Appendix 10

Habitat Suitability Criteria

Prepared by:

Rushing Rivers Institute

December 2009

Habitat Suitability

This appendix describes the details of determining habitat suitability criteria for selected species. These criteria were used to evaluate habitat quality within the study area and across the range of flow conditions mapped during this project. Habitat suitability criteria were established based on empirical data (for adult resident fish during summer), as well as literature reviews (spawning life stage). For each species, we identified criteria specifying not-suitable, suitable, and suitable-optimal habitat. Habitat models were created for each of the selected indicator species (common shiner, American eel, common white sucker, longnose dace, blacknose dace, brook trout and eastern pearlshell).

Methodology

The empirical set of criteria for rearing and growth (R&G) season had been developed from habitat use data collected in earlier studies. We developed a Microsoft Access database that included all the fish samples collected by our program in the Northeast. It includes observations from 15 rivers, presented in **Table 1**.

Table 1: Data sources used for calculation of logistic regression models.

PrjID	Name	# Grids	Description	Species used
1	Upper Souhegan Fish Data	91	Fish data for the Upper Souhegan River including species caught, individual lengths, grid habitat data, and fishing HMU's habitat data.	Longnose dace, White sucker, Fallfish, Blacknose dace
2	Upper West Branch Swift River	100	Fish data for the upper section of the West Branch Swift River in New Salem/Shutesbury, MA. Includes species, grid, and HMU habitat attributes and hydraulic data measurements.	Blacknose dace, Brook trout
3	Lower Souhegan Fish Data	33	Lower Souhegan River fish, grid, and fishing HMU data for Site 7 grids. Snorkeling survey for Sites 6, 8, 9, 10, and 11 include fish species with nearest flow HMU mapping.	Common shiner, Longnose dace, White sucker, Fallfish
4	Fort River Fish Data	81	Fish Data collected from the Fort River, Amherst in July of 2006. Fish were collected from the two case study sites of the MesoHABSIM 2006 Summer Course. Data includes species, grid, and HMU habitat data, and hydraulic measurements.	Common shiner, Longnose dace, White sucker, Fallfish, Tessellated darter, Brook trout
5	Pomperaug River	90	Fish Data collected from the mainstem of the Pomperaug River,	Common shiner, Longnose dace, White

PrjID	Name	# Grids	Description	Species used
			CT. Data includes fish species, lengths, grid, and HMU Habitat data, and hydraulic measurements.	sucker, Fallfish, Blacknose dace, Tessellated darter, Brook trout
6	Nonnewaug River	60	Fish data collected in 2004 on the Nonnewaug River (Upper Pomperaug Watershed), CT. Data includes fish species, lengths, HMU and grid habitat data, and hydraulic measurements.	Longnose dace, White sucker, Blacknose dace, Tessellated darter, Brook trout
7	Weekeepeemee River	7	Fish data collected in 2004 on the Weekeepeemee River, CT. Data includes species, lengths, grid, HMU habitat data and hydraulic measurements.	Longnose dace, White sucker, Blacknose dace, Tessellated darter, Brook trout
8	Lower Eightmile River Mainstem	97	Fishing survey data collected from the Lower Eightmile River mainstem, Connecticut in July of 2004. Data includes species caught, grid and HMU habitat data, and hydraulic data.	Common shiner, White sucker, Fallfish, American eel, Blacknose dace, Tessellated darter, Brook trout
9	Fenton River	508	Fish data collected on the Fenton River, Connecticut 2003.	Common shiner, White sucker, Fallfish, Brook Trout, Blacknose dace, Tessellated darter,
10	East Branch Eightmile River	117	Fishing survey data collected from the mainstem of the East Branch Eightmile River, Connecticut in July and August of 2004.	Common shiner, Longnose dace, White sucker, American eel, Blacknose dace. Tessellated darter, Brook trout
11	Upper Mainstem Eightmile	72	Fishing survey data collected from the Upper Mainstem Eightmile River, Connecticut in July and August of 2004.	Common shiner, Longnose dace, Fallfish, American eel, Blacknose dace, Tessellated darter
12	Stony Clove Creek	269	Fishing survey data from the Stony Clove Creek, NY (mainstem) collected in July of 2002.	Longnose dace, White sucker, Blacknose dace, Brook trout
13	Round Out River	106	Fishing survey data from Round Out,	Blacknose dace,

PrjID	Name	# Grids	Description	Species used
			NY collected in July of 2002.	Brook trout
14	Stewart Brook Fishing Survey Data 2002	16	Fishing survey data from Stewart Brook, NY collected in August of 2002.	Common shiner, Longnose dace, White sucker, Blacknose dace, Brook trout
15	Spring Brook	24	Fishing survey data from Spring Brook, NY collected in 2002.	Longnose dace, Blacknose dace, Brook trout

For each fish species (individuals one year of age and older), we analyzed habitat data obtained from rivers where the species has been observed in abundance higher than 5% of total observations of these species. We used a multivariate statistical model (logistic regression) to compute the habitat selection criteria for adult resident fish species and for Atlantic salmon. At each grid, the physical attributes of the HMU were recorded along with the species captured and their count.

To calculate the response functions for each species, we analyze the environmental characteristics at each grid location to describe 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 logistic regression model to identify the characteristics of habitat occupied, versus unused habitat, for each fish species. The model uses Akaike information criterion (Sakamoto et al 1986) to determine which parameters should be included in the following regression formula:

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

Where:

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

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 relative fish presence. Because the selection of attributes may have critical influence on modeling results, we applied very rigorous procedures for this process to increase model certainty.

In the first step, 20% of the randomly selected data is separated to be used for model validation. This data has the same proportion of occupied grids as the whole data set. The regression formula is developed with the remaining 80% of the data.

Subsequently, 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
- $z = b_1 x_1 + b_2 x_2 + ... + b_n x_n + a$
- $x_{1..n}$ = significant physical variables
- $b_{1..n}$ = regression coefficients
- a = constant

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 with actual observations. We created a Relative Operating Characteristic (ROC) curve for presence predictions (Metz, 1978). The curve examines the performance of the model over a range of threshold levels by plotting the proportion of grids correctly predicted to be occupied (sensitivity or *true positive rate*), versus 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, or the *cutoff*, that has the highest true positive rate and lowest false positive rate, and therefore, the best separation of occupied and unoccupied areas. In the following assessment, the habitats with a probability of presence greater than cutoff were classified as suitable.

To validate model strength, we applied the computed formula to the validation data (20%) and compared the number of the fish observations with predictions of suitable habitat. The proportion of correct predictions is recorded as a success rate.

This procedure is repeated 20 times and each time a new randomly selected data set is put aside for validation purposes. After 20 runs, the model generates a list of parameters that were selected in at least two runs and conducts one more run using only these parameters as input attributes. The success rate of this last model is reported together with the average of success rates from previous runs. If these numbers are relatively close and the average is not much higher than the current success rate, the result is considered satisfactory and the model is considered to be final.

To distinguish suitable habitat, we used binary dependent variables indicating presence and absence. In a second model, we focused on high and low abundances. The fish and data were 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 white suckers, more than three fish indicated high abundance. For common shiner, longnose dace, blacknose dace and tessellated darter, more than two individuals were needed. For brook trout the presence of more than one individual indicated high abundance. While we used all the available data for the development of presence models, only data from grids in which fish were caught were used to develop abundance models.

We calculated the probability of presence and of high abundance for every species. The observed presence and abundance at each grid was compared with the probability for the HMU where the grid was located. The suitable habitats with a probability of high abundance greater than the selected cutoff were deemed optimal. Those 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.

Results

Table 2 represents attributes of both models for brook trout that were established from 1,779 grids, including 270 grids where brook trout were captured and 95 grids with high abundance of this species. The presence model indicates an affinity to coarser substrate and a variety of structures; such as Undercut Banks, Boulders and shading. These conditions were found in HMUs with moderate depths and velocities. The abundance model describes deeper habitats, with Undercut Banks.

Table 2: 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 not suitable, suitable and optimal habitats. B represents regression coefficients of the logistic regression model. SE indicates standard error of the coefficients.

Presence		**SE**	**Abundance**		**SE**
calibration success	0.8746		calibration success	0.815	
estimated success	0.8396		estimated success	0.600	
area under roc	0.8514		area under roc	0.840	
Cutoff	0.28		Cutoff	0.32	
Attribute	B		Attribute	B	
Constant	-4.9077	0.672327	Constant	-1.23541	0.350152
Boulders	0.221156	0.109619	Undercut Banks	0.567583	0.224449
Riprap	-0.52298	0.190029	Shallow Margins	-0.5876	0.2074
Overhanging Vegetation	-0.41688	0.133932	Depth 50-75 cm	1.638982	0.998964
Submerged Vegetation	0.536989	0.14189	Depth 75-100 cm	-10.5662	5.309481
Canopy Shading	0.391847	0.118341	Depth 100-125 cm	58.19927	23.71426
Undercut banks	0.282916	0.132164	Velocity 60-75 cm/s	-3.80769	2.474808
Woody debris	-0.3058	0.125831	Velocity 90-105 cm/s	-39.2455	24.25826
Shallow Margins	0.258056	0.106536	Velocity>105	16.1355	8.061479
CASCADE	-1.13234	0.66529	MACROLITHAL	1.764645	0.579664
PLUNGEPOOL	2.342875	0.879931			
POOL	0.957866	0.309516			
RAPIDS	-0.96642	0.46914			
RUFFLE	-0.43917	0.292578			
SIDEARM	0.797638	0.478441			
Depth<25 cm	2.231489	0.541125			
Depth 25-50 cm	1.4334	0.597922			
Depth100-125 cm	2.928222	1.903718			
Velocity<15 cm/s	1.514037	0.456389			
Velocity 15-30 cm/s	1.830022	0.488066			
Velocity 45-60 cm/s	1.419752	0.763817			
Velocity 75-90 cm/s	-4.08722	2.325959			
AKAL	-5.95305	1.82001			
MICROLITHAL	-1.19115	0.485014			
PELAL	-5.63659	2.604593			
PSAMMAL	-3.97862	1.077552			

The following Table 3 presents attributes of the presence model for blacknose dace that was established from 1,580 grids, including 641 grids where this species was captured and 323 grids with a high abundance of this species. The presence model indicates affinity with fast-flowing but shallow HMUs. The trend of the abundance model is Runs and Glides with low and high velocity.

Table 3: Physical attributes correlating with presence and high abundance of blacknose dace. The area under ROC curve is a measure of the discrimination capacity of the model. Selected cut-off indicates the probability separating not suitable, suitable and optimal habitats. B represents regression coefficients of the logistic regression model. SE indicates standard error of the coefficients.

Presence		**SE**	**Abundance**		**SE**
calibration success	0.7385		calibration success	0.739	
estimated success	0.7176		estimated success	0.602	
area under roc	0.7868		area under roc	0.755	
Cutoff	0.52		Cutoff	0.52	
Attribute	B		Attribute	B	
Constant	-1.24447	0.355774	Constant	-0.98154	0.275681
Boulders	0.193665	0.084948	Shallow Margin	-0.41959	0.117845
Submerged Vegetation	-0.81736	0.109911	GLIDE	0.965715	0.231392
Canopy Shading	0.201932	0.096228	RUN	0.55303	0.219465
Shallow Margins	-0.44383	0.084351	Depth<25 cm	1.091835	0.340231
BACKWATER	-1.25794	0.419035	Depth 50-75 cm	-1.86913	0.714027
FASTRUN	0.971053	0.655397	Velocity<15 cm/s	1.667397	0.325075
RUFFLE	0.615082	0.210367	Velocity>105 cm/s	7.003962	4.262344
Depth<25 cm	2.231528	0.360643	PELAL	-4.81179	1.954896
Depth 25-50 cm	0.826431	0.415529	SAPROPEL	12.8045	5.841807
Depth 75-100 cm	-4.13257	1.908641			
Depth 100-125 cm	-5.28267	3.486069			
MEGALITHAL	-2.03757	0.467793			
PHYTAL	-10.6168	3.937434			

In Table 4, attributes of both models for white sucker were established from 1,481 grids, including 241 grids where white suckers were captured and 57 grids with high abundance of this species. The presence model indicates affinity to finer substrate and slower HMUs with moderate depths and open water (no shading). The abundance model describes slower habitats.

Table 4: 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 not suitable, suitable and optimal habitats. B represents regression coefficients of the logistic regression model. SE indicates standard error of the coefficients.

Presence			**Abundance**		**SE**
calibration success	0.8107		calibration success	0.776	
estimated success	0.7538		estimated success	0.702	
area under roc	0.7056		area under roc	0.775	
Cutoff	0.3		Cutoff	0.285	
Attribute	B		Attribute	B	
Constant	-1.20993	0.239712	Constant	-1.55664	0.325328
Submerged Vegetation	-0.21133	0.128	RIFFLE	-1.44878	0.765378
Canopy Shading	-0.31954	0.105	Velocity<15 cm/s	1.198989	0.480445
Shallow Margins	-0.14363	0.096	DETRITUS	-16.2095	11.37966
BACKWATER	1.070461	0.524	MEGALITHAL	-3.27947	2.080651
Depth 25-50 cm	0.685011	0.279			
Velocity 30-45 cm/s	-1.27962	0.430			
Velocity 60-75cm/s	-3.7866	1.024			
MEGALITHAL	-1.0125	0.509			
MICROLITHAL	1.5534	0.361			
PELAL	-3.8743	1.452			
PHYTAL	-11.1237	5.866			
PSAMMAL	0.972	0.437			
SAPROPEL	-4.1275	3.075			

Table 5 represents attributes of both models for longnose dace that were established from 900 grids, including 300 grids where longnose dace were captured and 100 grids with high abundance of this species. The presence model consists of a number of habitat attributes that describe fast flowing shallow HMUs with large gravel and boulders as cover. The abundance model describes Riffle habitats, but with lower velocities, finer substrate and vegetation cover.

Table 5: 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 not suitable, suitable and optimal habitats. B represents regression coefficients of the logistic regression model. SE indicates standard error of the coefficients.

Presence		**SE**	**Abundance**		**SE**
calibration success	0.7385		calibration success	0.739	
estimated success	0.7176		estimated success	0.602	
area under roc	0.7868		area under roc	0.755	
Cutoff	0.42		Cutoff	0.32	
Attribute	B		Attribute	B	
Constant	-0.53315	0.287153	Constant	-1.48292	0.427284
Boulders	0.257365	0.11813	Overhanging Vegetation	0.325044	0.219879
Woody Debris	-0.44096	0.124349	Submerged Vegetation	0.680344	0.36411
Clay	1.034639	0.331324	Clay	1.859982	0.47062
BACKWATER	-1.42394	1.120944	RIFFLE	0.608561	0.288994
RAPIDS	-1.4929	0.381401	Depth 25-50 cm	-1.25792	0.596418
RUN	-1.03868	0.225953	Velocity 15-30 cm/s	1.032572	0.663116
Depth<25 cm	1.310087	0.284931	MICROLITHAL	1.995228	0.798149
Depth 75-100 cm	-6.47203	2.594172			
Velocity<15 cm/s	-1.20313	0.332384			
Velocity 15-30 cm/s	-0.83122	0.388796			
MESOLITHAL	0.854006	0.380272			
PSAMMAL	-1.75763	0.762359			

Table 6 represents attributes of both models for common shiner that were established from 1,014 grids, including 148 grids where common shiner was captured and 71 grids with high abundance of this species. The presence model consists of a high number of habitat attributes that significantly correspond with observed fish. They describe swiftly flowing but shallow HMUs, accompanied by Shallow Margins and Woody Debris. The model also indicates more affinity to coarser substrate and woody deposits. We did not find many common shiners in shallow and slow areas with shading and Undercut Banks. The abundance model describes similar swift habitats, but with boulders as well as finer gravel and sandy substrate.

Table 6: Physical attributes correlating with presence and high abundance of common shiner. The area under ROC curve is a measure of discrimination capacity of the model (0-1). Selected cut-off indicates the probability separating not suitable, suitable, and optimal habitats. B represents regression coefficients of the logistic regression model. SE indicates standard error of the coefficients.

Presence		**SE**	**Abundance**		**SE**
calibration success	0.828		calibration success	0.711	
estimated success	0.8119		estimated success	0.502	
area under roc	0.7436		area under roc	0.776	
Cutoff	0.28		Cutoff	0.450	
Attribute	B		Attribute	B	
Constant	-2.00391	0.477857	Constant	-0.776	0.721
Riprap	0.437485	0.249462	Boulders	0.697	0.307
Canopy Shading	-0.31072	0.168139	Canopy Shading	-0.472	0.333
Undercut Banks	-0.27188	0.147118	Depth<25 cm	-1.189	0.739
Woody Debris	0.334613	0.135984	Velocity 30-45 cm/s	2.022	0.946
Shallow Margins	0.507682	0.134152	Velocity 75-90 cm/s	-3.783	2.541
FASTRUN	1.287691	0.657806	MEGALITHAL	-2.461	1.549
Depth<25 cm	-1.49854	0.351573	MICROLITHAL	1.407	0.891
Velocity<15 cm/s	-1.1891	0.353594	PSAMMAL	5.055	1.858
MEGALITHAL	2.540271	0.843162	XYLAL	-62.477	30.728
MESOLITHAL	1.045589	0.537611			
MICROLITHAL	2.264542	0.544196			
PSAMMAL	-1.36918	0.850235			
XYLAL	20.54929	5.431389			

Table 7 represents attributes of both models for eastern pearlshell that were established from 78 grids collected on the Wekepeke Brook, including 55 grids where mussel were found and captured and 22 grids with high abundance of this species. The presence model consists of a high number of habitat attributes that significantly correspond with observed fish. They describe slow flowing, moderately deep HMUs, accompanied by Undercut Banks. The model also indicates less affinity to organic substrate. The abundance model describes shaded habitats with finer gravel substrate.

Table 7: Physical attributes correlating with presence and high abundance of common shiner. The area under ROC curve is a measure of discrimination capacity of the model (0-1). Selected cut-off indicates the probability separating not suitable, suitable, and optimal habitats. B represents regression coefficients of the logistic regression model. SE indicates standard error of the coefficients.

Presence		**SE**	**Abundance**		**SE**
calibration success	0.833		calibration success	0.727	
estimated success	0.664		estimated success	0.475	
area under roc	0.930		area under roc	0.800	
Cutoff	0.59		Cutoff	0.49	
Attribute	B		Attribute	B	
Constant	2.5078	0.9532	Constant	-2.6719	1.2264
Overhanging Vegetation	-0.6733	0.4655	Overhanging Vegetation	-0.9208	0.6963
Undercut Banks	1.3342	0.7374	Canopy Shading	1.2092	0.5839
Shallow Margins	-2.0523	0.7327	MICROLITHAL	3.0764	1.9688
RAPIDS	-1.8428	1.6032			
RUN	-0.8122	0.8554			
Depth 25-50cm	6.8098	2.5881			
PHYTAL	-2.9586	2.5013			
SAPROPEL	-8.1739	3.6345			

Discussion

The models presented here all have a satisfying capacity to discriminate between occupied and not occupied habitats, which are indicated by high areas under ROC curves. The models also correspond well with empirical expectations. For example, all fluvial specialists show 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. The mussel model corresponds with observations of mussel abundances described in Appendix 3.