

Appendix 3

Habitat Suitability Criteria

This appendix describes the details of determining habitat suitability criteria for the selected species. These criteria were used to evaluate the habitat quality in the areas mapped during this project. They were established based on empirical fish capture data of adult resident fish in summer. For each species we identified criteria specifying not-suitable, suitable and optimal habitat. Habitat models were created for American eel, Atlantic salmon, blacknose dace, brook trout, common shiner, fallfish, longnose dace, tessellated darter, and white sucker.

Empirical data based model

The empirical set of criteria for rearing and growth (R&G) season had been developed from habitat use data collected in earlier studies. We analyzed fish habitat data from several locations where resident 1+ and older, American eel, Atlantic salmon, blacknose dace, brook trout, common shiner, fallfish, longnose dace, tessellated darter, and white sucker were captured. Data were collected on the Eightmile River (345 grids) and Fenton River (506 grids) in Connecticut as well as 455 grids collected from Stony Clove (269), Roundout River (106), Trout Brook (24), Spring Brook (24), Stewart Brook (16) and Willowemoc (16) in New York. Fish data collected on the Pomperaug River in Connecticut (194 grids) were used to supplement these data to generate the models for fallfish due to a lack of fallfish data from these previous rivers. For each species we used only data from rivers where they occurred in significant numbers (more than 5 individuals captured). We used a multivariate statistical model (logistic regression) to compute the habitat selection criteria for adult resident fish species and Atlantic salmon. At each grid the physical attributes of the HMU in which it was located were recorded together with the number of individuals and species captured.

To calculate the response functions for the species above, we described each grid that was sampled during the survey in terms of the same environmental characteristics used to develop the habitat database, as well as by the species presence and abundance. The environmental attributes were independent variables and the species were dependent variables in regression models describing habitat preference. We employed a stepwise forward logistic regression model (using SPSS) to identify the characteristics of habitat that is used versus habitat that is not used by each fish species. The model uses likelihood ratios to determine which parameters should be included in the following regression formula:

$$R=e^{-z}$$

where:

- e = natural log base
- $z = b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_n \cdot x_n + a$
- $x_{1..n}$ = significant physical variables
- $b_{1..n}$ = regression coefficients
- a = constant

We established the b coefficients using an iterative process. The b coefficients represent how the attributes influence the dependent variable in the data set. This mathematical formula is based on proportions of occupied/non-occupied areas observed during the survey and do not capture all the possible circumstances or represent mechanisms of fish behavior. Furthermore, fish presence is due largely in part to a combination of environmental factors; interpreting the influence of individual parameters in the formula is therefore somewhat limited in applicability.

To distinguish suitable habitat, we used binary dependent variables indicating presence and absence and, in a second model, high and low abundances. The fish data was separated to low and high abundance classes. The cut off value was calculated from observed abundances per grid and was different for each species depending on their behavior (solitary vs. gregarious) and size. For blacknose dace and white sucker three or more fish indicated high abundance. For common shiner, fallfish, longnose dace and tessellated darter two or more individuals were needed. For brook trout, American eel and Atlantic salmon the presence of more than one indicated high abundance. We used all the available data for the presence model, and for the abundance, we used only data from grids in which fish were caught. From the output of the logistic regression function, we obtained two important types of information: the environmental attributes that significantly correspond with species presence and abundance and the regression coefficients B-values. The B-values indicate the strength and direction (+ or -) of the association between each habitat attribute and fish presence and level of abundance.

In a subsequent step we determined the predictive strength of the model as well as identified thresholds between predictions for suitable and not suitable habitat by comparing probabilities of fish presence and high abundance for each grid and actual observations. The following procedure was used:

For each mesohabitat mapped during the biological survey, we calculated the probability of fish presence using computed regression equations and the following formula:

$$p = \frac{e^z}{(1+e^z)}$$

Where:

- p = probability of presence/high abundance
- e = constant
- $z = b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_n \cdot x_n + a$
- $x_{1..n}$ = significant physical variables
- $b_{1..n}$ = regression coefficients
- a = constant

We calculated the probability of presence and of high abundance for every species. The observed presence and abundance at each grid was associated with the probability for the HMU where the grid was located. We created a relative operating

characteristic (ROC) curve for presence and abundance predictions (Metz, 1978). The curve examines the discrimination performance of the model over a range of threshold levels, by plotting proportion of grids correctly predicted to be occupied (sensitivity or true positive rate), and the proportion of grids incorrectly predicted to be occupied (false positive rate). The area under the ROC curve defines the discrimination capacity of the model based on Mann-Whitney statistics (Pearce & Ferrier 2000). The inflection points on the ROC curve allow one to define the probability (Pt) that has the highest true positive rate and lowest false positive rate and therefore best separation of occupied and not occupied areas. Separate Pt values are selected for the presence and abundance models. In the following assessment the habitats with a probability of presence greater than Pt were classified as suitable. The suitable habitats with a probability of high abundance greater than selected Pt are deemed optimal. The areas under the curve and Pt values were selected and presented in the results section together with a list of significant parameters and B-values for both the presence and abundance models. The model was then applied to the data from the mapping survey to identify suitable and optimal habitat areas.

For the young-of-the-year (YOY) fish life stage habitat, which consists only of shallow margins, empirical criteria developed on the Quinebaug River were applied. Areas designated as shallow margins had an average depth of 12 cm (SD = 6 cm), and an average velocity of 15 cm.s⁻¹ (SD = 11). Substrate in these areas was generally small, ranging from sand to meso-lithal. Shallow margins are an attribute of a HMU and are mapped either as present or abundant. HMUs with abundant shallow margins were considered optimal.

Results

Table 1 represents attributes of both models for American eel established from 345 grids including 116 grids where American eel was captured and 38 grids with a high abundance of this species. The presence model indicates a negative correlation with pool HMU types, and mud and sand substrates. The abundance model shows a positive correlation with boulders, and a negative correlation with sand.

Table 1: Physical attributes correlating with presence and high abundance of American eel. The Area Under ROC curve is a measure of the discrimination capacity of the model. Selected cut-off indicates the probability separating un-suitable, suitable and optimal habitats. B represents regression coefficients of logistic regression model.

American eel			
Presence Model		Abundance Model	
Area under ROC curve	0.686	Area under ROC curve	0.668
Selected cut-off	0.38	Selected cut-off	0.31
Variables in the Equation	B	Variables in the Equation	B
Choriotop: Pelal	-5.777	Boulders	0.598
Choriotop: Psammal	-2.362	Choriotop: Psammal	-3.911
HMU Type: Pool	-1.009	Constant	-1.164

Table 2 represents attributes of the model for Atlantic salmon presence established using empirical habitat data from 345 grids including 34 grids where Atlantic salmon were captured. The presence model indicates a positive correlation between the presence of this species and riffle HMU types, MicroLithal choriotop (gravel substrate), and average Froude numbers (i.e. riffle habitat). Do to a low number of instances when this species was captured in high abundances resulting in a lack of empirical data for Atlantic salmon abundance, development of an abundance model is not possible for this species.

Table 2: Physical attributes correlating with presence of Atlantic salmon. The Area Under ROC curve is a measure of the discrimination capacity of the model. Selected cut-off indicates the probability separating un-suitable and suitable habitats. B represents regression coefficients of logistic regression model.

Atlantic salmon	
Presence Model	
Area under ROC curve	0.804
Selected cut-off	0.08
Variables in the Equation	B
HMU Type: Riffle	2.034
Choriotop: MicroLithal	2.712
Average Froude	3.345
Constant	-5.080

Table 3 represents attributes of both models for blacknose dace established from 851 grids from all rivers including 386 grids where blacknose dace was captured and 225 grids with high abundance of this species. The presence model shows a negative correlation with submerged vegetation, canopy shading, MegaLithal choriotop, and higher velocities. There is a positive correlation between the presence of this species and average Froude number. The abundance model shows a positive correlation with water velocity and glide HMU types, and a negative correlation with water depths between 50 and 75 cm deep.

Table 3: Physical attributes correlating with presence and high abundance of blacknose dace. The Area Under ROC curve is a measure of discrimination capacity of the model (0-1). Selected cut-off indicates the probability separating un-suitable, suitable and optimal habitats. B represents regression coefficients of the logistic regression model.

blacknose dace			
Presence Model		Abundance Model	
Area under ROC curve	0.790	Area under ROC curve	0.658
Selected cut-off	0.4	Selected cut-off	0.58
Variables in the Equation	B	Variables in the Equation	B
Current Velocity: 75-90 cm/s	-1.979	Current Velocity: 30-45 cm/s	1.094
Average Froude	0.981	Water Depth: 50-75 cm	-2.188
Submerged Vegetation	-0.671	HMU Type: Glide	0.934
Canopy Shading	-0.569		
Choriotop: MegaLithal	-1.489		
Constant	-0.605		

Table 4 represents attributes of both models for brook trout established from 963 grids including 122 grids where brook trout was captured and 17 grids with high abundance of this species. The presence model consists of a number of habitat attributes which correlate both positively and negatively with the presence of this species. Canopy shading and MacroLithal choriotop are both positively correlated with the presence of brook trout, while there are negative correlations with riprap, moderate to fast flowing water, and MicroLithal choriotop. The abundance model shows positive correlations with riffles, woody debris, and deep, moderately flowing water conditions. There is a negative correlation with higher velocities, and shallower depths.

Table 4: Physical attributes correlating with presence and high abundance of brook trout. The Area Under ROC curve is a measure of the discrimination capacity of the model. Selected cut-off indicates the probability separating un-suitable, suitable and optimal habitats. B represents regression coefficients of logistic regression model.

brook trout			
Presence Model		Abundance Model	
Area under ROC curve	0.825	Area under ROC curve	0.759
Selected cut-off	0.09	Selected cut-off	0.52
Variables in the Equation	B	Variables in the Equation	B
Current Velocity: 45-60 cm/s	-2.573	Current Velocity: 15-30 cm/s	5.990
Current Velocity: 60-75 cm/s	-4.475	Current Velocity: 30-45 cm/s	-6.017
Rip Rap	-0.742	Water Depth: 25-50 cm	-4.131
Canopy Shading	0.832	Water Depth: 75-100 cm	10.074
Choriotop: MacroLithal	1.184	Woody Debris	2.308
Choriotop: MicroLithal	-3.087	HMU Type: Riffle	3.100
Constant	-1.794	Constant	-4.294

Table 5 represents attributes of the presence model for common shiner established from 851 grids including 99 grids where common shiner was captured. The presence model indicates a positive correlation between the presence of this species and shallow margins, woody debris, MicroLithal and MegaLithal choriotops, and run HMU types. The model indicates negative correlations with submerged vegetation and shallow water. Due to the low number of instances when this species was captured in high abundances there is a lack of empirical data for common shiner abundance. Development of an abundance model is not possible for this species.

Table 5: Physical attributes correlating with presence of common shiner. The Area Under ROC curve is a measure of the discrimination capacity of the model. Selected cut-off indicates the probability separating un-suitable and suitable habitats. B represents regression coefficients of logistic regression model.

common shiner	
Presence Model	
Area under ROC curve	0.704
Selected cut-off	0.10
Variables in the Equation	B
Water Depth: 0-25 cm	-0.921
Woody Debris	0.253
Submerged Vegetation	-0.478
Shallow Margins	0.297
Choriotop: MegaLithal	2.018
Choriotop: MicroLithal	1.358
HMU Type: Run	0.516
Constant	-2.758

Table 6 represents attributes of both models for fallfish established from 1,045 grids including 155 grids where fallfish was captured and 95 grids with a high abundance of this species. The presence model indicates negative correlations with certain water velocities, submerged vegetation, and glide HMU types and positive correlations with another range of water velocities, moderate depth, rip rap, shallow margins, and Akal, MegaLithal, and Pelal choriotops. The abundance model shows a negative correlation with MesoLithal choriotop.

Table 6: Physical attributes correlating with presence and high abundance of fallfish. The Area Under ROC curve is a measure of the discrimination capacity of the model. Selected cut-off indicates the probability separating un-suitable, suitable and optimal habitats. B represents regression coefficients of logistic regression model.

fallfish			
Presence Model		Abundance Model	
Area under ROC curve	0.745	Area under ROC curve	0.600
Selected cut-off	0.14	Selected cut-off	0.65
Variables in the Equation	B	Variables in the Equation	B
Current Velocity: 0-15 cm/s	-0.986	Choriotop: MesoLithal	-1.630
Current Velocity: 60-75 cm/s	-8.896	Constant	0.930
Current Velocity: 90-105 cm/s	11.888		
Water Depth: 25-50 cm	0.988		
Submerged Vegetation	-0.581		
Rip Rap	0.495		
Shallow Margins	0.305		
Choriotop: Akal	2.867		
Choriotop: MegaLithal	1.401		
Choriotop: Pelal	2.346		
HMU Type: Glide	-1.107		
Constant	-2.011		

Table 7 represents attributes of both models for longnose dace established from 851 grids including 38 grids where longnose dace was captured and 8 grids with high abundance of this species. The data from the Fenton River was not used because this river is outside of the zoogeographic range of this species. The presence model consists of a number of habitat attributes that correlate positively with the presence of this species such as water velocities, and HMU types (e.g. riffles and ruffles). The abundance model also shows a positive correlation between low water velocities and the abundance of longnose dace.

Table 7: Physical attributes correlating with presence and high abundance of longnose dace. The Area Under ROC curve is a measure of the discrimination capacity of the model. Selected cut-off indicates the probability separating unsuitable, suitable and optimal habitats. B represents regression coefficients of logistic regression model.

longnose dace			
Presence Model		Abundance Model	
Area under ROC curve	0.890	Area under ROC curve	0.693
Selected cut-off	0.015	Selected cut-off	0.78
Variables in the Equation	B	Variables in the Equation	B
Current Velocity: 15-30 cm/s	2.613	Current Velocity: 0-15 cm/s	3.080
Current Velocity: 75-90 cm/s	3.774		
Current Velocity: 90-105 cm/s	12.029		
Water Depth: 25-50 cm	-1.934		
HMU Type: Riffle	3.238		
HMU Type: Ruffle	3.559		
Constant	-5.652		

Table 8 represents attributes of both models for tessellated darter established from 851 grids including 191 grids where tessellated darter was captured and 74 grids with a high abundance of this species. The presence model indicates positive correlations between the presence of this species and shallow margins and run HMU types. The model indicates negative correlations with depths between 75 and 100 cm and water velocities between 45 and 60 cm/s. The abundance model shows a correlation between slow water velocities and the abundance of this species.

Table 8: Physical attributes correlating with presence and high abundance of tessellated darter. The Area Under ROC curve is a measure of the discrimination capacity of the model. Selected cut-off indicates the probability separating unsuitable, suitable and optimal habitats. B represents regression coefficients of logistic regression model.

tessellated darter			
Presence Model		Abundance Model	
Area under ROC curve	0.943	Area under ROC curve	0.658
Selected cut-off	0.15	Selected cut-off	0.36
Variables in the Equation	B	Variables in the Equation	B
Current Velocity: 45-60 cm/s	-2.146	Current Velocity: 0-15 cm/s	1.013
Water Depth: 75-100 cm	-9.134	Constant	-0.831
Shallow Margins	0.455		
HMU Type: Run	0.402		
Constant	-1.726		

Table 9 represents attributes of both models for white sucker established from 851 grids including 167 grids where white sucker was captured and 15 grids with high abundance of this species. The presence model indicates a positive correlation between the presence of this species and water depths between 25 and 50 cm. The model indicates a negative correlation with water velocities between 60 and 75 cm/s. The abundance model specifies a positive correlation with submerged vegetation.

Table 9: Physical attributes correlating with presence and high abundance of white sucker. The Area Under ROC curve is a measure of the discrimination capacity of the model. Selected cut-off indicates the probability separating un-suitable, suitable and optimal habitats. B represents regression coefficients of logistic regression model.

white sucker			
Presence Model		Abundance Model	
Area under ROC curve	0.672	Area under ROC curve	0.620
Selected cut-off	0.16	Selected cut-off	0.15
Variables in the Equation		Variables in the Equation	
	B		B
Current Velocity: 60-75 cm/s	-4.385	Submerged Vegetation	0.597
Water Depth: 25-50 cm	0.938	Constant	-2.612
Constant	-1.700		

Discussion

The empirical-based models presented here have a high capacity to discriminate between occupied and not occupied habitats, which is indicated by high areas in the under ROC curves. However, the B-coefficients should not be interpreted as representing the affinity of species with specific environmental attributes alone, but rather, as factors affecting the shape of the regression lines. The models include parameters that correspond with empirical expectations. For example, fluvial specialists show a clear affinity towards fast flowing, riffle habitats. The habitat for fluvial dependent species such as white sucker and common shiner is characterized by swift but deeper areas. Brook trout habitat is diverse but also clearly associated with fast flowing areas. Limited abundances of species such as common shiner and Atlantic salmon do not allow for the development of precise models. The data was not sufficient to identify optimal habitats for these species.

References

Metz, C.E., 1986. ROC methodology in radiologic imaging. *Invest. Rad.* 21, 720–733.

Pearce, J. & S. Ferrier. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, Volume 133(3):225-245.