

GIS-Based Atlantic Salmon Habitat Model

**DRAFT
2008**

Jed Wright¹, John Sweka², Alex Abbott¹, Tara Trinko³

¹U.S. Fish and Wildlife Service, Gulf of Maine Coastal Program

²U.S. Fish and Wildlife Service, Northeast Fishery Center

³NOAA's National Marine Fisheries Service, Maine Field Station

Appendix C

GIS-Based Atlantic Salmon Habitat Model

Jed Wright¹, John Sweka², Alex Abbott¹, Tara Trinko³

Introduction

Fisheries management agencies have traditionally utilized field surveys to develop estimates of Atlantic salmon habitat in Maine rivers. While providing detailed information, field surveys are expensive to conduct and to-date cover only a small portion of the range of historic habitat of Atlantic salmon. A GIS-based habitat model was developed to predict the amount of Atlantic salmon rearing habitat in un-surveyed salmon rivers. The model was developed using data from habitat surveys conducted in the Machias, Sheepscot, Dennys, Sandy, Piscataquis, Mattawmkeag, and Soudabscook Rivers. The model uses reach slope derived from contour and digital elevation model (DEM) datasets, cumulative drainage area, and physiographic province to predict the total amount of rearing habitat within a reach. The variables included in the model explain 73% of the variation in rearing habitat. Maps and data from the model will help inform the proposed listing of critical habitats. This GIS based model will also be used for a variety of management activities including stocking, removing barriers, and prioritizing in-stream habitat restoration projects. The maps below show the extent of the area modeled by the project and detailed GIS output that is available from the model.

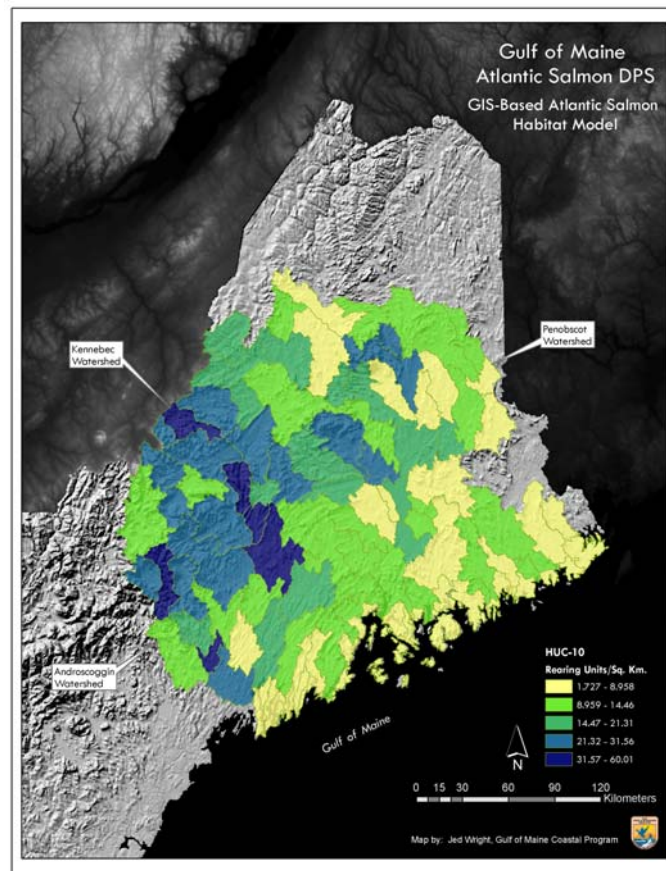


Figure 1: Extent of area included in GIS model.

¹ U.S. Fish and Wildlife Service, Gulf of Maine Coastal Program

² U.S. Fish and Wildlife Service, Northeast Fishery Center

³ NOAA Fisheries Service, Maine Field Station

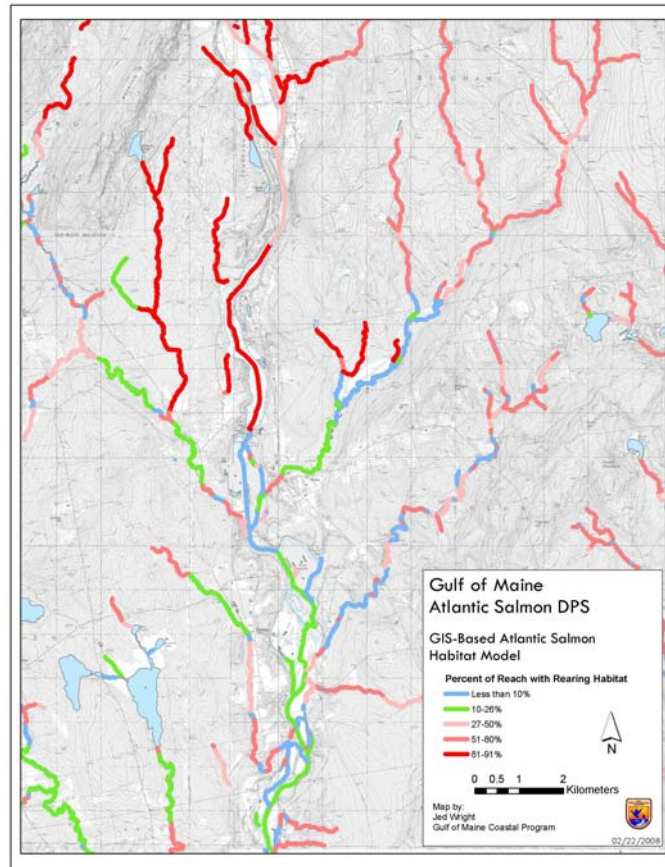


Figure 2: The GIS model predicts the amount of habitat within each stream reach.

Methods

Stream Segment Selection Methods

ArcGIS software version 9.2 (Environmental Systems Research Institute 2006) was used to process datasets used in the analysis. The National Hydrography High Resolution Dataset (NHDH) was used to identify potential habitat within the expanded Atlantic Salmon Gulf of Maine Distinct Population Segment (GOM-DPS). NHDH flowlines that were either perennial streams/rivers or that were located within 1:24,000 double line river segments were selected for use in the model. Selected stream and river reaches were then dissolved by GNIS-ID using the dissolve command in ArcTool Box. XTools Pro Version 5.0 was used to convert multipart selected stream segments to single parts and editing was conducted to remove short artificial path segments.

Stream Segment Slope Determination

Using the selected stream segments, XTools Pro was used to split the selected set of NHDH polylines with a 1:24,000 contour coverage. X Tools Pro was then used to create TO and FROM endpoints from the newly split line segments. A spatial join was used to obtain an elevation value for the TO and FROM points from contour lines. In addition, a distance to the nearest contour line was calculated for each point. Hawth's Tools Version 3.27 was then used to obtain digital elevation model (DEM) elevation values to each TO and FROM point. DEM values were obtained from both 10 and 30 meter DEM datasets as a 10 meter DEM was not available for the entire study area. After values had been obtained from contour and DEM datasets, a final elevation was calculated for each point.

A point located within 1 meter of the nearest contour line was given a final elevation based on contour values. All remaining points were then coded with a final elevation of the corresponding DEM value, 10 meter values were used if available otherwise 30 meter DEM dataset values were used. Final elevations were calculated in meters.

TO and FROM points were joined by attribute back to corresponding selected NHDH stream segments based on either From ID or To ID and Object ID. The NHDH line was then coded with the FROM and TO elevation of the points. A field was added to NHD lines called "vertical" and a value was calculated as FROM elevation- TO Elevation. All lines were then examined for negative slopes and edited for errors. In addition, segments that intersected contour lines multiple times or segments that intersected contour lines and identical FROM and TO values were dissolved. Finally, a "slope" field was added to the selected NHD stream segments and calculated as $(\text{Vertical} / \text{ShapeLength}) * 100$ to give the percent slope. All data sets were edited to contain less than 5% negative or zero slopes as calculated by total stream length. All negative and zero slope values were removed from the data set for later regression analyses. A final processing step involved identifying reaches that were located in tidal river reaches. National Wetlands Inventory (NWI) datasets were used to select and delete reaches that were located in either in estuarine or riverine tidal areas. The final reach dataset included over 148,010 reaches.

Cumulative Drainage Area

The original dataset used to develop the habitat model used Arc Hydro for ArcGIS 9 (version 1.1) and both 10 and 30 meter DEMs to obtain a cumulative drainage area for the downstream end of each reach. Unfortunately, there was not enough processing time to create cumulative drainage areas for almost 150,000 points representing the downstream end of each potential habitat line segment. Instead, cumulative drainage area was calculated where possible for all segments using NHDPlus datasets (<http://www.horizon-systems.com/nhdplus/index.php>). NHDPlus provides a cumulative drainage area (as well as other attributes such as flow and Strahler stream order) for each reach through a tabular join (through the ComID field) to the flowlineattributesflow.dbf table.

Cumulative drainage area was calculated (in the CumDrnSqKM field) for each potential habitat segment where its original NHDH ReachCode matched the ReachCode of the NHDPlus lines. Each of the matching lines received a MatchCode of 1 for ease of identification throughout processing. All line segments were run through the FLoWs (Colorado State University; v. 9.2) Snap Points to Landscape Network Edges Pre-Processing tool using ArcGIS 9.2 software to assign a reach identifier (rid) and a distance ratio value (ratio) to the centroids of each potential habitat segment. FLoWs snaps each input point within a specified distance to the NHDPlus lines ("Network Edges"), and gives the ratio of the distance that point sits along the NHDPlus reach line from downstream to upstream. To avoid the large number of errors that can occur when the tool snaps points to the lines the downstream TO points were not used as inputs to the tool. Instead, the segments' centroids were substituted. There is a difference in distance between the TO points and the centroids of the same line segments and this process provides only the approximate ratio of the distance of each TO point along the original reach line. Yet, as there are normally several potential habitat line segments within each NHDPlus reach, this process provides a reasonable ratio of the distance for use in calculating cumulative drainage areas.

The next step was to assign catchment areas to each NHDPlus reach through a join to the NHDPlus catchment shapefile via the ComID field. The ratio calculated above was then used to calculate the segments' approximate catchment area, take its inverse, and subtract that from the CumDrnSqKM value for each segment with a MatchCode = 1, but not including any headwater stream segments with a ratio > 0.1 (these segments are generally in smaller catchments that receive the default cumulative drainage area value applied to other segments without matching NHDPlus reaches). A selection was made of all segments of MatchCode = 1 AND CumDrnSqKM = Catchment AND Ratio > 0.1, and all selected records had a new cumulative drainage area field, CumDrain2, calculated = -99 (No Data). The selection was then switched to its reciprocal, and values calculated using the formula:

$$\text{CumDrain2} = \text{CumDrnSqKM} - (\text{Ratio} * \text{Catchment})$$

Next, all records of MatchCode not equal to 1 were selected and calculated = -99. Finally, a new field, DA, was calculated to hold the value of cumulative drainage area in square miles.

Cumulative drainage area for all streams without matching NHDP ReachCodes (MatchCode = -99) were set at a fixed value of one square mile after calculation of sample drainage areas from various watersheds within the SHRUs.

Reach Width

A width for each stream reach was calculated using regional hydraulic geometry curves for Maine rivers based on Dudley (2004) and the cumulative drainage area obtained from the steps outlined above.

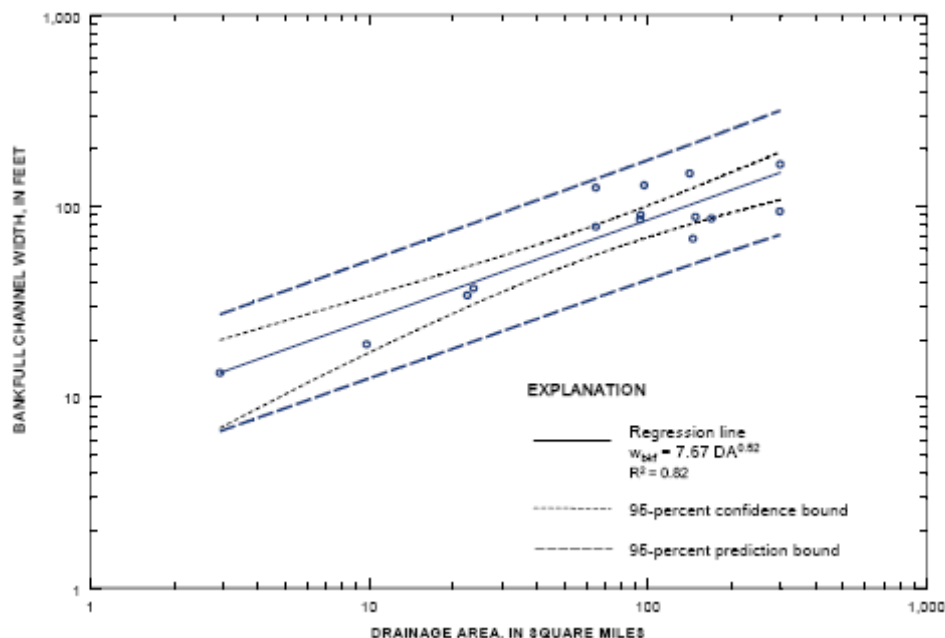


Figure 3: Regional relation of bankfull channel width to drainage area for rivers in coastal and central Maine. [wbkf, channel width associated with the bankfull streamflow; DA, drainage area; R², fraction of variance explained by regression] (Dudley 2004).

A cursory analysis was undertaken to examine the relationship between predicted bankfull widths and widths measured in the field during habitat surveys. This

examination showed that habitat widths were approximately 80% of predicted bankfull widths.

Physiographic Provinces

Maine Atlantic salmon rivers span a diverse set of geologies, climates and elevations. In order to account for these differences we incorporated a physiographic variable into the model. Each river reach was classified by physiographic divisions based on Fenneman, N.M., and Johnson, D.W. (1946).

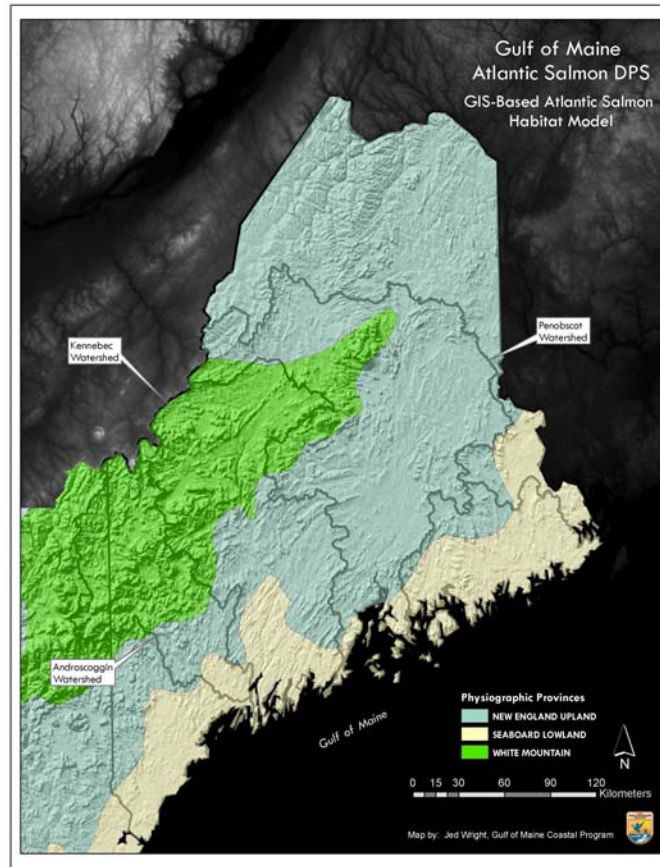


Figure 4: Physiographic provinces included in GIS model.

Final Dataset

The final dataset included the following variables:

Variable	Definition
Unique ID	A unique ID for each stream reach in the SHRU
Source	Elevation source (DEM or contour)
Physiographic Province	Physiographic province from Fenneman
HUC10_Code	USGS HUC 10 code
Length	Length of each reach in meter
Reach Slope	Slope calculated from vertical elevation and reach length
Cumulative Drainage Area	Drainage area in square meters at downstream end of reach
Width	80% of width calculated using regional hydraulic

	geometry curves and cumulative drainage area
Access	N if the reach was not historically accessible to salmon

Regression Tree Analysis

Regression tree analysis is a modern statistical technique that has advantages over classical multiple regression techniques in that there are no assumptions about the error structure of the data and is robust to highly correlated predictor variables (De' Atth and Fabricius 2000). The regression tree is constructed by repeatedly splitting the data into two mutually exclusive groups which are as homogeneous as possible. A group of data is referred to as a node and nodes are further split into additional nodes creating a graphical tree explaining the variability in the data. For numeric predictor variables, the values of a predictor are ranked and trial splits are made moving across all possible division points. The variance of the resulting nodes is calculated and the splitting point which results in the most homogeneous groups (minimized variance) is retained. This process is then repeated for each of the other predictor variables and the best split for any predictor variable is used to perform the actual split on the node. Thus, the optimal split on any given node may be performed by any one of the predictor variables. The regression tree process can result in an overly complex tree as resulting nodes are split further and further. Breiman et al. (1984) recommended V-fold cross-validation as a means to find the best single tree for description and predictive purposes.

The computer software DTREG® (Sherrod 2006) was used to build the regression tree describing the variation in percent rearing habitat within a stream reach. A total of 332 stream reaches were used in the analysis. Predictor variables included valley width cumulative drainage, reach slope, and physiographic province. An initial split based on physiographic province was specified in the model because of the apparent differences between streams of different physiographic provinces.

The optimal tree based on V-fold cross validation contained predictor variables of physiographic province, cumulative drainage area, and reach slope and explained 73% of the variation in percent rearing habitat (Figure 5). Valley width was dropped from the set of predictors because it provided little additional explanatory power. The final tree contained 12 terminal nodes. In general, there was a tendency for percent rearing habitat to increase with greater slope, but there was also an apparent interaction between reach slope and cumulative drainage area (Figure 5).

This model was then used to predict the percent rearing habitat and absolute amount of rearing habitat in 148,010 reaches throughout Maine rivers. Predictions of percent rearing habitat were made by running the data through the DTREG® software and assigning each reach to one of the terminal nodes of the regression tree. The absolute amount of habitat in a reach was estimated by multiplying the area (length x mean width) of the stream reach by the mean percent rearing habitat of the terminal node. The variance associated with the estimate of rearing habitat equaled the variance of the terminal node (Standard Deviation in Figure 5 squared) multiplied by the area² of the reach. The total rearing habitat within river basins was estimated by summing estimates of reach habitat and associated variances.

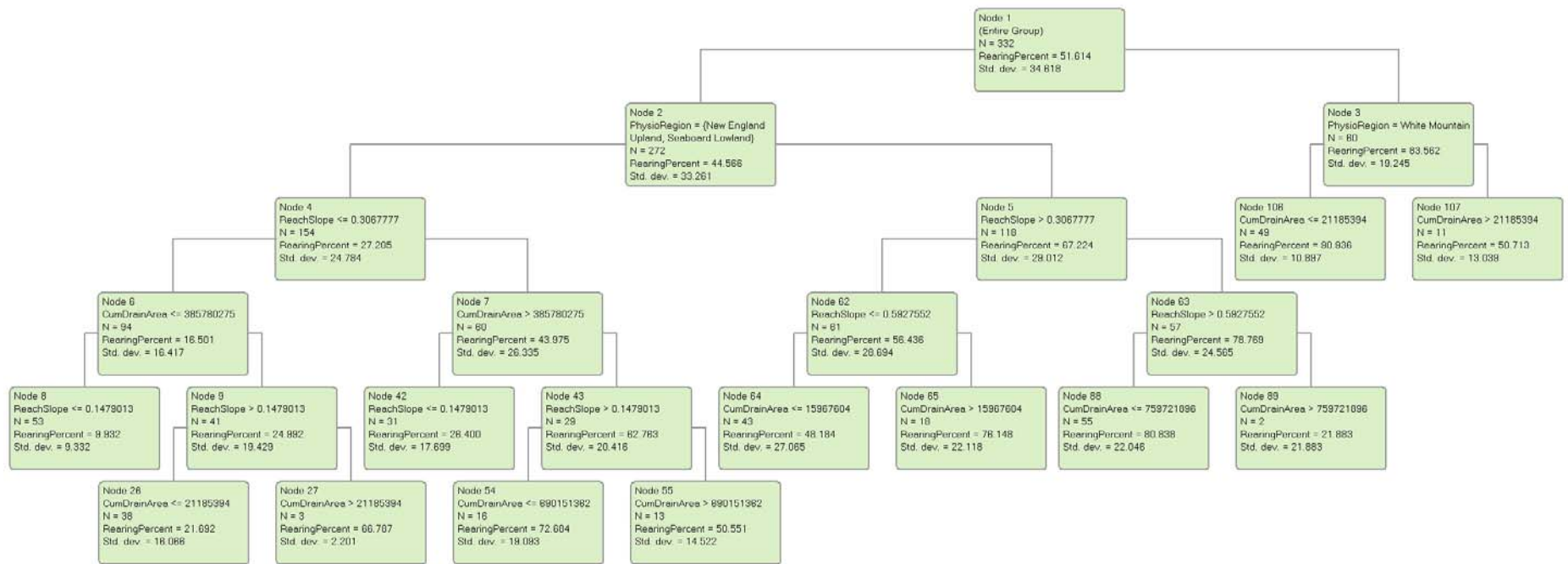


Figure 5: Regression tree model to predict the percent rearing habitat in a steam reach. The model explained 73% of the variation in percent rearing habitat from the 332 reaches used to create the model.

References:

Breiman et al. 1984. Classification and Regression Trees. Wadsworth & Brooks/Cole.

De'ath G, Fabricius KE (2000) Classification and regression trees: a powerful yet simple technique for the analysis of complex ecological data. *Ecology* 81 (11): 3178–3192

Dudley, Robert W. —Hydraulic-Geometry Relations for Rivers in Coastal and Central Maine—USGS Scientific Investigations Report 2004-5042

Fenneman, N.M. and Johnson, D.W. (1946). Physical Division of the United States. U.S. Geological Survey, scale 1:7,000,000. Available at <http://tapestry.usgs.gov/physiogr/physio.html>.