driven factors included water quality parameters, substrate type standardized to Coastal and Marine Ecological Classification Standard, tidal B-IBI, and functions of distance accumulation to the following defined features: submerged aquatic vegetation, hardened shoreline structures, and protected areas.

All species distribution model methods (21; see **Appendix Table A5**) available within the SDM package were evaluated equally in a preliminary step and tested to ensure accuracy based on thresholds (AUC > 0.7 and TSS> 0.6). Selected models were then bootstrapped and combined into an ensemble model based on the weighted average (maximized [sensitivity + specificity]) to create POC based on habitat suitability. This analysis was conducted for all cells that fell within the tidal waterbody (the boundaries of which are described in section *3.1.2 Framework, Watershed, and Waterbody Extents).* Because the tidal waters provided a smaller extent and reduced computational processing limitations compared to modeling the entire HUC-8 basin, statistical evaluation, model selection, and ensemble modeling was conducted for all three spatial resolutions (1000m, 100m, and 10m raster cell sizes) without the need to restrict finer resolutions to the area of predicted presence from the previous resolution size (see **Figure 9** for the tidal-bound non-nested modeling workflow).

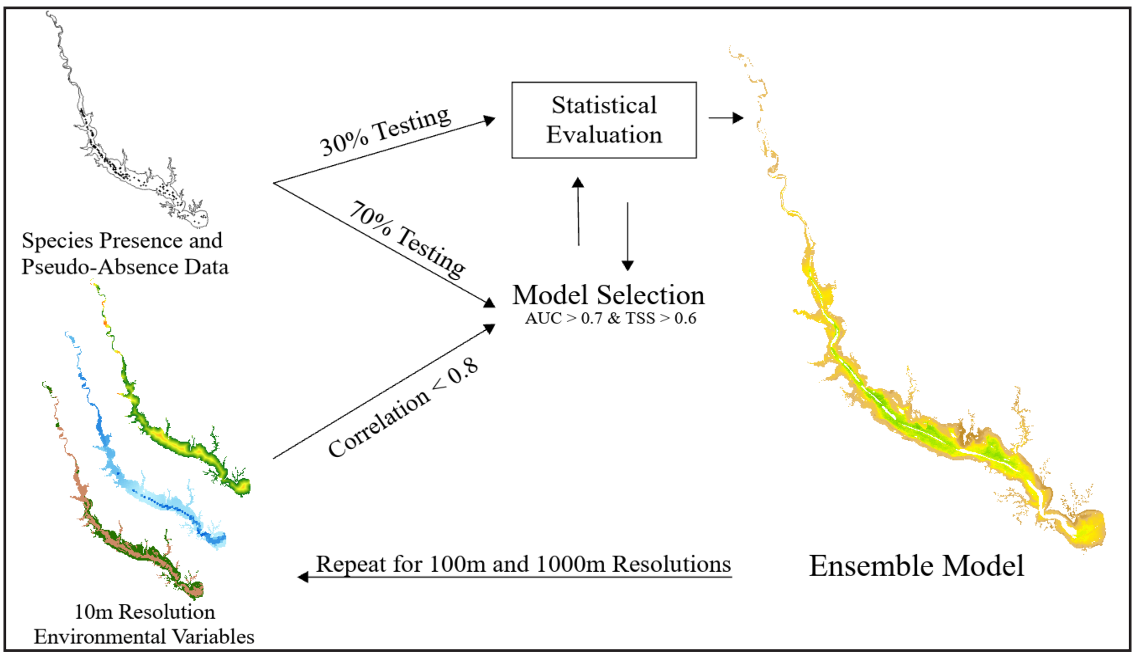

Figure 9. Conceptual model for the tidal-bound non-nested ensemble species distribution model.

Species occurrence and pseudo-absence data are split into 70% training and 30% testing datasets. Preliminary tests using all model methods (21) are run using 20 bootstrap repetitions and 20 replicates of random resamples. Models that met our inclusion criteria (AUC > 0.7 and/or TSS > 0.6) were included in the final ensemble model run where bootstrap and replicates were increased to 200. Individual model results were then assembled based on a weighted average that maximized sensitivity and specificity. POC was then used to determine areas of probable presence throughout the tidal extent of the Patuxent waterbody for all resolutions (1000m, 100m, and 10m)

3.4.2 Preliminary Results from the Tidal-bound Non-nested Ensemble Model

Using data for the tidal-bound region of the Patuxent, model outputs were generated for white perch at 1000m, 100m, and 10m raster scales to assess the spatial distribution patterns. Again, it is important to emphasize that caution must be exercised in drawing explicit conclusions or making connections between the identified variables of importance at each scale.

Variables included in the white perch models were systematically ranked based on the AUC, with the top 10 variables of importance presented in **Table 8** below. Also included are outputs for the non-nested model runs, incorporating species distribution modeling methods that exhibited sufficient accuracy. Method performance for each model was quantified through categorical ratings, reflected by AUC and TSS metrics, as detailed in **Table 9** below.

Table 8. Top 10 Variables of Importance for white perch (Non-nested Approach)

Derived from the Tidal-bound Non-nested Modeling Efforts at each raster grid scale. Rank is shown on the left side ranging from 1 (more influential) to 10 (less influential).

White Perch (Non-nested)			
VIP Rank	**1000m**	**100m**	**10m**
1	Temperature Seasonality (ts)	Bottom Elevation	Bottom Elevation
2	Bottom Elevation	Temperature Seasonality (ts)	Distance to Hardened Shoreline (m)
3	% Slope (geomorphon)	% Coverage of Mud	% Coverage of Sand
4	% of Grid Cell that is Tidal Forested Wetlands	% Slope (geomorphon)	% Coverage of Mud
5	% Coverage of Mud	Distance to Hardened Shoreline (m)	% of the Watershed that is Tree Cover Over Impervious
6	% Coverage of Biogenic Oyster Reef	% Coverage of Biogenic Oyster Rubble	% Coverage of Sandy Mud
7	Flow Accumulation	Slope	% of the Watershed that is Tidal Forested Wetlands
8	% Coverage of Sand	Precipitation Seasonality (ps)	Mean Temperature During the Warmest Quarter (mtwq)
9	% of the Watershed that is Terrene Wetlands	% of the Watershed that is Terrene Forested Wetlands	% of the Watershed that is Tidal Wetlands
10	Distance to SAV Beds (m)	% of the watershed that is Tidal Forested Wetlands	Temperature Annual Range (tar)

Regarding the species distribution modeling methods incorporated into the final ensemble, a statistically determined selection of BRT, CART, MaxEnt, RBF, and RF revealed their appropriateness for the tidal-specific analysis across different scales, as delineated in **Table 9**. The same categorical AUC and TSS ratings applied to the NESDM performance were used for the non-nested individual models and final ensemble model at each given scale. For TSS values, values from 0.2 to 0.5 indicate poor model fit, values from 0.6 to 0.8 denote adequate/fair model fit, and values greater than 0.8 are considered excellent model fit (Coetzee et al., 2009). Models with AUC values <0.5 are considered worse than random (poor), values from 0.5 to 0.7 are considered fair, 0.7–0.9 are considered good, and values >0.9 are considered an excellent fit (Swets, 1988).

Table 9. White Perch Tidal-bound Non-nested Model Performance

	White Perch					
	1000m		100m		10m	
Model	**AUC**	**TSS**	**AUC**	**TSS**	**AUC**	**TSS**
BRT	Good	Poor	Good	Fair	Good	Fair
CART			Good	Fair		
MAXENT	Good	Poor	Good	Fair		
RBF	Good	Poor			Good	Poor
RF			Excellent	Fair		
Ensemble	Good	Fair	Good	Fair	Good	Fair

For the output layers representing probability of occurrence for white perch at each scale, values ranged from 0 to 1. Similarly to the NESDM occurrence outputs for each species, we categorized the probability of occurrence values into either low (<= 0.33), medium (>0.33 & <=0.66), or high (>0.66) bins based on the individual grid cell value (**Figure 10**).

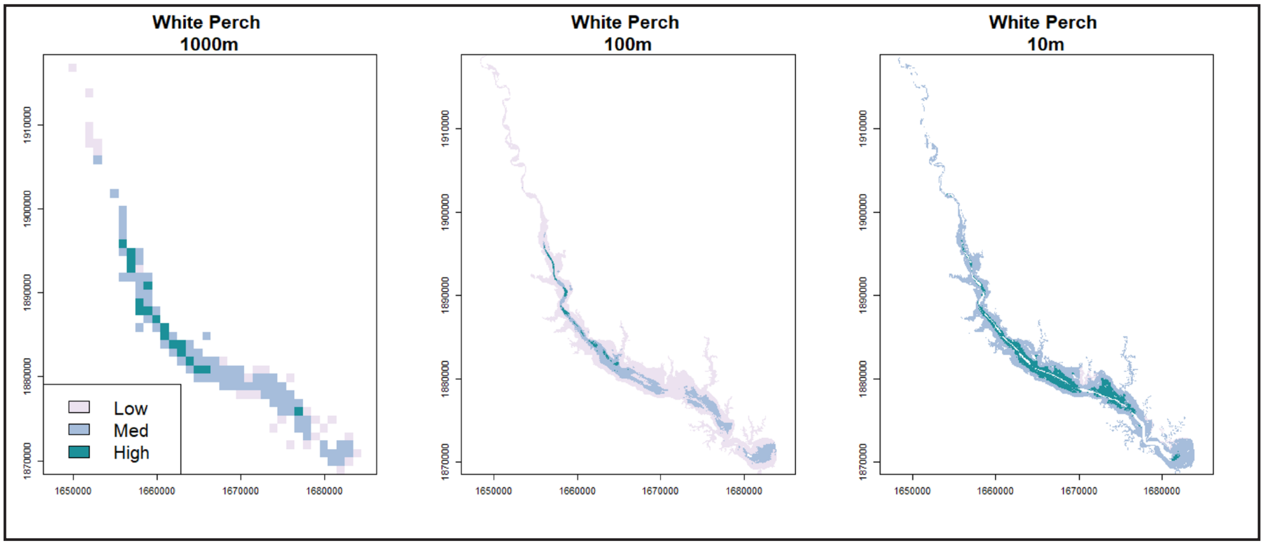

Figure 10. Tidal-bound Non-nested outputs of probability of occurrence layers for white perch at three spatial resolutions (1000m, 100m, and 10m). The probability of presence areas, particularly for the 1000m and 100m scales, are reflective of the predominance of sampling locations for white perch in open waters of the river (see **Figure 1**).

Differences in predicted area (in m^2) for white perch between scales is shown as a percentage of predicted area respectively from the 100m to the 1000m and 10m to the 1000m (**Table 10**).

Table 10. Difference in Predicted Area between Scales (Tidal-bound Non-nested Approach)
Area (in m^2) of the Patuxent tidal waterbody where a species is indicated as having possible presence according to the trained ensemble model output at the 1000m scale. Percentage of predicted tidal area respectively from the 100m to the 1000m and 10m to the 1000m.

White Perch (Non- Nested)		
Resolution	**m^2**	**% of lowest resolution**
1000m	41,000,000	
100m	15,980,000	38.98%
10m	34,820,100	84.93%

3.5 Comparison Between the Nested and Non-nested White Perch Models

Contrasting the forecasted presence areas for white perch using two modeling methods, NESDM and non-nested, helps pinpoint overlapping suitable habitats predicted by both models. This particular analysis also allows us to explore instances where one model may be less effective in predicting habitat presence based on scale.

Examining the predicted areas of presence that exceeded the AOT for white perch in both NESDM and non-nested models at different resolutions (1000m, 100m, and 10m) reveals a larger habitat prediction area at the 1000m and 100m scales for the NESDM outputs. However, a substantial shift in favor of the non-nested model is observed at the 10m scale (refer to **Figure 11**).

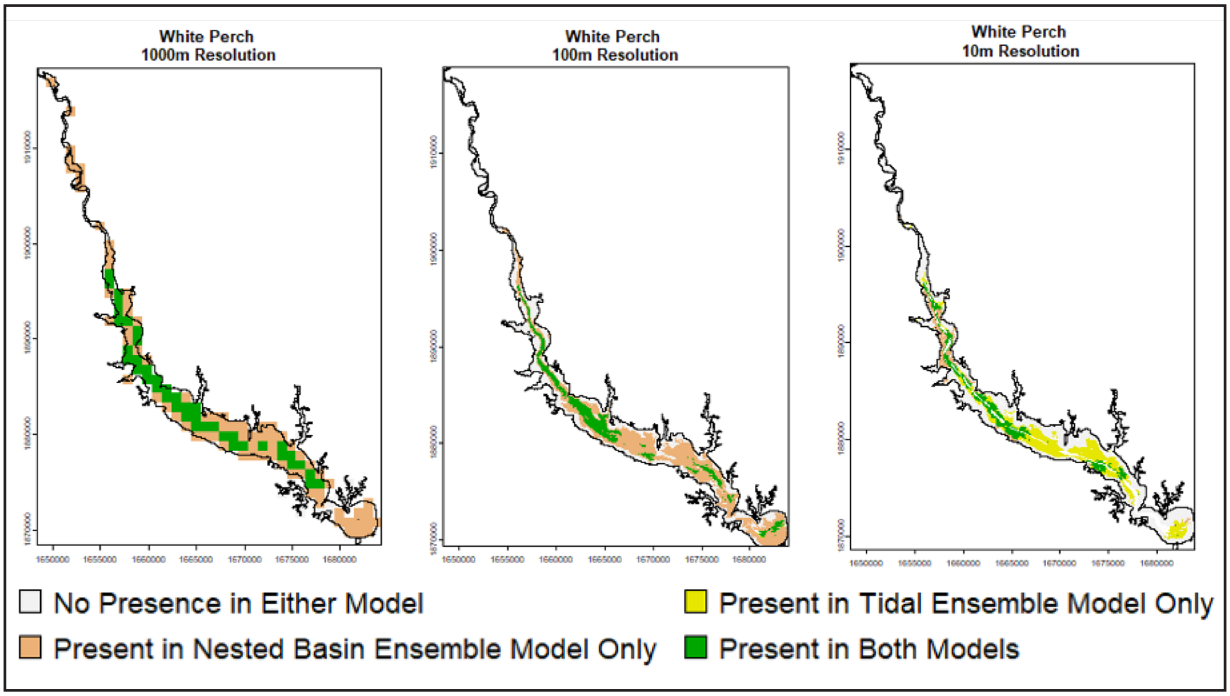

Figure 11. Overlap of predicted areas of presence that exceeded the AOT in white perch for the NESDM and non-nested model at varying resolutions (1000m, 100m, and 10m).

Comparing nested and non-nested modeling approaches for our predominantly tidal species offers valuable insights, of which include a possible limitation associated with geographic restrictions of habitat preferences within the finer resolutions of the nested approach. Another constraint stems from the exclusion of environmental factors in the basin-wide analysis, especially those driven by tidal dynamics, due to insufficient data in upland streams and rivers. This limitation has implications for potential oversight of environmental factors that are highly predictive of a largely tidal habitat species. In the context of our basin-wide nested modeling approach, the absence of tidal-driven variables could have potentially limited the accuracy and completeness of our assessment, impacting the understanding of species distribution patterns within that specific region for a given species. Therefore, the parallel analysis provides valuable insights into the trade-offs associated with spatial constraints and exclusion of data layers only available in the tidal waters, offering a more nuanced perspective for our largely tidal species.

4.0 Discussion of Outcomes

4.1 Framework Utility

The gridded framework provided an excellent structure for organizing data across the entire waterbody, integrating the influence of landscape stressors with local in-water factors, and providing the structure to seamlessly predict fish habitat distribution across freshwater and tidal environments. We were able to effectively summarize a broad set of variables from various spatial file structures into the framework. Further, the use of a continuous raster grid layer over the entire HUC-8 watershed allowed the application of several data summary approaches, such as Flow-Condition Parameter Grids (Barnhart et al. 2021) and Inverse Distance Weighted Accumulation (Peterson and Pearse 2017) to land-based features. These summarization methods attempt to account for the potential influence of distance on the relative importance of various environmental conditions.

4.2 Utility of Nested Ensemble Modeling Approach

Ensemble modeling provided a robust test of the proposed framework and a relatively new statistical approach for predicting fish presence based on multiple models and underlying environmental predictor variables. The use of a nested modeling approach, starting with the largest spatial scale (1000m in this case), and iteratively moving to finer scales, provided potentially important insights into different environmental influences at different spatial magnitudes. In fact, the variables of importance influencing the models changed substantially between the different resolutions (see **Tables 1, 2, and 3**). Although there were some differences in the variables included at the different raster scales, as selected by Variance Inflation Factor (VIF) testing (see **Appendix Table A4** for a list of all variables included in each model iteration), some very generalized trends were noted. For example, the coarsest resolution (1000m) analysis tended to emphasize variables that were more broad in geographical scale (e.g. climate and large landscape features like % geomorphon classes). The intermediate resolution (100m) more prominently featured riparian areas and was more influenced by percent land classes surrounding the stream area. The finest resolution (10m) had the greatest importance from energetic variables (solar radiation), from meso-scale features like aspect and bottom elevation, and from accumulation of LULC variables. Variables of importance also changed between species being tested. American eel was most impacted by climate variables (**Table 1**), tessellated darter was more impacted by riparian influences (**Table 2**), and the white perch was most impacted by land-based features, estimated by either flow accumulation and percent water basin variables (**Table 3**).

Predicted area (i.e. area predicted as suitable or above the AOT) also changed substantially among the NESDM outputs. The 1000m models predicted on average 10 times the amount of suitable area as the 10m models, and 3-4 times the area as the 100m models (see **Table 7**). This may be influenced by the nested approach removing large portions of the basin prior to predictions of the higher resolution datasets. However, greater than 92% of all occurrence data were included at all resolutions for all species with the exception of the tessellated darter 100m model which still contained 83% of occurrences. Therefore, only the range in the randomly generated pseudo-absences changed substantially. When compared to the tidal models, which did not remove area between resolutions, we see that predicted area of presence decreased in a similar fashion to the NESDM from the 1000m to 100m scales, but not of the same magnitude as you moved to the finest (10m) resolution (see **Table 10**). In further comparison between nested and non-nested models, the nested method predicts white perch presence in more waterbody areas than the tidal-bound non-nested models at 1000m and 100m scales (see **Figure 11**). However, at the 10m scale, there's a meaningful shift, with the tidal-bound model predicting presence in far more waterbody areas than the nested model. This shift may be due to the removal of critical habitat areas during the nesting process or the exclusion of environmental factors unique to the estuarine waterbody and only included in the tidal-bound analysis.

When comparing between the three fish species modeled in the nested approach, the spatial area predicted as suitable was not substantially different between the American eel and tessellated darter, however the location of areas predicted to be suitable differed, as expected (see **Figure 7**). The white perch predicted area was lower than the two basin-wide species, but given its dependence on estuary waters, this is not unexpected.

Model accuracy and the number of models included in the ensemble decreased with increased spatial resolution (see **Tables 4, 5, and 6**). The 1000m model had the highest model accuracy and number of model types included while the 10m had the lowest number of models and accuracy. This trend was seen in all three of the focal species. The decreasing model accuracy could be due to limitations of including pseudo-absences instead of true absence data. The higher resolution models are only predicted in areas of high suitability based on the previous scale. Therefore, pseudo-absence data created within the nested model extent is within high suitability areas. Therefore, the finer scale models are refining areas already deemed as suitable from the higher scale, thus they are modeling only the residual "suitable" habitat. Modeling of unsuitable areas will therefore be less informative than the previous scale models.

The differences in variables of importance were likely influenced by the nested model removing areas that were not suitable and therefore removing variables that would have impacted predicted presence. For example, the cumulative number of dams was important for the American eel at the lowest resolution (1000m). In using the nested model approach to remove areas that were not deemed suitable based on this characteristic and others, the importance of the number of dams substantially decreased at the other resolutions to the point where removing it did not influence the model accuracy at either the 100m and 10m resolution. This shows the conceptual framework of the hierarchical model at work. The importance of a variable at one resolution was used to remove areas impacted by this feature. The models at the higher resolutions then focused on variables that were more impactful within the new spatial extent without the influence of variables impacting them at greater scales and could then be used to increase our understanding of the variables influencing habitat suitability at a higher resolution.

4.3 Considerations of the Framework and Modeling Approaches

- Because fish presence data were aggregated from multiple surveys using different collection methods and gear, true absences were not produced for all surveys. The generation of pseudo-absences was modeled instead and may have influenced the predicted fish distributions.
- A common understanding of modeling is summarized well by the notion that "all models are wrong, but some are useful", largely attributed to George Box (1976). For example, all models are limited by the data inputs. For this study, this was apparent in the lack of some variables (such as water temperature and benthic habitat type) throughout the entirety of tidal and non-tidal waters, discrepancies in the calculation and subsequent use of IBI between tidal and non-tidal waters, and the sparse fish survey data in shallow waters of the tidal river. However, the framework allowed for the exploration of many different variables and the identification of data limitations.
- The example modeling approach employed in this pilot has the advantage of exploring potential statistical relationships of a broad suite of environmental factors at multiple spatial scales. However, the use of the nested approach does have the potential for the exclusion of some areas of possible presence. In our study, most of the observed presence locations remain even when areas were interactively selected for finer scale model runs, suggesting that the potential exclusion of preferred areas is limited.
- Perhaps the largest limitations in the example model runs conducted for this study were the flattening of time and the focus on juvenile and adult fish. Inter-annual and decadal variability in species population dynamics occur and are important in identifying relationships between species and their environment. However, the framework and modeling approach allows for testing differences in predicted occupancy at various times and life stages, if sufficient data are used to support those analyses.
- The model results in this pilot study represent the probability of occurrence based on habitat suitability as indicated by known observations and associated environmental conditions. As such, the predicted areas of presence indicate the likelihood of habitat associations and do not indicate definitive areas where presence would be expected, even where predicted highly suitable. Additional testing of the model strength in predicting presence would need to be provided by subsequent sampling. However, our approach could be used as a guide for areas most suggestive of supporting habitats for the target species based on environmental factors and known previous occurrence.
- We acknowledge that if we were just doing estuarine modeling, we would be including other dynamic environmental conditions, such as thermal stratification, dissolved oxygen patterns at varying depths, or bidirectional (tidal) flow. Our approach simplifies the estuarine environment in order to couple methods suitable for both inland and estuarine waters. However, our results do largely agree between the estuarine only models and the basin-wide nested approach.

5.0 Framework and Model Enhancements and Applications

5.1 Extending the analytical framework

The analytical framework design could easily be expanded to other tributaries of the Chesapeake Bay and the entire Bay itself, but with several important caveats. The first caveat is with a larger spatial extent, more computer processing power would be required, and more importance would need to be placed on the nested modeling approach. For example, an analysis of the Potomac River or Susquehanna River Watersheds at the 10m raster scale would potentially require a higher amount of processing power (RAM) and time than required for our Patuxent River watershed analysis. For comparison, the tidal waters 10m analysis for white perch took approximately seven days to run on 48 processors within the Microsoft® Azure cloud environment. The second caveat is there is far less fish survey data available for some areas of the Chesapeake Bay. Availability of fish data may be due to lack of sampling or data sharing restrictions. For non-tidal waters, especially, the number and density of fish sampling stations varies between states.

5.2 Extending the statistical analysis

An important next step would be to look at particular seasons and life stages, as data availability allows. Lumping data across seasons and life stages likely results in prediction of more generalized habitat connections, whereas particular environmental variables are likely to be especially important for certain life stages of a particular species. For the tidal-specific modeling, we restricted the analysis to areas with benthic habitat structure data, which proved to be an important decision as several benthic classes were predicted to be selected as important by the white perch models at all three raster sizes. However, this may have impacted the importance of other variables, such as shoreline types, in the models.

5.3 Data Expansion and Data Gaps

5.3.1 Best Management Practices

The USGS team is developing summaries of best management practices (BMPs) for the Chesapeake Bay watershed. These BMPs could potentially inform selection and use of environmental variables within our practical framework as well as become factors in the framework itself.

5.3.2 Sediment Dynamics

Members of the USGS are also assessing the sediment dynamics, such as sediment fluxes and erodible banks, within rivers of the Chesapeake Bay. The Patuxent Pilot team met with USGS sediment dynamic experts several times to understand if any sediment movement data could be incorporated into this pilot study. Ultimately, it was decided that in order to produce the initial set of analyses described in this report, more work was necessary to understand sediment dynamics at a relatively fine spatial scale within the Patuxent River (Greg Noe, USGS Water Mission Area, USGS Chesapeake Bay Workshop session, oral communication, June 2022).

5.3.3 Water Quality Data

Due to computer processing demand and the relatively large file size of interpolated water quality, we selected a handful of years to represent average conditions in tidal portions of the Patuxent River. Maryland DNR have summarized data for each of the 100+ water quality sampling events conducted by MD DNR at each of the Eyes on the Bay[24] monitoring stations. Future modeling efforts could leverage these data to summarize water quality based on seasons or particular time periods.

5.3.4 Fish Data

There are many areas of the Chesapeake Bay where fish sampling has been sparse or nonexistent. A broad assessment of all tributaries could be done using the currently available data, but would likely be biased by the lack of fish survey data in some areas. However, the amount of fish survey information in the Chesapeake Bay would allow for the application of the joint pilot approach for certain species in other tributaries or, given the constraints described, for the Bay in general. Ideally, an assessment of habitat would be relevant for all managed fish species. However, data availability varies considerably between species. It would be possible to also advance the framework beyond using primarily presence and absence data by including surveys that provide consistent information on species abundance or catch rates in a particular benthic or water column habitat(s) to define a range of species habitat affinities (Monaco et al. 1998).

Recently the Chesapeake Bay Program's Science and Technical Advisory Committee released the Comprehensive Evaluation of System Response (CESR) report (STAC, 2023), which highlights the need to evaluate changes in living resource populations and the various habitats they use, rather than just focusing on the water quality goals that support them. In particular, there is an emphasis on "shallow water" habitat, which generally refers to the 'edges of the waterbody'. The effects of sparse fish survey locations in shallow waters of the tidal Patuxent River may have contributed to the predominance of white perch predicted presence areas in the central channel, as seen in this current study. While additional fish data exist for shallow waters in some other tributaries, there remains a bias towards open water fish sampling for long term monitoring studies in tidal waters of the Chesapeake Bay. Further implications of this study relative to the recommendations in the CESR report are discussed in section 5.4 below.

5.4 How the Framework and Modeling Approach Might Be Used

Our goal was to design and test an analytical framework for a tributary of the Chesapeake Bay in order to see if this would provide a more holistic approach than past tidal and non-tidal assessments that were conducted separately. The key use of this pilot project was to inform and direct subsequent studies, resulting from the success of this initial project. However, there are some preliminary takeaways for both fisheries managers and land use managers from this pilot. For example, the influence of different types of environmental factors at different spatial scales, though perhaps not surprising, suggests the need for different management approaches depending on the scale of the management action. Running a full Baywide analysis at different scales allows managers to assess and target specific geographic locations based on their particular needs and goals. Further, the pilot highlighted that there are ongoing needs to address data gaps, such as water quality monitoring in non-tidal waters and fish survey data from other stream reaches and tributaries. However, our modeling approach in the Patuxent lays the foundation for a more flexible and integrated framework of modeling the connection between upland land uses and the Bay's receiving waters. This ability of the framework to estimate potential predictor variables from a broad suite of data types (raster, polygon, point) across the entire watershed also demonstrates the importance of a coupled inland-estuarine framework versus having separate tidal and non-tidal assessments as has been standard practice in the past.

The multi-scale, nested modeling approach used in this study might also have some application to recommendations provided in the recently released CESR report[25] developed by the Chesapeake Bay Program's Science and Technical Advisory Committee (STAC, 2023). The report calls for improved modeling of living resource populations and the various habitats they use, rather than just the previous Chesapeake Bay total maximum daily load (TMDL) emphasis on the water quality goals that support them, stating that "finer spatial scale modeling and monitoring could further identify high nutrient loss areas and operations and be used to consider more effective treatment options." Modeling of nutrients, dissolved oxygen and suspended sediments in shallow, tidal waters are improving (Vargas-Nguyen et al., 2023)) and could be used as input to the fine-scale framework for fish habitat assessments. The framework and modeling described in this study also helps improve upon existing living resource modeling by incorporating fine-scale data, such as the recently developed 1m land use land cover data release for the Bay (Chesapeake Bay Program Land Use/Land Cover Data Project[26]). In particular, the estimation of distance metrics for land-based features, rather than a simple summarization of conditions throughout the catchment, could be advantageous for modeling shallow water habitats. Further, improved water quality models could be coupled with fish survey information (e.g., abundance, size, condition) using the framework and modeling approaches described in this study to address the information gap mentioned in the CESR report that "...the CBP [Chesapeake Bay Program] does not use models to relate changes in dissolved oxygen and habitat to the composition or abundance of living resources."

Literature Cited

Allan, J. D. (2004). Landscapes and riverscapes: the influence of land use on stream ecosystems. *Annual Review of Ecology, Evolution, and Systematics* 35:257–284.

Bahner, L. (2001). The Chesapeake Bay and tidal tributary volumetric interpolator, software VOL3D version 4.0. NOAA Chesapeake Bay Office, Annapolis, MD.

Barbet-Massin, M., F. Jiguet, C.H. Albert, W. Thuiller (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3: 327-338. https://doi.org/10.1111/j.2041-210X.2011.00172.x.

Barnhart, T.B., A.R. Schultz, S.A. Siefken, F. Thompson, T. Welborn, T.R. Sando, A.H. Rea, P.M. McCarthy (2021). Flow-Conditioned Parameter Grids for the Contiguous United States: A Pilot, Seamless Basin Characteristic Dataset: U.S. Geological Survey data release, https://doi.org/10.5066/P9HUWM6Q.

Box, G. (1976). Science and statistics. Journal of the American Statistical Association, 71 (356): 791–799, doi:10.1080/01621459.1976.10480949.

Buto, S.G., and Anderson, R.D., 2020, NHDPlus High Resolution (NHDPlus HR)---A hydrography framework for the Nation: U.S. Geological Survey Fact Sheet 2020-3033, 2 p., https://doi.org/10.3133/fs20203033.

Coetzee, B. W., Robertson, M. P., Erasmus, B. F., Van Rensburg, B. J., & Thuiller, W. (2009). Ensemble models predict Important Bird Areas in southern Africa will become less effective for conserving endemic birds under climate change. *Global Ecology and Biogeography*, 18(6), 701-710.

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. http://assessment.fishhabitat.org/

Danielson, J., S. Poppenga, D. Tyler, M. Palaseanu-Lovejoy, D. Gesch (2018) Coastal National Elevation Database. USGS Fact Sheet. 2pp. 10.3133/fs20183037.

Esselman, P.C., D.M. Infante, L. Wang, D. Wu, A.R. Cooper, and WW. Taylor. (2011). An index of cumulative disturbance to river fish habitats of the conterminous United States from landscape anthropogenic activities. *Ecological Restoration* 29: 133-151.

Fish Habitat Action Team. (2022). Fish Habitat 2022-2024 Action and Logic Plan. Action 2.1. https://www.chesapeakeprogress.com/abundant-life/fish-habitat/logic-action-plan.

Frissell, C.A., W.J. Liss, C.E. Warren, M.D. Hurley (1986) A hierarchical framework for stream habitat classification: viewing streams in a watershed context. *Environmental Management*, 10, 199-214.

Greene, C.M., K. Blackhart, J. Nohner, A. Candelmo, and D.M. Nelson. (2015). A National Assessment of Stressors to Estuarine Fish Habitats in the Contiguous USA. *Estuaries and Coasts* 38:782-799.

Hijmans, R.J., S. Phillips, J. Leathwick, J. Elith, M.R.J. Hijmans (2017). Package 'dismo'. *Circles*, 9(1), 1-68.

Hijmans, R.J., R. Bivand, K. Forner, J. Ooms, E. Pebesma, M.D. Sumner (2022). Package 'terra'.

Hunt, G., D. Bilkovic, S. Faulkner, T. Ihde, M. McGinty, M. Monaco, T. O'Connell, P. Tango, B. Vogt, K. Maloney, J. Young, L. Williamson, A. Leight, R. Dixon (2018). Factors Influencing the Headwaters, Nontidal, Tidal, and Mainstem Fish Habitat Function in the Chesapeake Bay Watershed: Application to Restoration and Management Decisions. STAC Publication Number 18-006, Edgewater, MD. 112 pp.

Hynes, H. B. N. (1975). The stream and its valley: With 4 figures and 2 tables in the text. Internationale Vereinigung für theoretische und angewandte Limnologie: Verhandlungen, 19(1), 1-15.

Jasiewicz, J., T.F. Stepinski (2013). Geomorphons—a pattern recognition approach to classification and mapping of landforms. *Geomorphology*, 182, 147-156.

Jenson, S. K., and J. O. Domingue. (1988). "Extracting Topographic Structure from Digital Elevation Data for Geographic Information System Analysis." *Photogrammetric Engineering and Remote Sensing* 54 (11): 1593–1600.

Kiser, A. H., K.S. Cummings, J.S. Tiemann, C.H. Smith, N.A. Johnson, R.R. Lopez, C.R. Randklev (2022). Using a multi-model ensemble approach to determine biodiversity hotspots with limited occurrence data in understudied areas: An example using freshwater mussels in México. *Ecology and Evolution*, 12, https://doi.org/10.1002/ece3.8909.

Leight, A., G. Hunt, E. Markin (2019). Chesapeake Bay Fish Habitat: Summary of Stakeholder Feedback and Identified Needs. White paper delivered to Chesapeake Bay Program, Fish Habitat Action Team. 29pp. Chesapeake Bay Fish Habitat Summary of Stakeholder Feedback/Needs.

Leight, A., S. Skelley, J. Christensen, D. Dorfman, C. Jeffrey, D. Bruce (2021). Recommendations for Conducting Fish Habitat Assessments in Tidal Waters of the Chesapeake Bay. Report to the Chesapeake Bay Program, Fish Habitat Action Team. 83pp. Recommendations for Conducting Fish Habitat Assessments in Tidal Waters of the Chesapeake Bay.

Liu, C., M. White, G. Newell (2013). Selecting thresholds for the prediction of species occurrence with presence only data. *Journal of Biogeography*, 40(4), 778-789.

Maloney, K.O., K. Krause, M. Cashman, W. Daniel, B. Gressler, D. Wieferich, J. Young (2022).
Using fish community and population indicators to assess the biological condition of streams and rivers of the Chesapeake Bay watershed, USA. *Ecological Indicators,* 134, https://doi.org/10.1016/j.ecolind.2021.108488.

Martin, E., K. Hoenke, L. Havel (2020). Fish Habitat Conservation Area Mapping and Prioritization Project. https://www.atlanticfishhabitat.org/wp-content/uploads/2020/08/ACFHP-Mapping-and-Prioritization-Final-Report.pdf, Last accessed 2/16/2024.

Monaco, M.E., Weisberg, S.B., T.A. Lowery. (1998). Summer habitat affinities of estuarine fish in U.S. mid-Atlantic coastal systems. Fisheries Management and Ecology. (5) 161-171.

Naimi, B., M.B. Araújo (2016). sdm: a reproducible and extensible R platform for species distribution modeling. *Ecography*, 39: 368-375. https://doi.org/10.1111/ecog.01881.

Naimi, B., N.A.S Hamm, T.A. Groen, A.K. Skidmore, A.G. Toxopeus (2014). Where is positional uncertainty a problem for species distribution modeling? *Ecography* 37 (2): 191-203.

Patuxent River Commission. (2019). 2019-2020 Patuxent River Commission Annual Action Plan. Approved 9/11/2019. https://planning.maryland.gov/Documents/OurWork/PRC/PRC-Annual-Action-Plan.pdf

Peterson, E.E., A.R. Pearse. (2017). IDW-Plus: An ArcGIS toolset for calculating spatially explicit watershed attributes for survey sites. *Journal of the American Water Resources Association* 53(5): 1241-1249.

Poff, N.L. (1997). Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. *Journal of the North American Benthological Society*, 16(2), 391-409.

Scientific and Technical Advisory Committee (STAC). (2023). Achieving water quality goals in the Chesapeake Bay: A comprehensive evaluation of system response (K. Stephenson & D. Wardrop, Eds.). STAC Publication Number 23-006, Chesapeake Bay Program Scientific and Technical Advisory Committee (STAC), Edgewater, MD. 129 pp.

Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*(4857), 1285-1293.

Tetra Tech, 2020. Inventory & Evaluation of Environmental and Biological Response Data for Fish Habitat Assessment. Report to the Sustainable Fisheries Goal Implementation Team, Chesapeake Bay Program. 68pp. https://d18lev1ok5leia.cloudfront.net/chesapeakebay/documents/FishHabitatInventoryFinalReport2020.pdf

U.S. Geological Survey (USGS) Gap Analysis Project (GAP), 2022, Protected Areas Database of the United States (PAD-US) 3.0: U.S. Geological Survey data release, https://doi.org/10.5066/P9Q9LQ4B.

Vargas-Nguyen, A. Fries, J. Edgerton, B. Dennison, S. Anderson, L. Badri, V. Lucchese, K.M. Laumann, H. Kelsey. (2023). 2022 Chesapeake Bay and Watershed Report Card. 8pp. https://ian.umces.edu/publications/2022-chesapeake-bay-and-watershed-report-card/

Vose, R.S., S. Applequist, I. Durre, M.J. Menne, C.N. Williams, C. Fenimore, K. Gleason, D. Arndt. (2014). Improved Historical Temperature and Precipitation Time Series For U.S. Climate Divisions. Journal of Applied Meteorology and Climatology. DOI: 10.1175/JAMC-D-13-0248.1

Appendices

Table A1. List of Fish Surveys Included in this Study

Metadata in this table includes the lead agency, the source at which the data are maintained/available, the program under which the survey was conducted, the study area, study length, and sampling method.

Agency	Data Source	Program	Primary Geography	Study Area	Study Length	Sampling Method
University of Maryland Center for Environmental Science - Chesapeake Biological Laboratory	Unpublished data, 2023	CBL Seine vs Trawl	Tidal Waters	Choptank, Nanticoke, Patuxent, Upper Bay	2006-2013	Seine
University of Maryland Center for Environmental Science - Chesapeake Biological Laboratory	Unpublished data, 2023	CBL Seine Cruises	Tidal Waters	Choptank, Patuxent, Upper Bay	2011-2013	Seine
Maryland Department of Natural Resources	Juvenile Striped Bass Survey	Juvenile Striped Bass Program	Tidal Waters	Choptank River, Tuckahoe Creek, Nanticoke River, Potomac River, Sassafras River, Worton Creek, Ches Bay, Northeast River, Elk River, Bohemia River, Susquehanna Flats, Patuxent River	1999-2018 collected (entirety 1954-present)	Seine
University of Maryland Center for Environmental Science - Chesapeake Biological Laboratory	https://hjort.cbl.umces.edu/cfdata.html, 2023	TIES CHES-FIMSPAX-FIMS	Tidal Waters	Chesapeake Bay mainstem	2001-2005	Midwater trawl
University of Maryland Center for Environmental Science - Chesapeake Biological Laboratory	https://hjort.cbl.umces.edu/cfdata.html, 2023	TIES CHES-FIMSPAX-FIMS	Tidal Waters	Chesapeake Bay mainstem	1995-2000	Midwater trawl
University of Maryland Center for Environmental Science - Chesapeake Biological Laboratory	https://hjort.cbl.umces.edu/cfdata.html, 2023	TIES CHES-FIMSPAX-FIMS	Tidal Waters	Patuxent River	2004	Midwater trawl and otter trawl
Maryland DNR*	MBSS Data Request Page	MDE 319	Non Tidal Waters	Patuxent River	2012	MBSS
Maryland DNR*	MBSS Data Request Page	Fisheries	Non Tidal Waters	Patuxent River	2014	MBSS
Maryland DNR*	MBSS Data Request Page	MBSS Random	Non Tidal Waters	Patuxent River	1995-2021	MBSS
Maryland DNR*	MBSS Data Request Page	National Park	Non Tidal Waters	Patuxent River	2006	MBSS
Maryland DNR*	MBSS Data Request Page	Restoration Monitoring	Non Tidal Waters	Patuxent River	2011-2019	MBSS
Maryland DNR*	MBSS Data Request Page	Sentinel	Non Tidal Waters	Patuxent River	1997-2021	MBSS
Maryland DNR*	MBSS Data Request Page	Special Project	Non Tidal Waters	Patuxent River	1995-2021	MBSS
Maryland DNR*	MBSS Data Request Page	Targeted	Non Tidal Waters	Patuxent River	2007	MBSS
Maryland DNR*	MBSS Data Request Page	Tier II	Non Tidal Waters	Patuxent River	2007-2018	MBSS

*Data included in this document were provided by the Maryland Department of Natural Resources Monitoring and Non-tidal Assessment Division.

Table A2. Cross Reference of Layers Referenced in Various Assessments

Atlantic Coast Fish Habitat Partnership (ACFHP); Science and Technical Advisory Committee (STAC); The Nature Conservancy (TNC); Stakeholder (Stakeholder)

Layer	ACFHP	STAC	TNC	Stakeholder
Agricultural practices	NO	YES	NO	NO
Benthic inverts	NO	YES	YES	NO
Bottom substrate	NO	YES	NO	NO
Channelization/ditching/dredging	NO	YES	NO	YES
Phytoplankton	NO	YES	NO	NO
Climate change	NO	YES	NO	NO
Critical/protected habitat	YES	NO	NO	NO
Development	YES	YES	NO	YES
DO	NO	YES	YES	NO
Episodic events (droughts, flooding)	NO	YES	NO	YES
Erosion	NO	YES	NO	YES
Eutrophication	NO	YES	NO	NO
Fishing/boating	NO	YES	NO	NO
Flow alteration	YES	YES	NO	NO
Forage	NO	YES	NO	NO
Forest loss	NO	YES	NO	YES
Habitat loss	NO	YES	NO	NO
Harmful algal blooms	NO	YES	NO	YES
Heavy metals	NO	YES	NO	NO
Housing density	NO	YES	NO	NO
Impervious surface	YES	YES	NO	YES
Increased mortality	NO	YES	NO	NO
Invasive species	NO	YES	NO	YES
Land use	YES	YES	YES	NO
Loss of feeding habitat	NO	YES	NO	NO
Loss of riparian vegetation	YES	YES	NO	NO
Nitrogen	NO	YES	NO	YES
Nutrients	NO	YES	NO	YES
Oyster reef loss	YES	YES	YES	NO
Pesticides	NO	YES	NO	NO
Phosphorous	NO	YES	NO	YES
Population shift	NO	YES	NO	NO
Population density	NO	YES	NO	NO
Predator-prey interactions	NO	YES	NO	NO
Range shift	NO	YES	NO	NO
River flow variability	NO	YES	NO	NO
Road crossings	YES	YES	NO	YES
Runoff	NO	YES	NO	NO
Salinity	NO	YES	YES	YES
Sea level Rise	NO	YES	YES	NO
Sedimentation	NO	YES	YES	YES
Septic	NO	YES	NO	NO
Shoreline change/amoring	YES	YES	YES	YES
Source pollution	YES	NO	NO	NO
Species access	YES	NO	YES	NO
Species shifts	NO	YES	NO	NO
Stormwater runoff	NO	YES	NO	YES
SAV	YES	YES	YES	YES
Surface water withdrawal	NO	YES	NO	NO
Temperature	NO	YES	NO	YES
Toxicants	NO	YES	NO	NO
Trophic effects	NO	YES	NO	NO
Turbidity/light	NO	YES	NO	NO
Wastewater treatment plants	NO	YES	NO	NO
Water temperature	NO	YES	NO	YES
Water use/ including withdrawal	NO	YES	NO	NO
Wetland loss	YES	YES	YES	NO
Woody structures	NO	YES	NO	NO

Table A3. Land Use/Land Cover Categories

Used by the Chesapeake Conservancy[***] (LULC Original) and for the Patuxent Pilot.

New Code	Old Code	LULC Original	LULC Patuxent Pilot
2	3	Estuarine/Marine	Estuarine Marine
19	4	Lakes and Reservoirs	Water
19	5	Riverine Ponds	Water
19	6	Terrene Ponds	Water
7	7	Lotic Water (fresh)	Lotic Water
8	8	Bare Shore	Natural Succession
13	9	Roads	Roads
6	10	Structures	Impervious
6	11	Other Impervious	Impervious
16	12	Tree Canopy Over Roads	Tree Canopy over Impervious
16	13	Tree Canopy Over Structures	Tree Canopy over Impervious
16	14	Tree Canopy Over Other Impervious	Tree Canopy over Impervious
17	15	Tree Canopy Over Turf Grass	Tree Canopy over Turf Grass
18	16	Turf Grass	Turf Grass
11	17	Transitional Barren	Pervious Developed
5	18	Harvested Forest Herbaceous	Harvested Forest
11	19	Solar Field Herbaceous	Pervious Developed
3	20	Extractive Barren	Extractive
3	21	Extractive Impervious	Extractive
4	22	Forest	Forest
4	23	Other Tree Canopy	Forest
11	24	Suspended Succession Barren	Pervious Developed
11	25	Suspended Succession Herbaceous	Pervious Developed
11	26	Suspended Succession Scrub/Shrub	Pervious Developed
8	27	Natural Succession Barren	Natural Succession
8	28	Natural Succession Herbaceous	Natural Succession
8	29	Natural Succession Scrub/Shrub	Natural Succession
20	30	Riverine Wetlands Barren	Wetlands, Riverine Non-forested
20	31	Riverine Wetlands Herbaceous	Wetlands, Riverine Non-forested
20	32	Riverine Wetlands Scrub/Shrub	Wetlands, Riverine Non-forested
12	33	Riverine Wetlands Tree Canopy	Riverine Wetlands Forest
12	34	Riverine Wetlands Forest	Riverine Wetlands Forest
21	35	Terrene Wetlands Barren	Wetlands, Terrene Non-forested
21	36	Terrene Wetlands Herbaceous	Wetlands, Terrene Non-forested
21	37	Terrene Wetlands Scrub/Shrub	Wetlands, Terrene Non-forested
14	38	Terrene Wetlands Tree Canopy	Terrene Wetlands Forest
14	39	Terrene Wetlands Forest	Terrene Wetlands Forest
1	40	Cropland Barren	Cropland
1	41	Cropland Herbaceous	Cropland
1	42	Pasture/Hay Barren	Cropland
1	43	Pasture/Hay Herbaceous	Cropland
10	44	Pasture/Hay Scrub/Shrub	Scrub
9	45	Orchard/Vineyard Herbaceous	Orchard
9	46	Orchard/Vineyard Scrub/Shrub	Orchard
22	47	Tidal Wetlands Barren	Wetlands, Tidal Non-forested
22	48	Tidal Wetlands Herbaceous	Wetlands, Tidal Non-forested
22	49	Tidal Wetlands Scrub/Shrub	Wetlands, Tidal Non-forested
15	50	Tidal Wetlands Tree Canopy	Tidal Wetlands Forest
15	51	Tidal Wetlands Forest	Tidal Wetlands Forest

[***] **Chesapeake Bay Program Land Use/Land Cover Data Project**

Table A4. Table of Variables Utilized

Variables that were used in Nested and Non-nested Ensemble Models at each given scale, as selected by variance inflation factor testing (VIF) . 'X' denotes if that variable was used in the final ensemble model at that given resolution and species.

	Basin-wide Nested Model									Tidal-bound Non-nested Model		
	American Eel			Tessellated Darter			White Perch			White Perch		
Variable	1000m	100m	10m	1000m	100m	10m	1000m	100m	10m	1000m	100m	10m
aspect	X	X	X	X	X	X	X	X	X	X	X	X
CMECS_Anthropogenic_Shell_Rubble										X	X	X
CMECS_Biogenic_Oyster_Reef										X	X	X
CMECS_Biogenic_Oyster_Rubble										X	X	X
CMECS_Gravel_Mixes										X	X	X
CMECS_Mud										X	X	X
CMECS_Muddy_Sand										X	X	X
CMECS_Sand										X	X	X
CMECS_Sandy_Mud										X	X	X
CMECS_Unclassified										X	X	X
cropland	X	X	X	X	X	X	X	X	X			
cropland_shed_exp_mosaic			X		X	X						
cropland_shed_per_mosaic	X	X		X			X	X	X	X	X	X
Distance_m_to_combined_SAV_beds										X	X	X
Distance_m_to_hardened_shoreline										X	X	X
Distance_m_to_PADUS3_0										X	X	X
est_marine		X	X		X	X		X	X			
est_marine_fac_exp_mosaic			X		X	X						
est_marine_fac_per mosaic	X	X		X			X	X		X	X	X
est_marine_shed_exp_mosaic		X	X					X	X			
est_marine_shed_per_mosaic	X			X	X	X	X			X	X	X
extractive	X	X		X	X	X	X	X	X			
extractive_fac_exp_mosaic			X		X							
extractive_fac_per_mosaic	X	X		X			X					
extractive_shed_exp_mosaic			X			X		X				
extractive_shed_per_mosaic	X	X		X	X		X		X	X	X	X
flat	X	X	X	X	X	X	X		X	X		
footslope	X	X	X	X	X	X	X	X	X		X	X
forest	X	X	X	X	X	X	X	X	X			
forest_shed_exp_mosaic		X	X			X		X	X			
forest_shed_per_mosaic	X			X	X		X			X	X	X
harvest_forest	X	X		X	X		X	X				
harvest_forest_fac_exp_mosaic	X			X			X	X				
harvest_forest_fac_per_mosaic		X								X	X	
harvest_forest_shed_exp_mosaic	X	X		X	X		X	X				
harvest_forest_shed_per_mosaic										X	X	
harvested_forest			X			X			X			
harvested_forest_fac_per_mosaic			X						X			X
harvested_forest_shed_exp_mosaic						X						
hollow	X	X	X	X	X	X	X	X	X	X	X	X
impervious			X			X			X			
impervious_shed_exp_mosaic			X			X			X			
impervious_shed_per_mosaic												X
impervious_surface	X	X		X	X		X	X				
impervious_surface_shed_exp_mosaic	X			X			X	X				
impervious_surface_shed_per_mosaic		X			X						X	
iso	X	X	X	X	X	X	X					
lotic_water	X	X	X	X	X	X	X	X	X			
lotic_water_fac_per_mosaic										X		
lotic_water_shed_exp_mosaic			X						X			
lotic_water_shed_per_mosaic	X	X		X	X	X	X	X		X	X	X
major_dams_1000m	X			X			X					
major_dams_100m		X			X			X				
major_dams_10m			X			X			X			
minor_dams_1000m	X			X			X					
minor_dams_100m		X						X				
minor_dams_10m			X			X			X			
mtcm								X				
mtwetm	X	X	X	X	X	X	X		X			
mtwq	X	X	X	X	X	X	X			X	X	X
natural_succession			X			X			X			
natural_succession_shed_exp_mosaic			X						X			
natural_succession_shed_per_mosaic						X						X
natural_sucession	X	X		X	X		X	X				
natural_sucession_shed_exp_mosaic					X			X				

	Basin-wide Nested Model									Tidal-bound Non-nested Model		
	American Eel			Tessellated Darter			White Perch			White Perch		
Variable	1000m	100m	10m	1000m	100m	10m	1000m	100m	10m	1000m	100m	10m
natural_sucession_shed_per_mosaic	X	X		X			X			X	X	
orchard	X	X	X	X	X	X	X	X	X			
orchard_fac_exp_mosaic	X			X			X					
Orchard_fac_exp_mosaic		X	X									
Orchard_fac_per_mosaic					X	X		X			X	X
Orchard_shed_exp_mosaic			X									
orchard_shed_ln_mosaic	X			X			X					
orchard_shed_per_mosaic	X			X			X			X		
Orchard_shed_per_mosaic		X			X	X		X	X		X	
pax_con_1000m										X		
pax_con_100m		X						X			X	
pax_con_10m			X			X			X			X
pax_fac_1000m	X			X			X			X		
pax_fac_100m		X			X			X			X	
pax_fac_10m			X			X			X			X
pax_fd8_1000m	X			X			X			X		
pax_fd8_100m		X			X			X			X	
pax_fd8_10m			X			X			X			X
pdm	X	X	X	X	X	X	X	X				
peak	X	X	X	X	X	X	X	X	X			
pervious_developed	X	X	X	X	X	X	X	X	X			
pervious_developed_shed_exp_mosaic	X	X	X	X	X	X	X	X	X			
previous_developed_shed_per_mosaic											X	X
pit	X	X	X	X	X	X	X	X	X			
ps								X	X	X	X	X
pwarmq									X	X	X	X
pwm	X	X	X	X	X	X	X	X				
ridge		X	X		X	X			X		X	X
river_wetland_forest	X	X	X	X	X	X	X	X	X			
river_wetland_forest_shed_exp_mosaic		X	X		X				X			
river_wetland_forest_shed_per_mosaic	X			X		X	X	X		X	X	X
roads	X	X	X	X	X	X	X	X	X			
roads_fac_exp_mosaic						X						
roads_shed_exp_mosaic			X		X	X		X	X			
roads_shed_per_mosaic	X	X		X			X			X		X
rough						X			X			X
roughness	X			X			X			X		
scrub	X	X	X	X	X	X	X	X	X			
scrub_fac_exp_mosaic								X				
scrub_fac_per_mosaic									X		X	
scrub_shed_exp_mosaic		X	X		X							
scrub_shed_per_mosaic						X		X	X		X	X
shoulder		X	X		X	X		X	X	X	X	X
slope	X	X		X			X	X			X	
slope_geom	X		X	X		X	X			X		X
slope_geomorphon		X			X			X			X	
spur	X	X	X	X	X	X	X	X	X		X	X
srad_range		X			X	X		X	X			
srad_range_10												X
srad_sum_10			X			X						X
srad_sum_100		X			X			X			X	
srad_sum_1000										X		
srad_win_10			X			X			X			X
srad_win_100		X			X			X			X	
srad_win_1000	X			X			X					
str_pwr_index	X	X	X	X	X	X	X	X	X	X	X	X
Summer_Surface_DO										X	X	X
Summer_Surface_Salinity										X		
Summer_Surface_Temp										X	X	X
tap	X		X	X	X	X	X					
tar									X		X	X
terr_wetlands			X			X			X			
terr_wetlands_forest	X	X		X	X		X	X				
terr_wetlands_forest_shed_exp_mosaic	X	X		X			X					
terr_wetlands_forest_shed_per_mosaic					X			X			X	
terr_wetlands_shed_per_mosaic			X			X			X			X
tidal_wetlands_forest	X	X	X	X	X	X	X	X	X	X	X	X
tidal_wetlands_forest_fac_per_mosaic						X		X			X	
tidal_wetlands_forest_shed_exp_mosaic		X	X		X	X		X				
tidal_wetlands_forest_shed_per_mosaic	X			X			X		X	X	X	X
topo_wet_index	X	X	X	X	X	X	X	X	X	X	X	X

	Basin-wide Nested Model									Tidal-bound Non-nested Model		
	American Eel			Tessellated Darter			White Perch			White Perch		
Variable	1000m	100m	10m	1000m	100m	10m	1000m	100m	10m	1000m	100m	10m
tpi	X	X	X	X	X	X	X	X	X	X	X	X
tree_over_impervious	X	X	X	X	X	X	X	X	X	X	X	X
tree_over_impervious_shed_exp_mosaic			X		X	X			X			
tree_over_impervious_shed_per_mosaic		X						X		X	X	X
tree_over_turf		X	X		X	X		X	X		X	X
tree_over_turf_fac_per_mosaic		X										
tree_over_turf_shed_exp_mosaic		X	X									
tree_over_turf_shed_per_mosaic	X			X	X	X	X		X			X
tri	X	X	X	X	X		X					
ts	X	X		X	X	X	X			X	X	X
turf	X	X		X	X		X	X		X	X	
turf_shed_exp_mosaic	X			X	X		X	X				
turf_shed_per_mosaic		X								X	X	
valley		X	X		X	X		X	X		X	X
VERSAR_AVE_IBI_Score										X	X	X
water	X	X	X	X	X	X	X	X	X	X	X	X
water_fac_exp_mosaic	X		X	X			X					
water_fac_per_mosaic											X	
water_shed_exp_mosaic		X	X		X			X				
water_shed_per_mosaic	X			X		X	X		X		X	X
wetland_river	X	X	X	X	X	X	X	X				
wetland_river_fac_per_mosaic												X
wetland_river_shed_exp_mosaic					X	X						
wetland_river_shed_per_mosaic	X	X	X	X			X	X	X		X	X
wetland_terr	X	X	X	X	X	X	X	X	X			
wetland_terr_shed_exp_mosaic	X			X			X		X			
wetland_terr_shed_per_mosaic		X	X		X			X		X	X	X
wetland_tidal	X	X	X	X	X	X	X	X	X			
wetland_tidal_fac_exp_mosaic	X			X			X					
wetland_tidal_fac_per_mosaic									X			
wetland_tidal_shed_exp_mosaic		X	X		X							
wetland_tidal_shed_per_mosaic	X			X		X	X	X	X	X	X	X

Table A5. SDM Models included in the Ensemble Package

Bioclim	Bioclimatic classification based on the type of regime approach
Bioclim.dismo	The Bioclim model performed by the Dismo package in R
BRT	boosted regression tree
CART	classification and regression trees
Domain.dismo	Domain (computes the Gower distance between variables)
FDA	factorial discriminant analysis
GAM	generalized additive models
GLM	generalized linear models
GLMnet	generalized linear model fitted with elastic net
GLMpoly	generalized linear model polynomial
Mahal.dismo	Mahalanobis model
MARS	multivariate adaptive regression splines
MaxEnt	Maximum Entropy
MaxLike	Maximum Likelihood
MDA	Multiple Discriminant Analysis
MLP	multi-layer perceptron ensemble
Ranger	The Ranger implementation of Random Forest
RBF	radial basis functions
RF	Random Forest
Rpart	recursive partitioning and regression trees
SVM	Support Vector Machines

Endnotes

1. https://www.atlanticfishhabitat.org/wp-content/uploads/2020/08/ACFHP-Mapping-and-Prioritization-Final-Report.pdf
2. https://www.usgs.gov/national-hydrography/about-national-hydrography-products
3. https://www.usgs.gov/search?keywords=Flow-Conditioned%20Parameter%20Grids
4. https://www.r-project.org/
5. https://www.usgs.gov/special-topics/coastal-national-elevation-database-applications-project
6. World Topo Map Attributed to Esri, DeLorme, HERE, TomTom, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, and the GIS User Community
7. https://coast.noaa.gov/digitalcoast/data/cusp.html
8. https://www.usgs.gov/tools/usgs-chesapeake-bay-studies-data-catalog
9. DOI (USGS Data Release): 10.5066/P13MLH6N. DOI (NOAA NCCOS Data Release): 10.25921/wpwv-zj89
10. https://www.chesapeakeconservancy.org/conservation-innovation-center/high-resolution-data/lulc-data-project-2022/
11. https://prism.oregonstate.edu
12. https://eyesonthebay.dnr.maryland.gov/contmon/ContinuousMonitoring.cfm
13. At the time of publication, the interpolated data were not publicly available from the Chesapeake Bay Program.
14. https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/resample.htm
15. World Topo Map Attributed to Esri, DeLorme, HERE, TomTom, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, and the GIS User Community
16. https://iocm.noaa.gov/standards/cmecs-home.html
17. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/tabulate-area.htm
18. https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/calculate-field.htm
19. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/distance-accumulation.htm
20. https://www.vims.edu/research/units/programs/sav/reports/index.php
21. https://www.vims.edu/ccrm/research/inventory/
22. https://www.usgs.gov/programs/gap-analysis-project/science/pad-us-data-download?qt-science_center_objects=0#qt-science_center_objects
23. https://baybenthos.versar.com/
24. https://eyesonthebay.dnr.maryland.gov/
25. https://www.chesapeake.org/stac/cesr/
26. https://www.chesapeakeconservancy.org/conservation-innovation-center/high-resolution-data/lulc-data-project-2022/

U.S. Department of Commerce

Gina M. Raimondo, *Secretary*

National Oceanic and Atmospheric Administration

Richard W. Spinrad, *Under Secretary for Oceans and Atmosphere*

National Ocean Service

Nicole LeBoeuf, *Assistant Administrator for Ocean Service and Coastal Zone Management*