R-WD-07-33

Coldwater Fish Assemblage Index of Biotic Integrity for New Hampshire Wadeable Streams

New Hampshire Department of Environmental Services
PO Box 95
29 Hazen Drive
Concord, NH 03302-0095
(603) 271-3503

Thomas S. Burack
Commissioner

Michael J. Walls
Assistant Commissioner

Harry T. Stewart
Director
Water Division

Prepared By:

David E. Neils
Biomonitoring Program Coordinator

May 10, 2007

TABLE OF CONTENTS

1. **INTRODUCTION……………………………………………………** 1

2. **GENERAL PROCESS FOR CWIBI DEVELOPMENT……….** 1

3. **METHODS…………………………………………………………** 2

   - **3.1 Identification of reference sites……………………….** 2
   - **3.2 Definition of a cold water fish community……………** 3
   - **3.3 Biological response indicators (metrics)………………** 4
   - **3.4 CWIBI scoring and threshold identification…………** 4
   - **3.5 Dataset…………………………………………………….** 5

4. **RESULTS…………………………………………………………** 5

   - **4.1 Coldwater fish community definition………………….** 5
   - **4.2 Biological response indicators (metrics)……………….** 8
   - **4.3 Redundancy testing and final metric selection………..** 11
   - **4.4 Metric and CWIBI scoring……………………………..** 15
   - **4.5 IBI threshold determination……………………………** 18
   - **4.6 Validation testing……………………………………….** 18
   - **4.7 IBI score vs. level of human disturbance………………** 22

5. **SUMMARY AND RECOMMENDATIONS……………………...** 23

6. **REFERENCES……………………………………………………** 28

7. **APPENDIX A……………………………………………………..** 29

8. **APPENDIX B……………………………………………………..** 30

9. **APPENDIX C……………………………………………………..** 31

1. INTRODUCTION

The following document describes the development of a coldwater fish assemblage Index of Biotic Integrity (CWIBI) for New Hampshire wadeable streams. The CWIBI is a numeric interpretation of the narrative water quality criteria as stated in New Hampshire Department of Environmental Services Administrative Rules Env – WS1700 covered under the statutory authority given in RSA 485-A:8, VI. Specifically, the narrative standard is detailed in section 1703.19 as:

> Env-Ws 1703.19 Biological and Aquatic Community Integrity.
>
> (a) The surface waters shall support and maintain a balanced, integrated, and adaptive community of organisms having a species composition, diversity, and functional organization comparable to that of similar natural habitats of a region.
>
> (b) Differences from naturally occurring conditions shall be limited to non-detrimental differences in community structure and function.

The product of the CWIBI development process detailed in this document will ultimately be used to assess, in part, the health of applicable aquatic communities. Assessments under this authority will be made for aquatic life use determinations as required for 305(b)/303(d) reporting to the United States Environmental Protection Agency, as well as for other state-level regulatory actions, and general water quality planning activities.

As a two-part narrative criterion, the goal of index development was to first identify the natural structure and function of the fish assemblages residing in the pertinent natural habitats [1703.19(a)] and second determine when a detrimental departure from the natural condition has occurred [1703.19(b)]. The basic approach taken for CWIBI development was the identification of a suitable reference condition and establishment of a natural range of variation within this reference condition (=identification of natural structure and function). Once identified, a reference condition threshold was established below which the biological condition includes detrimental changes in overall aquatic community structure and function (=departure from natural condition). Coldwater fish assemblages not meeting or exceeding the reference condition threshold would be considered to demonstrate significant unnatural community structure and function alterations and consequently not attaining the narrative water quality standard in 1703.19.

2. GENERAL PROCESS FOR CWIBI DEVELOPMENT

The development of a numeric index that interprets the biological condition of coldwater fish assemblages included five basic steps:

1) ***Identification of the reference sites*****:** An *a-priori* process used to select sites with minimal human impacts in order to establish the reference biological condition.

2) ***Definition of coldwater fish assemblage*****:** The determination of non-biological characteristics that describe streams that contain indicator coldwater fish species.

3) ***Identification of biological response indicators (metrics)*****:** The selection of the best ecological measures of community structure and function. Generally known as metric selection.

4) ***Establishment of index scoring criteria and thresholds*****:** A comparison of reference and non-reference biological conditions for the purpose of determining when substantial unnatural impacts to ecological structure and function have occurred.

5) ***Validation of index:*** Testing of metric responses, comparison of reference and non-reference conditions, and testing of proposed threshold with an independent dataset.

The end result of the development process is a numeric index that includes multiple response indicators (i.e. multi-metric) that are considered cumulatively to quantify the biological condition of applicable streams. The index should be sensitive to human disturbance in that it demonstrates declining biological conditions in response to increasing anthropogenic impacts.

3. METHODS

3.1 Identification of Reference Sites

The DES biomonitoring program uses an objective geospatial process in conjunction with basic habitat assessment to identify candidate reference sites. The geospatial process utilizes geographic information systems (GIS) to identify and compile data from a total of 10 variables at the watershed and local (1-mile upstream radius) scale (Appendix A). In addition, the EPA rapid bioassessment protocol habitat assessment form (Barbour et al. 1999) is used as an immediate-scale indicator of the level of human disturbance. Each of these indicators is then combined to produce a final Human Disturbance Gradient (HDG) score [0 = minimum (low disturbance); 33 = maximum (high disturbance)]. The HDG score serves as a relative gross indicator of the amount of human activity within the upstream watershed at any given sampling point. In conjunction with the HDG, best professional judgment is used to characterize the natural state of each sampling point and to make a final determination of whether or not it is reference quality. Sites that were considered to be less than reference quality were further evaluated and placed into intermediate and moderately disturbed categories for IBI development purposes (see 3.5 below). Based on the coarse *a-priori* geospatial process described above, none of the sites included in the dataset were considered to be highly impacted by human disturbance.

3.2 Definition of a Coldwater Fish Assemblage

The DES biomonitoring unit has produced a general fish assemblage type classification system based on multivariate statistical techniques known as nonmetric multidimensional scaling (NMS) in conjunction with cluster analysis (See Appendix B for NMS and cluster analysis results). NMS and cluster analysis was based only on fish species that occurred at less than or equal to five percent of sites and were of native origin to New Hampshire. A total of 13 fish species met these criteria. The outcome of the analysis identified three primary fish assemblage types based on reference site native fish species composition utilizing presence/absence data collected from 1st through 4th order wadeable streams distributed across the state. Presence/absence data from each site was then used to determine the overall frequency of occurrence of individual fish species for each assemblage type. The grouping of sites deemed to be "coldwater" fish assemblages had Eastern brook trout (*Salvelinus fontinalis*) and slimy sculpin (*Cottus cognatus*) as the most frequently occurring species at 100 percent and 67 percent, respectively (Table 1). Coldwater fish assemblages were further defined by low average species richness (2.5 species / site) with 75% of the sites characterized as having four or fewer native fish species.

Table 1. Frequency of occurrence of fish species most commonly encountered at reference sites used to identify 3 preliminary fish assemblage types identified in 1st to 4th order wadeable streams in New Hampshire using NMS and cluster analysis. (N = number of sites).

Common Name	Scientific Name	Preliminary Fish Group (% of sites)		
		Coldwater* (N=43)	***Warm Water**** (N=4)	***Cool Water**** (N=7)
Longnose sucker	*Catostomus catostomus*	5	0	29
Common white sucker	*Catostomus commersoni*	12	75	0
Slimy sculpin	*Cottus cognatus*	67	0	100
Lake chub	*Couesius plumbeus*	5	0	43
Tessellated darter	*Etheostoma olmstedi*	2	0	0
Burbot	*Lota lota*	5	0	43
Common shiner	*Luxilus cornutus*	0	100	14
Northern redbelly dace	*Phoxinus eos*	0	25	14
Blacknose dace	*Rhinichthys atratulus*	23	75	57
Longnose dace	*Rhinichthys cataractae*	16	75	100
Eastern brook trout	*Salvelinus fontinalis*	100	25	0
Creek chub	*Semotilus atromaculatus*	2	0	0
Fallfish	*Semotilus corporalis*	0	25	0

* Preliminary descriptive names for fish assemblage types. Names are based on fish species composition and basic knowledge of the stream environments in which the dominant species was most often found.

Once sites belonging to the coldwater fish assemblage type were assigned, static (non-changing) environmental descriptors were compared among all remaining fish assemblage types. Where possible, environmental variable breakpoints were identified and used, in part, to establish criteria for defining each fish assemblage type.

3.3 Biological Response Indicators (metrics)

Candidate metrics were selected from previously developed CWIBIs (Hughes et al. 2004; Karr 1981; Langdon 2001; Leonard and Orth 1986; Lyons et al. 1996; Mundahl and Simon 1999) and tested for their ability to respond to sites with varying levels of human disturbance. Metrics were computed based on species composition, thermal regime preferences, trophic classes, and tolerance to pollution. Species common names, scientific names, and respective thermal regime preferences, trophic class, and tolerance to pollution for commonly encountered fish species used in CWIBI construction are presented in Table 2.

Table 2. Common and scientific names with respective thermal regime preference, trophic class, and tolerance to pollution for common fish species in New Hampshire coldwater wadeable streams. Sources for table were Halliwell et al. 1999 and Langdon 2001.

Common Name	Scientific Name	Thermal Regime Preference	Trophic Class	Tolerance to Pollution
Brook Trout	*Salvelinus fontinalis*	Coldwater	Top Carnivore	Intolerant
Brown Trout	*Salmo trutta*	Coldwater	Top Carnivore	Intolerant
Burbot	*Lota lota*	Coldwater	Top Carnivore	Moderate
Common Shiner	*Luxilus cornutus*	Eurythermal	Generalist Feeder	Moderate
Creek Chub	*Semotilus atromaculatus*	Eurythermal	Generalist Feeder	Tolerant
Eastern Blacknose Dace	*Rhinichthys atratulus*	Eurythermal	Generalist Feeder	Tolerant
Fallfish	*Semotilus corporalis*	Eurythermal	Generalist Feeder	Moderate
Lake Chub	*Couesius plumbeus*	Coldwater	Generalist Feeder	Moderate
Longnose Dace	*Rhinichthys cataractae*	Eurythermal	Benthic Insectivore	Moderate
Longnose Sucker	*Catostomus catostomus*	Coldwater	Benthic Insectivore	Moderate
Northern Redbelly Dace	*Phoxinus eos*	Coldwater	Generalist Feeder	Moderate
Rainbow Trout	*Oncorhynchus mykiss*	Coldwater	Top Carnivore	Intolerant
Slimy Sculpin	*Cottus cognatus*	Coldwater	Benthic Insectivore	Intolerant
White Sucker	*Catostomus commersoni*	Eurythermal	Generalist Feeder	Tolerant

The distributions of metric values for reference and non-reference (test) sites were determined and used as the basis for metric testing. Reference and test site differences in metric values were compared using the Mann-Whitney *U* test for continuous metrics and contingency tables for discrete metrics. An initial set of metrics was selected and tested for redundancy using the Spearman correlation coefficient. Final CWIBI metric selection attempted to minimize metric redundancy, maximize the selection of metrics with the greatest separation between reference and test sites, and include several different metric types that captured broad structural and functional ecological categories.

3.4 CWIBI Scoring and Threshold Identification

CWIBI scores were the summation of the individual metric scores. Scores for individual metrics were established by reviewing the frequency distribution of reference and test sites. Three scoring categories were established to be consistent with previously developed CWIBIs. Specifically, a decision was made to implement a scoring system similar to that used for the CWIBI developed by the Vermont Department of Environmental Conservation (VT DEC) (Langdon 2001). The CWIBI index developed by DES used three scoring categories (7.5, 4.5, 1.5) with higher scores representing better condition. For each metric, raw values were assigned logical breakpoints to determine scoring categories. A final CWIBI threshold for aquatic life use attainment was based on the 25th percentile score for all reference sites.

3.5 Dataset

The CWIBI development phase included 81 sites where fish data was collected from 1997 through 2005 in 1st through 4th order wadeable streams. The entire dataset was subsequently broken into calibration and validation subsets. The calibration dataset included 27 reference sites, defined as having HDG scores less than or equal to three, and eight 'test' sites (presumed moderately impacted), defined as having HDG scores greater than or equal to eight. The validation dataset included 10 reference sites, five test sites, and 25 intermediate (impact unknown) sites, defined as having HDG scores from four to six. The validation dataset was used to confirm the ability of the CWIBI to discriminate between reference, intermediate, and test sites, and if the frequency of threshold attainment differed between these three site types.

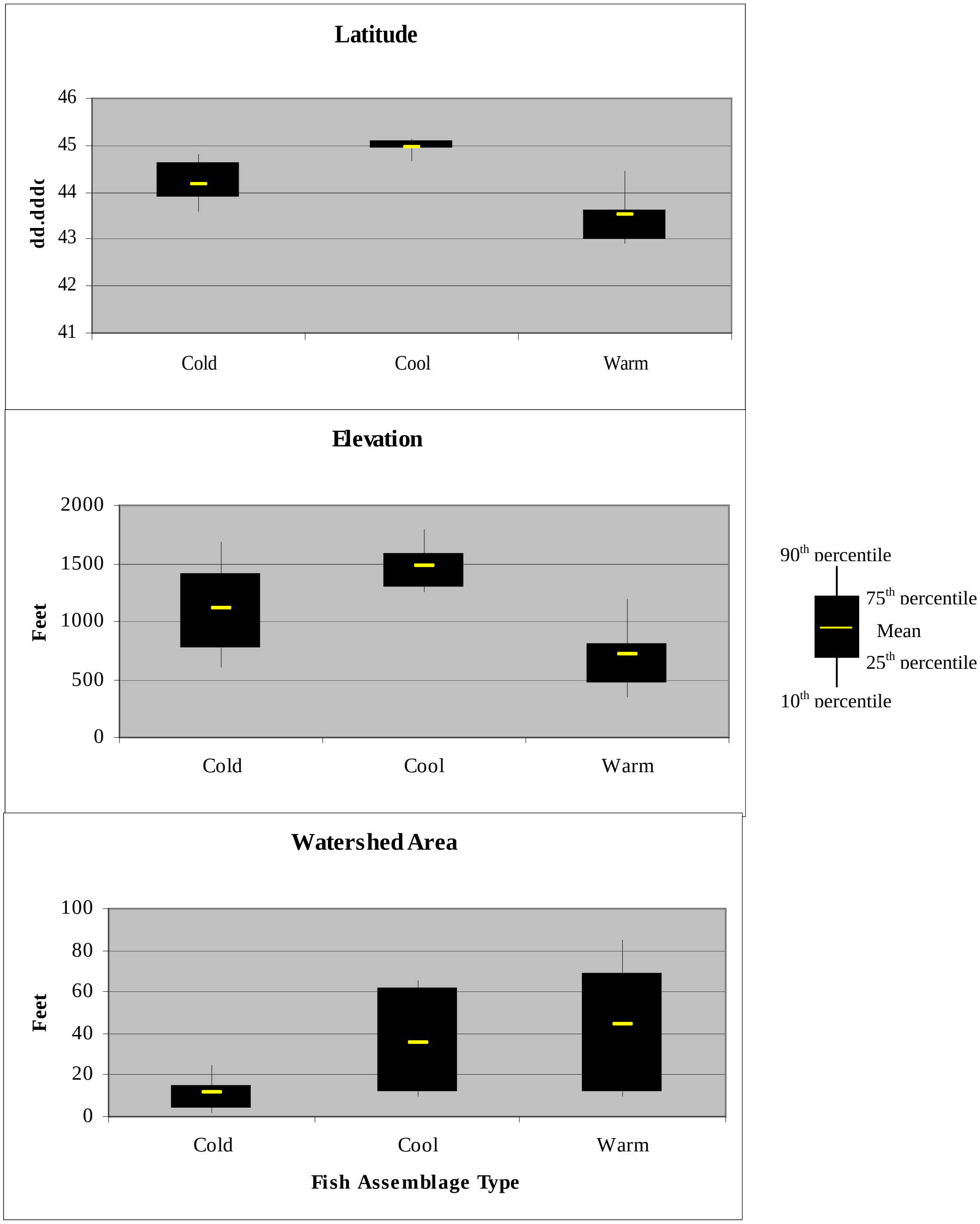
Fish data included the number of individuals of each species captured at all sites. A single pass (150-meter) backpack electrofishing effort was employed to capture fish. Stunned fish were captured, identified, enumerated, and returned to the stream. Actual shocking time was recorded to produce catch-per-unit-effort (CPUE) estimates. For inclusion into index computation a minimum of two individuals was required and all Atlantic salmon were excluded from the dataset since they subsist through stocking efforts alone (Langdon 2001). Little information was available or documented prior to 2002 regarding the origination (stocked or natural reproduction) of other native and non-native salmonids. As a result no distinction of their origin was attempted and all salmonids, exclusive of Atlantic salmon, were considered in metric computation and scoring.

4. RESULTS

4.1 Coldwater Fish Assemblage Definition

A total of nine environmental variables were compared among three primary fish assemblage types. Of the nine variables, latitude, elevation, and watershed size were selected as important *a-priori* variables in defining coldwater fish assemblages and proved to be important in distinguishing between fish assemblage types (Figure 1). In addition, the selected environmental variables are static (not subject to human impacts or natural changes) and are relatively easy to obtain. In general, coldwater fish assemblages occurred most frequently north of 43.75 degrees latitude, in excess of 775 feet in elevation, and in streams with watersheds smaller than 15 square miles (minimum watershed area = 1 square mile). The selection of environmental variable breakpoints was based on logical separations between fish assemblage types. Primary river basin was added as a fourth criterion to take into account small, lower elevation coldwater streams in the Connecticut River basin known to have naturally occurring coldwater fish assemblages.

Figure 1. Box and whisker plots of important environmental variables at reference sites for three preliminary fish assemblage types identified in 1st to 4th order wadeable streams in New Hampshire.



When combined the final *a priori* criteria for defining coldwater fish communities based on environmental variables were:

1) All sites north of 43.75 degrees latitude and a watershed area less than 15 square miles (minimum 1 square mile).

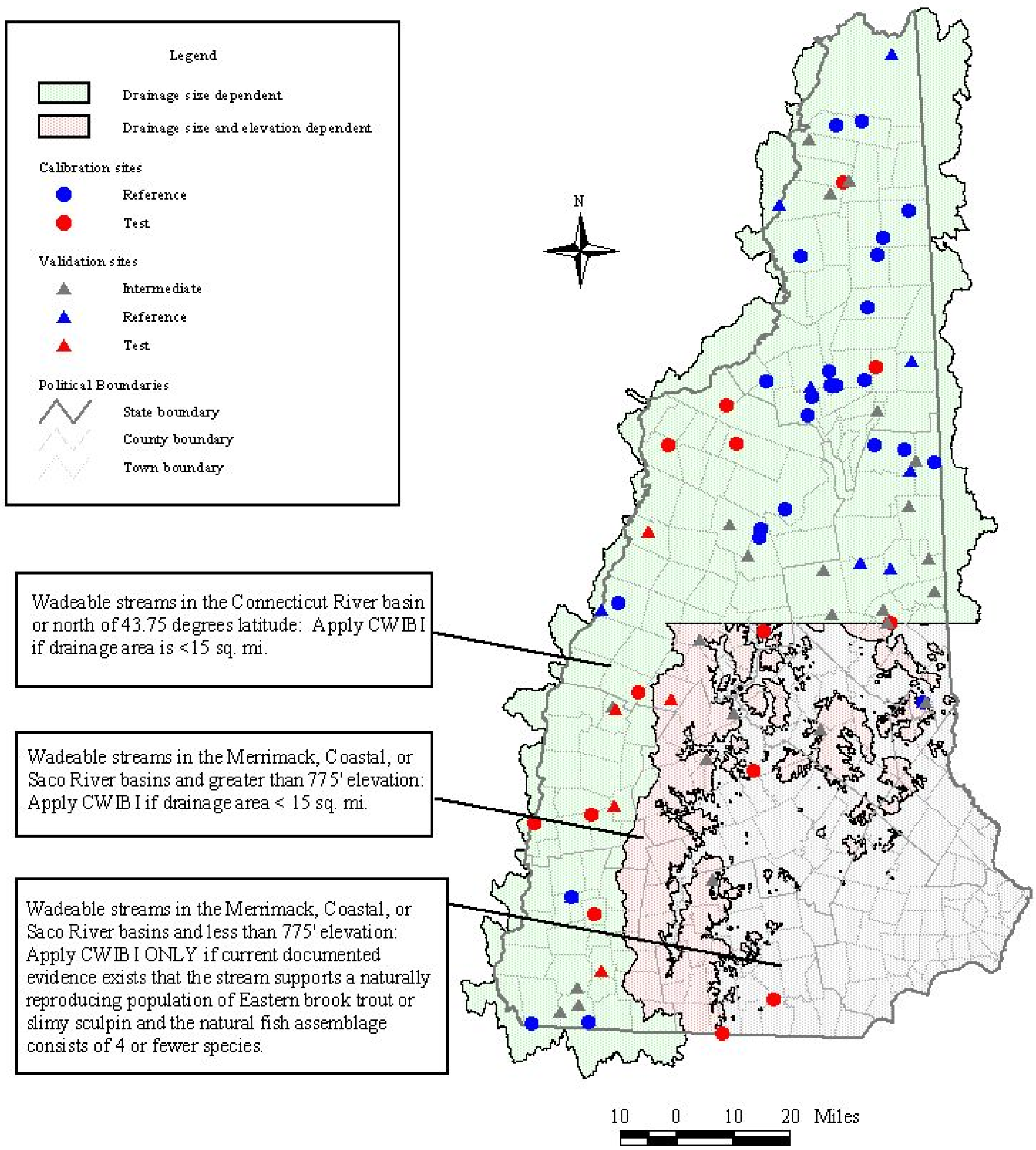
2) All sites in the Connecticut River basin south of 43.75 degrees latitude and a watershed area less than 15 square miles (minimum 1 square mile).

3) All sites in the Merrimack, Saco, or Piscataqua Basins south of 43.75 degrees latitude, greater than 775 feet in elevation, and a watershed area less than 15 square miles (minimum 1 square mile).

4) Other sites not meeting the criteria if documented naturally reproducing populations of Eastern brook trout or slimy sculpin exist and the expected natural species richness is between two and four.

The resultant geographic stratification of applicable CWIBI areas and associated sites included in the development process based on these criteria are illustrated in map 1.

Map 1. Geographic distribution of coldwater fish assemblage areas and the location of sampling sites included in the CWIBI development process.



4.2 Biological Response Indicators (metrics)

A total of 15 candidate metrics were selected for testing between reference and test sites from the calibration dataset (Table 3). A candidate metric's ability to discriminate between reference and test sites was based on seven measures of success (Table 4).

Table 3. List of candidate metrics included in the CWIBI development process for 1st – 4th order wadeable streams in New Hampshire. Trophic status and tolerance categories for individual fish species as defined in Table 2.

Metric Name	Metric Description	Expected Response*
per_CW	% of individuals as coldwater specialists relative to total number individuals in sample	+
EBT_CPUE	Catch per unit effort of Eastern brook trout. Number of individuals caught per minute of shocking time	+
per_BI	% of individuals as benthic insectivores relative to total number of individuals in sample	+
per_GF	% of individuals as generalist feeders relative to total number of individuals in sample	-
per_OI	% of individuals as obligate insectivores relative to total number of individuals in sample	+
per_TC	% of individuals as top carnivores relative to total number of individuals in sample	+
num_intol_sp	number of intolerant species present in sample	+
EBT_age_class	0 = eastern brook trout absent; 1= no young-of-year (YOY); 2= YOY only; 3= YOY and adults	+
per_I	% of intolerant individuals relative to total number of individuals in sample	+
per_M	% of moderately tolerant individuals relative to total number of individuals in sample	-
per_T	% of tolerant individuals relative to total number of individuals in sample	-
num_tol_sp	number of tolerant species present in sample	-
Total_CPUE	total catch per unit effort. Number of individuals caught per minute of shocking time	+
per_ebt	% of individuals as Eastern brook trout relative to total number of individuals in sample	+
per_ss	% of individuals as slimy sculpin relative to total number of individuals in sample	+

* (+) metric increases with increasing biological condition; (-) metric decreases with increasing biological condition

Primary among these measures was if the metric produced the expected response between the reference and test sites and if the response was statistically significant. Additional measures of success for individual metrics were based on the magnitude of separation or overlap between selected percentiles from the population reference and test sites. Raw values for means, medians, and selected percentiles of reference and test site populations are included in Appendix C.

Correct expected metric responses were observed for 13 out of 15 metrics (Table 5). Only the percent of obligate insectivorous individuals and total catch-per-unit-effort (CPUE) metrics displayed responses counter to expectations and were excluded from further consideration. Ten of 12 continuous metrics were significantly different between reference and test sites (Mann-Whitney U; $p<0.05$; Table 5). One of three discrete metrics were significantly different between reference and test sites (chi-square; $p<0.05$; Table 5).

Table 4. Metric response measures and associate descriptions employed to estimate the level of separation between coldwater fish assemblage reference and test sites for the calibration dataset.

Metric Response Measure	Description
Correct response	A first cut for continued metric inclusion. For (+) metrics the mean for reference sites must be higher than test sites. For (-) metrics the mean for reference site must be lower than test sites.
outer_percentile_overlap	A measure of amount of box plot outer quartile overlap. Negative indicates overlapping present. Positive indicates no overlap. Magnitude of overlap expressed as difference. The 25th and 75th percentiles of reference and test sites, respectively, are used for (+) metrics. The 75th and 25th percentiles of reference and test sites, respectively, are used for (-) metrics.
medain_true_outer_percentile_overlap:	A measure of the amount of overlap between reference site median and test site 75th percentile for (+) metrics and test site median and reference site 75th percentile for (-) metrics. Negative indicates overlapping present. Positive indicates no overlap. Magnitude of overlap expressed as difference.
mean_true_outer_percentile_overlap	Same as above except uses respective means rather than medians.
mann-whitney_test_result	Mann-Whitney U test performed to determine if difference present between reference and test sites. Applied only to continuous metrics.
Chi-square_test_result	Chi-square test performed to determine if difference was present in distribution of reference and test sites among categories. Applied only to discrete variables.
Discrimination efficiency	The percentage of test sites outside the 25th [(+) metrics] or 75th [(-) metrics] percentile of reference sites (Stribling et. al 2000).

Additional measures of metric success using selected percentiles were evaluated by comparing the number and magnitude of positive (indicating separation) or negative (indicating overlap) values obtained by computing differences between individual metrics. Of the remaining 10 correct response metrics that showed a significant difference between reference and test sites, seven (per_CW, per_GF, per_TC, per_I, per_T, per_EBT, Num_tol_sp) had varying levels of separation at all three levels of metric comparison (Outer percentile, median/outer percentile, mean/outer percentile; Table 4). The remaining three metrics (per_M, per_ss, EBT_CPUE) had a moderate amount of overlap between the outer percentiles. A summary of final metric selection and rationale for inclusion or exclusion is provided in Table 6.

Following the initial metric response screening, the selection of metrics chosen for redundancy testing was based on summing and ranking the performance measures of all metrics (Table 5). Of the seven remaining candidate metrics, each performed equally well with six out of six correct performance measures. The discrimination efficiency, defined as the percentage of test sites below the 25th percentile of reference sites for positive metrics or above the 75th percentile of reference sites for negative metrics, was considered as a final measure in selecting metrics for redundancy testing. Discrimination efficiencies were in excess of 75 percent for all remaining candidate metrics.

For the Eastern brook trout age class metric, a subjective decision was made to override objective measures of success and further consider it for inclusion into the final IBI. While undetectable through statistical comparisons, this metric is considered to be important for the construction of a robust IBI that includes a measure of reproductive success for an important indicator species (Halliwell et al. 1999). In addition, the dataset used only included reliable data on the age class structure of brook trout populations for four of nine years. Therefore, the apparent similarity between reference and test sites (i.e. lack of significant chi-square test) is partially attributed to incomplete data.

Table 5. Results of metric response measures for comparisons between reference and test coldwater fish assemblage sites from the calibration dataset. Positive (+; blue background) results indicate reference and test site separation; Negative (-; orange background) results indicate reference and test sites overlap.

Metric	Metric Type	Correct response (mean)	Correct response (median)	Outer percentile overlap	Median / outer percentile overlap	Mean / outer percentile overlap	Mann-Whitney *U* test*	chi-square test*	Total # correct	% correct	Discrimination efficiency (%)
CONTINUOUS METRICS											
per_CW	+	Y	Y	17	53.41	33.49	1	------	6	100	87.5
per_BI	+	Y	Y	-22.37	1.02	0.38	0	------	4	66.7	50
per_GF	-	Y	Y	3.36	6.58	15.84	1	------	6	100	87.5
per_OI	+	N	N	-62.31	-62.31	-46.12	1	------	0	0	12.5
per_TC	+	Y	Y	1.01	35.56	33.74	1	------	6	100	75
per_I	+	Y	Y	8.01	53.41	30.86	1	------	6	100	75
per_M	-	Y	Y	-2.7	11.39	10.62	1	------	5	83.3	62.5
per_T	-	Y	Y	29.23	47.14	42.45	1	------	6	100	87.5
per_ebt	+	Y	Y	2.82	38.97	32.72	1	------	6	100	75
per_ss	+	Y	Y	-0.34	13.82	17.13	1	------	5	83.3	75
EBT_CPUE	+	Y	Y	-0.16	0.78	0.97	1	------	5	83.3	62.5
Total_CPUE	+	N	N	-6.53	-5.17	-4.53	0	------	0	0	0
DISCRETE METRICS											
Num_intol_sp	+	Y	Y	0.75	0.75	0.53	------	0	5	83.3	75
EBT_age_class	+	Y	Y	0	0	-0.48	------	0	2	33.3	62.5
Num_tol_sp	-	Y	Y	0.75	1	1	------	1	6	100	75

* Indicates significance; 1 = $p < 0.05$, 0 = $p > 0.05$.

4.3 Redundancy testing and final metric selection

In an attempt to avoid redundant metrics, Spearman's correlation coefficients were generated for all of the selected metric combinations. A cut point of $r < 0.80$ was utilized as a tentative criterion for determining the lack of strong autocorrelation between metrics. Eleven out of 28 possible total metric combinations had correlation coefficients in excess of 0.80 (Table 7). The percentage of generalist feeders and Eastern brook trout age class metrics were the only metrics that met the tentative correlation criterion with all other metrics. The percentage of intolerant individuals and number of tolerant species

Table 6. Candidate CWIBI metric disposition and rationale for inclusion or exclusion for further testing.

Metric	Include / Exclude	Reason for Inclusion / Exclusion
	CONTINUOUS METRICS	
per_CW	Include	6 out of 6 correct responses; excellent separation between test vs. reference sites; 87.5% discrimination efficiency; included in VT DEC coldwater IBI.
per_BI	Exclude	4 out of 6 correct responses; minimal separation between test vs. reference sites [ns Mann-Whitney test, ($p>0.05$)]; lower discrimination efficiency (50%) when compared to all candidate metrics.
per_GF	Include	6 out of 6 correct responses, good separation between test vs. reference sites, minimal variation in per_GF individuals among reference sites; 87.5% discrimination efficiency; included in VT DEC coldwater IBI; not correlated with candidate metrics.
per_OI	Exclude	Incorrect mean and median response.
per_TC	Include	6 out of 6 correct responses; good separation between test vs. reference sites; 75% discrimination efficiency; included in VT DEC coldwater IBI.
per_I	Exclude	6 out of 6 correct responses; excellent separation between test vs. reference sites; 75% discrimination efficiency; included in VT coldwater IBI; metric is represented in NH CWIBI through per_CW, per_TC, and per_ebt metrics.
per_M	Exclude	5 out of 6 correct responses; minimal separation between test sites vs. reference sites, 62.5% discrimination efficiency; representative of fish species that frequently occur in reference and test sites at densities that may be unrelated to habitat and/or water quality.
per_T	Exclude	6 out of 6 correct responses; excellent separation between test vs. reference sites; 87.5% discrimination efficiency; opted to included num_tol_sp_metric in IBI instead in order to have richness metric with significant response between test vs. reference sites.
per_ebt	Include	6 out of 6 correct responses; good separation between test vs. reference sites; 75% discrimination efficiency; identified as 1 of 2 statewide indicator species of coldwater fish communities.
per_ss	Exclude	5 out of 6 correct responses; good separation between test vs. reference sites; 75% discrimination efficiency; excluded because several reference sites were naturally devoid of this species and would have resulted in an unnecessarily low metric score.
EBT_CPUE	Exclude	5 out of 6 correct metric responses; minimal separation between test vs. reference sites; 62.5% discrimination efficiency; CPUE estimates unreliable as estimate of density.
Total_CPUE	Exclude	Incorrect mean and median response.
	DISCRETE METRICS	
Num_intol_sp	Exclude	ns chi-square test ($p>0.05$) between test vs. reference sites.
EBT_age_class	Include	Response testing showed little difference between metric and test sites, included in VT DEC coldwater IBI, subjective decision made to include metric because provides measure of reproductive success for important coldwater species.
Num_tol_sp	Include	6 out of 6 correct responses; excellent separation between test vs. reference sites; 75% discrimination efficiency; represents only richness metric that displayed strong response between test vs. reference sites [sig. chi-square test ($p<0.05$)].

metrics were the most redundant metrics, each with five out of seven possible metric combinations having correlation coefficients in excess 0.80. While the tentative criterion

Table 7. Spearman's correlation coefficients, level of significance, and number of calibration sites (N) for CWIBI candidate metrics.

		per_CW_ metric	per_GF	per_TC	per_I	per_T	per_ebt	num_tol_sp	EBT_age_ class
per_CW_ metric	Correlation Coefficient	1.000	-.622(**)	.829(**)	.950(**)	-.930(**)	.808(**)	-.912(**)	.201
	Sig. (2-tailed)	.	.000	.000	.000	.000	.000	.000	.246
	N	35	35	35	35	35	35	35	35
per_GF	Correlation Coefficient		1.000	-.525(**)	-.628(**)	.534(**)	-.586(**)	.689(**)	-.318
	Sig. (2-tailed)		.	.001	.000	.001	.000	.000	.063
	N		35	35	35	35	35	35	35
per_TC	Correlation Coefficient			1.000	.864(**)	-.759(**)	.961(**)	-.777(**)	.199
	Sig. (2-tailed)			.	.000	.000	.000	.000	.253
	N			35	35	35	35	35	35
per_I	Correlation Coefficient				1.000	-.900(**)	.845(**)	-.882(**)	.228
	Sig. (2-tailed)				.	.000	.000	.000	.187
	N				35	35	35	35	35
per_T	Correlation Coefficient					1.000	-.745(**)	.942(**)	-.098
	Sig. (2-tailed)					.	.000	.000	.577
	N					35	35	35	35
per_ebt	Correlation Coefficient						1.000	-.776(**)	.282
	Sig. (2-tailed)						.	.000	.101
	N						35	35	35
num_tol_s p	Correlation Coefficient							1.000	-.206
	Sig. (2-tailed)							.	.235
	N							35	35
EBT_age_ class	Correlation Coefficient								1.000
	Sig. (2-tailed)								.
	N								35

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

proved useful for screening purposes, final metric selection required the acceptance of some metric redundancy in order to construct an IBI that had a sufficient number of metrics and different metric types. Of the metrics selected for inclusion (Table 6), only four out of 15 combinations had correlation coefficients in excess of 0.80, with three of the four coming from the coldwater specialist metric. The strong correlations associated with this metric were recognized, but were nonetheless accepted to include a metric that accounted for the presence of the only non-carnivorous coldwater specialist species (i.e slimy sculpin). A target of six total metrics was decided upon to be consistent with previously constructed CWIBIs (Langdon 2001). There was also a desire to have both positive and negative response metrics, as well as a species richness metric, and one or more metrics related to Eastern brook trout and their reproductive success. The final suite of six metrics best fitting the above requirements while concurrently minimizing the level of inter-metric correlation were as follows:

1) **Percentage of coldwater specialist individuals** (+) **–** expected to be highest in minimally impacted streams (reference quality) (greater than 70%) and decrease with increasing anthropogenic activity through the removal of riparian buffers or related activities. The ultimate impact to the fish assemblage is realized through an unnatural increase in a stream's thermal regime. Native coldwater specialist species in New Hampshire wadeable streams are represented by Eastern brook trout, slimy sculpin, longnose sucker, burbot, and lake chub.

2) **Percentage of generalist feeders** (-) **–** expected to comprise a minimal (<10%) proportion of individuals in minimally impacted streams and increase with increasing anthropogenic activity, especially through instream physical habitat modification. Generalist feeders are able to capitalize on a variety of different food sources and often increase dramatically with habitat degradation. For the calibration dataset, this metric showed the most dramatic difference between reference and test sites.

3) **Percentage of top carnivores (+) –** expected to be highest in minimally impacted streams (>30%) and decrease with instream impacts. In New Hampshire coldwater wadeable streams Eastern brook trout are the dominant top carnivore, but naturally reproducing populations of non-native brown and rainbow trout are also occasionally found, especially in the Androscoggin and upper Connecticut River drainages.

4) **Percentage of Eastern brook trout** (+) **–** expected to be highest in minimally impacted streams (~50%) and decrease with increasing anthropogenic activities. Selected as an important indicator species for coldwater streams. A reduced percentage of Eastern brook trout can reflect unnatural changes in physical habitat quantity and quality, potential effects of acid deposition, thermal regime modification, or other unknown human related perturbances.

5) **Eastern brook trout age class** (+) **–** The presence of adult and young-of-the-year (YOY) are expected to occur at minimally impacted sites and decrease in frequency as anthropogenic impacts increase. The inclusion of this categorical metric [(1) YOY absent; (2) YOY present; (3) Adults and YOY present] provides a measure of reproductive success, an important attribute for consideration when assessing the overall ecological integrity of biological

communities. The lack of YOY and / or adults might result from several types of human impacts including physical habitat degradation or thermal regime alteration.

6) **Number of tolerant species** (-) **–** Expected to be absent or consist of a single species comprising a low proportion of the total individuals captured in minimally impacted streams and increase with increasing anthropogenic activities. Selected as the only taxa richness metric that was significantly different between reference and test sites. The presence of tolerant species may reflect several types of human impacts. The Eastern blacknose dace is the most commonly occurring tolerant species associated with coldwater fish assemblages and not necessarily indicative of significant negative impacts by itself when comprising a low percentage of the entire fish assemblage. However, the addition of other tolerant species is not common for high quality coldwater fish assemblages.

As a final check on the degree of separation between reference and test sites for the calibration dataset, box and whisker plots were constructed for the six metrics selected for inclusion into the CWIBI (Figure 2). Five of six metrics box plots showed a moderate to high level of separation between reference and test sites confirming initial metric testing results. For the Eastern brook trout age class metric, the level of separation was less clear; however, adult and YOY were present at 78 percent of reference sites (21 out of 27) and only 38 percent of test sites (three out of eight) justifying its inclusion into the CWIBI. The generalist feeder metric showed a low level of variability among reference sites (25 of 27 sites less than 5 percent) and was consistently lower than test sites.

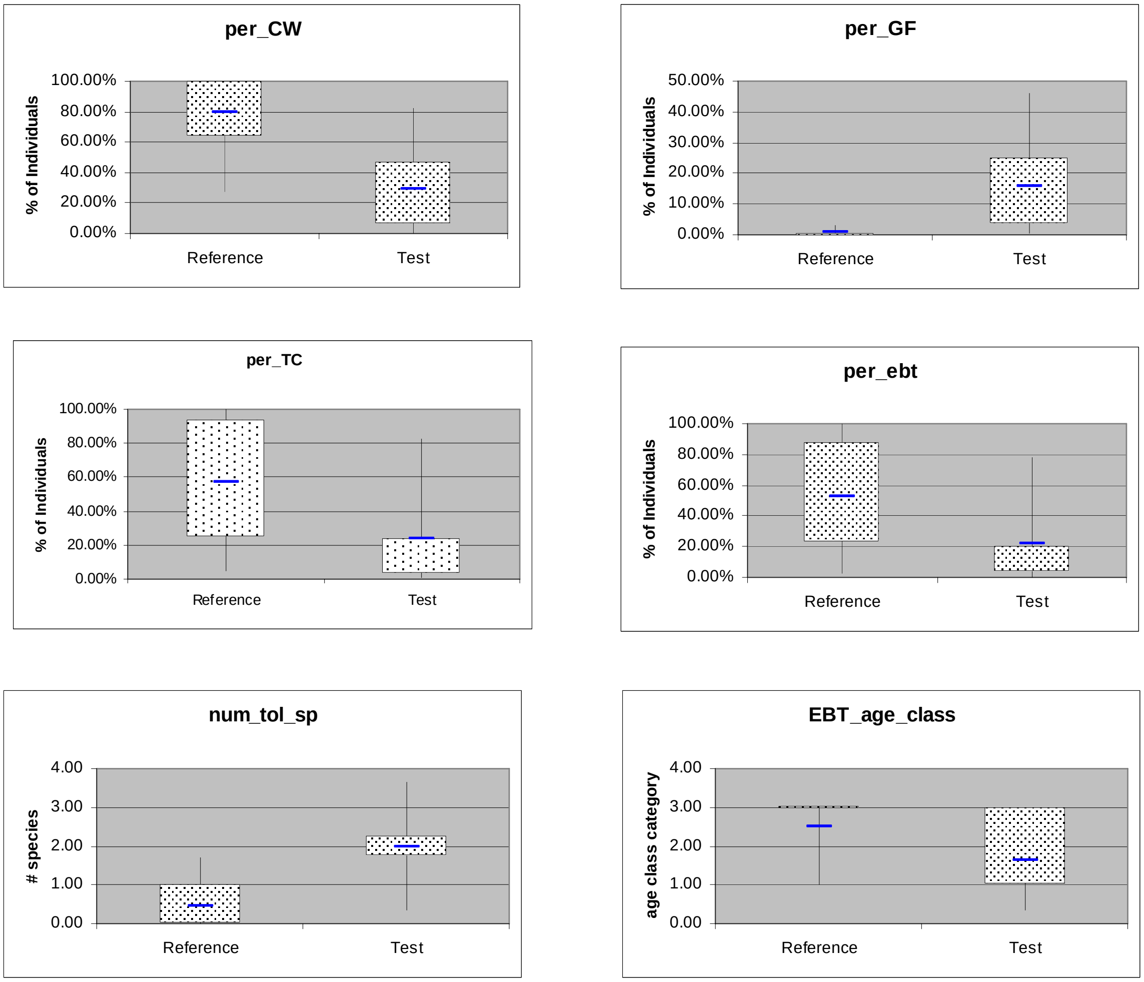
4.4 Metric and CWIBI Scoring

Raw metric values were converted to a numeric score based on the IBI scoring categories as established by the VT DEC (Langdon 2001). Each metric from an individual site was eligible for three scoring categories (1.5, 4.5, 7.5) depending on the raw metric result. Scoring categories were selected to be consistent with the VT DEC CWIBI. Low metric scores indicate poorer condition than high metric scores.

Metric score categories and corresponding raw metric thresholds were established by examining the frequency distributions of reference sites and test sites. An attempt was made to have 20 reference sites within the highest score category (7.5) with logical raw metric break points for the 4.5 and 1.5 scoring categories (Table 8).

For the Eastern brook trout age class metric, scoring categories mimicked those utilized in the VT DEC CWIBI, granting the highest score (7.5) for streams where both adult and young-of-year individuals are captured, an intermediate score (4.5) where only young-of-year are captured, and the lowest score (1.5) where young-of-year are absent. The

Figure 2. Box and whisker plots (upper whisker = 95th percentile; lower whisker = 5th percentile; upper boundary of box = 75th percentile; lower boundary of box = 25th percentile; blue line = mean) of selected metrics for coldwater fish assemblage reference and test sites from the calibration dataset. NOTE: 5th, 25th, and 75th percentiles for per_GF metric = 0%.



absence of young-of-year Eastern brook trout in a given stream is considered to be indicative of streams where natural reproduction is lacking or severely hampered due to anthropogenic causes. However, reproductive failure can be a result of natural phenomena (i.e. flooding, drought, severe winter, etc.) leading to an incorrect assumption of human-based impacts. Natural event impacts, in all likelihood, would be temporally limited [i.e.occurring the particular year(s) of the event(s)] and widespread across all streams regardless of the level of human impact. Our data indicate otherwise, with the presence of young-of-year Eastern brook trout unrelated to a specific year(s) of sampling and aligned closely with minimally impacted sites (reference sites).

Table 8. CWIBI scoring categories, proposed criteria for raw metric values, and frequency distribution of reference and test sites.

	%GF		
Score=>	7.5	4.5	1.5
Raw Metric Threshold =>	<3%	3-6%	>6%
# Reference Sites	24	2	1
# Test Sites	2	2	4

	%TC		
Score=>	7.5	4.5	1.5
Raw Metric Threshold =>	>35%	20-35%	<20%
# Reference Sites	19	3	5
# Test Sites	2	0	6

	%CW		
Score=>	7.5	4.5	1.5
Raw Metric Threshold =>	>70%	50-70%	<50%
# Reference Sites	20	1	6
# Test Sites	1	1	6

	%EBT		
Score=>	7.5	4.5	1.5
Raw Metric Threshold =>	>25%	10-25%	<10%
# Reference Sites	19	4	4
# Test Sites	2	1	5

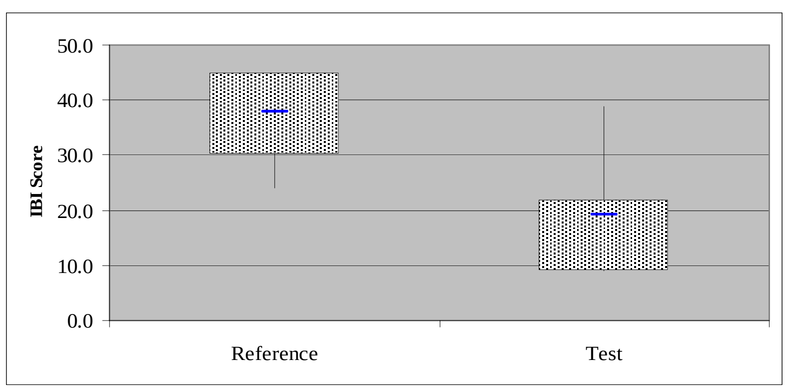
	EBT_Age_class		
Score=>	7.5	4.5	1.5
Raw Metric Threshold =>	YOY & adults	YOY only	YOY absent
# Reference Sites	21	0	6
# Test Sites	3	0	5

	# Tolerant species		
Score=>	7.5	4.5	1.5
Raw Metric Threshold =>	0	1	>=2
# Reference Sites	19	6	2
# Test Sites	1	1	6

Final CWIBI scores were computed by summing individual metric scores. The minimum score was nine and the maximum score was 45. The mean CWIBI score for reference sites was significantly higher than test sites (Mann-Whitney *U* test, $p<0.001$). The 25th percentile of reference sites was clearly separated from the 75th percentile of test sites (Figure 3). Mean reference and test site scores were separated by 18 points. Six of eight test sites scored below the 25th percentile of reference sites (75 percent).

Figure 3. CWIBI score summary (upper whisker = 95th percentile; lower whisker = 5th percentile; upper boundary of box = 75th percentile; lower boundary of box = 25th percentile; blue line = mean) for reference and test sites from the calibration dataset. Note: 95th and 75th pecentiles are the same for references sites; 5th and 25th percentiles are the same for test sites.

Site Type	Statistic	Result
Reference	95th	45.0
	75th	45.0
	mean	37.9
	25th	30.0
	5th	24.0
Test	95th	38.9
	75th	21.8
	mean	19.1
	25th	9.0
	5th	9.0



4.5 IBI threshold determination

The threshold for binomial Aquatic Life Use attainment status (full support, non-support) was established by reviewing the distribution of CWIBI scores for calibration reference and test sites. The existing DES benthic IBI threshold was set at the 25th percentile of the reference site distribution. To be consistent the same logic was applied to the CWIBI. The 25th percentile CWIBI score of the calibration reference sites was 30.0 (Figure 3). A final CWIBI threshold of 30 is recommended for ease of communication. This compares favorably to the VT DEC coldwater IBI threshold for sites considered in “fair” condition (IBI score = 27). When a CWIBI score threshold of 30 was applied, 22 of 27 (82 percent) reference sites and 2 of 8 (25 percent) test sites for the calibration dataset exceeded the criteria (Table 9). Contingency tables indicated that the distribution of reference and test site exceeding and failing to meet the recommended criteria were significantly different (chi-square; $p<0.05$).

Table 9. Observed and expected frequency of IBI attainment (above threshold) and failure (below threshold) for reference and test sites from the calibration dataset. Expected values were computed as [Total # sites per Site Type / (Total # Reference + Test sites)] * (# observed Reference sites + # observed Test sites Above or Below threshold). Example: # expected Reference sites Above threshold = [27/(27+8)]*(22+2)=18.5.

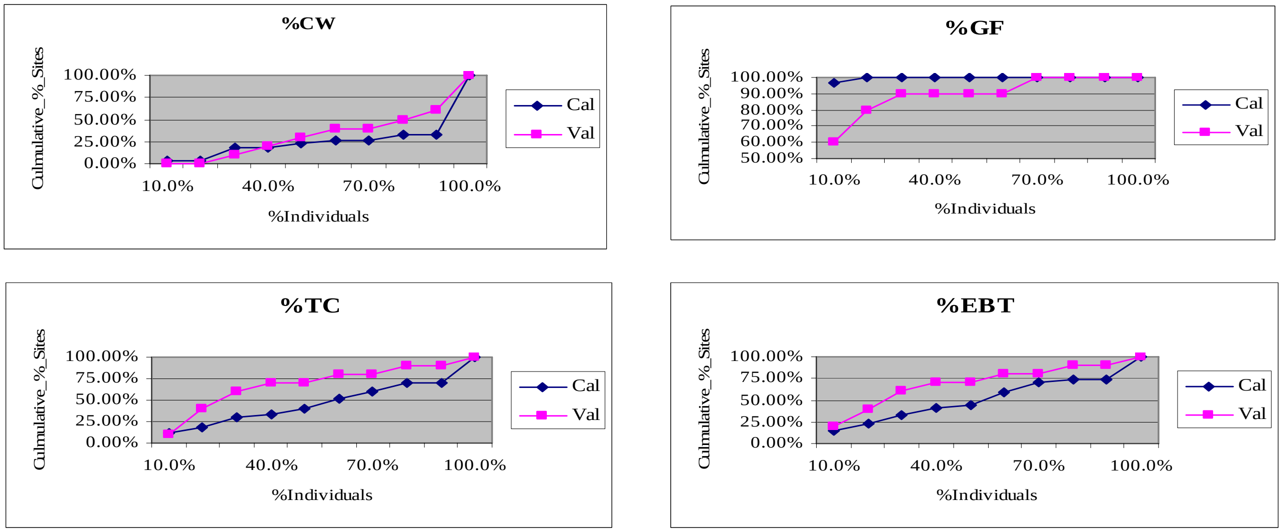
Site Type		**# Above threshold**	**# Below threshold**	**Total**	**Chi-square**	**Critical value** (1df, 0.05)
Reference	# observed	22	5	27	6.40	3.841
	# expected	19	8			
Test	# observed	2	6	8		
	# expected	5	3			
	total	24	11	35		

4.6 Validation testing

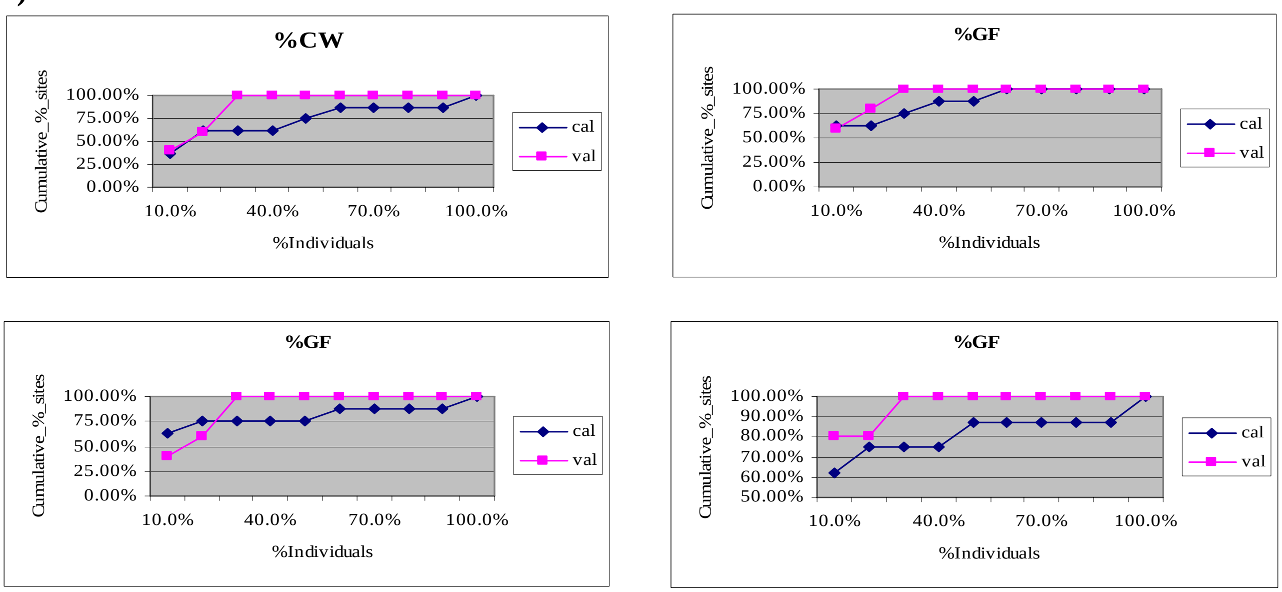
A total of 40 sites were excluded from the CWIBI development process for the purposes of validating the performance of the CWIBI. As an initial check on the similarity of raw values for continuous metrics between reference and test sites for calibration and validation datasets, cumulative distribution plots were constructed (Figure 4). For all metrics, only slight differences

Figure 4. Cumulative percentage of calibration (cal) and validation (val) datasets for reference (A) and test (B) sites for individual metrics.

A) Reference



B) Test



were observed in the distribution of sites between calibration and validation datasets for reference and test sites, indicating they likely came from the same distribution. Contingency testing confirmed these observations with non-significant differences (chi-square; $p>0.05$) between frequency distributions of reference and test sites for all metrics in the calibration and validation datasets (Table 10). Once it was confirmed that raw metric values were similar for the validation and calibration datasets, CWIBI scores were computed for the reference and test sites from the validation dataset.

Table 10. Chi-square contingency table testing results between calibration and validation datasets for reference (A) and test (B) sites for individual metrics.

A) Reference

Metric	Variable Type	Chi-square	df	0.05 critical value	Sig. (Y/N)
%CW	Continuous	1.948	4	9.488	N
%GF	Continuous	5.709	4	9.488	N
%TC	Continuous	4.083	4	9.488	N
%EBT	Continuous	2.606	4	9.488	N
EBT_age_class	Discrete	5.580	2	5.991	N
num_tol_sp	Discrete	5.880	2	5.991	N

B) Test

Metric	Variable Type	Chi-square	df	0.05 critical value	Sig. (Y/N)
%CW	Continuous	5.078	4	9.488	N
%GF	Continuous	0.794	4	9.488	N
%TC	Continuous	4.550	4	9.488	N
%EBT	Continuous	2.860	4	9.488	N
EBT_age_class	Discrete	0.627	2	5.991	N
num_tol_sp	Discrete	1.477	2	5.991	N

Reference and test site CWIBI scores showed clear separation for the validation dataset (Figure 5). Reference and intermediate (disturbance) sites had similar CWIBI scores with little separation. Reference site and intermediate sites had similar CWIBI score means (32-reference; 31.1-intermediate). There was minimal overlap between the 25th percentile (18.0) of reference site CWIBI scores and the 75th percentile (21.0) of test site CWIBI scores. Three of five test sites scored below the 25th percentile of reference site CWIBI scores (60 percent discrimination efficiency).

For the validation dataset, 5 of 10 (50 percent) reference sites and none of the test sites exceeded the recommended criteria (Table 11). For intermediate sites, 17 of 25 (68 percent) exceeded the recommended criteria. Contingency table testing indicated that the distribution of reference and test sites exceeding and failing to meet the recommended criteria were significantly different for the validation dataset (chi-square; $p<0.05$). A final comparison between calibration and validation sites using box and whisker plots indicated that reference site CWIBI score means both exceeded the recommended criteria of 30 (Figure 6). Mean test site CWIBI scores from the validation dataset were similar to the calibration dataset and substantially lower than the recommended criteria for validation than the calibration sites. Sites with an intermediate level of disturbance from the calibration and validation dataset were nearly evenly distributed above and below the recommended criteria and with means more comparable to reference than test sites.

Figure 5. CWIBI score summary (upper whisker = 95th percentile; lower whisker = 5th percentile; upper boundary of box = 75th percentile; 25th percentile = low boundary of box; blue line = mean) for reference, intermediate, and test sites from the validation dataset.

Site Type	Statistic	Result
Reference	95th	45.0
	75th	43.5
	mean	30.3
	25th	18.8
	5th	18.0
Intermediate	95th	45.0
	75th	39.0
	mean	30.2
	25th	24.0
	5th	9.0
Test	95th	21.0
	75th	21.0
	mean	16.2
	25th	12.0
	5th	12.0

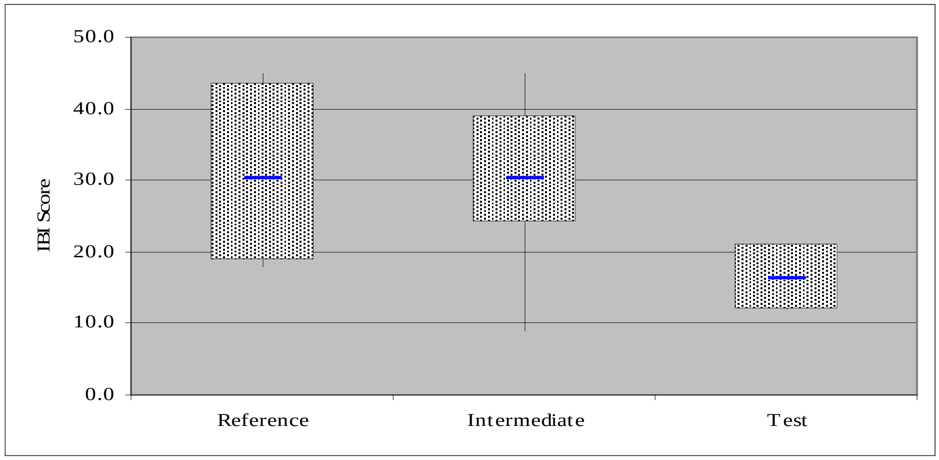
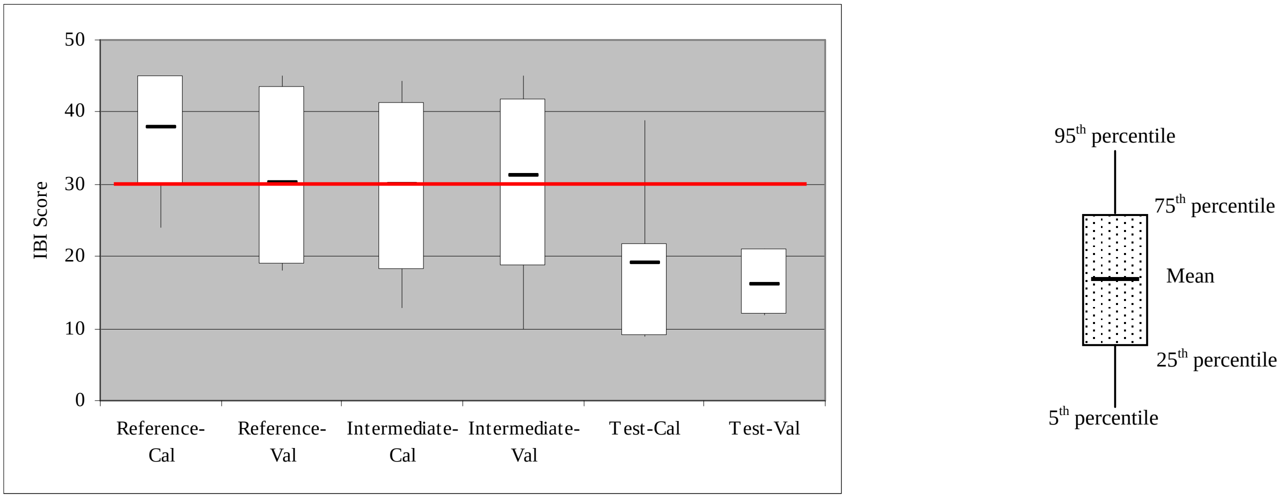


Table 11. Observed and expected frequency of IBI attainment (above threshold) and failure (below threshold) for reference and test sites from the validation dataset. Expected values computed as described in Table 8.

Site Type		# Above threshold	# Below threshold	Total	Chi-square	Critical value (1df, 0.05)
Reference	# observed	5	5	10	5.238	3.841
	# expected	3	7			
Test	# observed	0	5	5		
	# expected	2	3			
	total	5	10	15		

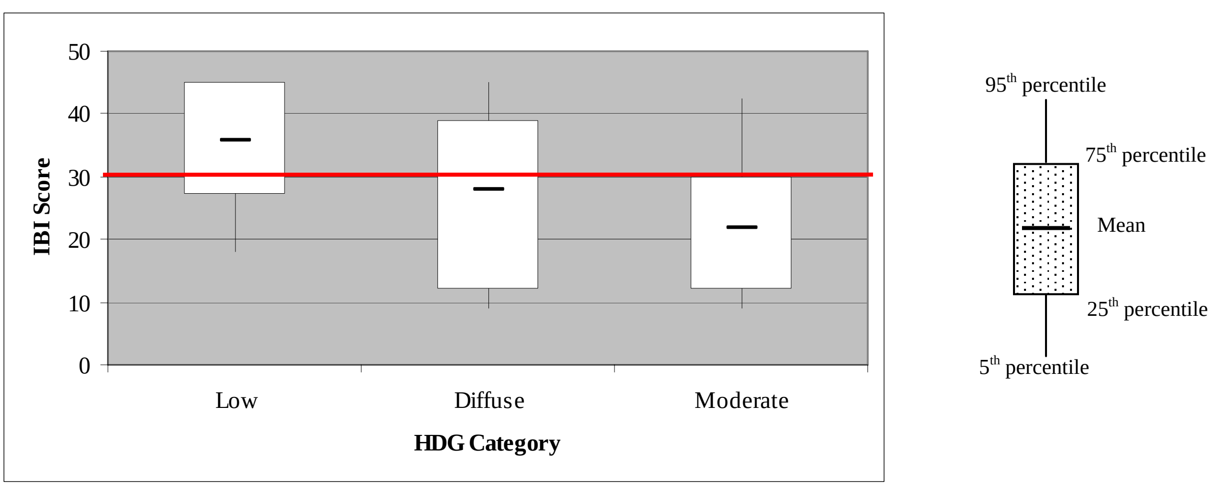
Figure 6. Box and whisker plot of CWIBI scores for sites in the calibration (cal) and validation (val) datasets. Red line indicates proposed CWIBI threshold.



4.7 IBI score vs. level of human disturbance

Following calibration and validation, all data was combined to examine the level of comparability between CWIBI scores and the level of human disturbance. While the determination of reference and test sites for CWIBI development was based on the level of human disturbance, it did not include direct comparisons to coarse HDG categories. The HDG determination process includes the identification of four primary disturbance categories (low, diffuse, moderate, and high). In general, CWIBI scores were higher in the low HDG category as compared to the diffuse or moderate categories (Figure 7). None of the coldwater fish assemblage sites were in the high HDG category.

Figure 7. Box and whisker plot of CWIBI scores associated with HDG categories (Low, Diffuse, Moderate) (includes calibration and validation sites). Red line indicates proposed CWIBI threshold.



The overall percentage of sites in each HDG category with CWIBI scores above or below the recommended threshold indicated that the probability of not attaining the threshold increased from 27 percent to 74 percent for sites with low and moderate levels of human disturbance, respectively

(Table 12). In other words, sites with a moderate level of human disturbance are 2.7 times more likely to fail the CWIBI as proposed than sites with a low level of human disturbance. Conversely, the probability of threshold attainment was much higher for sites with a low (73 percent) vs. moderate (26 percent) level of human disturbance (Table 11). Sites with a low level of human disturbance were 2.8 times more likely to attain or exceed the CWIBI threshold as proposed than sites with a moderate level of human disturbance.

Table 12. Frequency distribution (x100 for percentage of sites) of CWIBI attainment status within HDG categories (A) and probability of CWIBI failure and attainment in a given HDG category (B).

A) Frequency Distribution

	HDG Category			
IBI Status	*Low*	*Diffuse*	*Moderate*	*Total*
IBI attain	0.41	0.14	0.06	0.60
IBI non-attainment	0.15	0.07	0.17	0.40
Total	0.56	0.21	0.23	

B) Non-Attainment and Attainment Probabilities

Probability of IBI Non-Attainment Given HDG Category	
P(IBI nonattainment, given HDG Low)	0.27
P(IBI nonattainment, given HDG Diffuse)	0.35
P(IBI nonattainment, given HDG Moderate)	0.74
Probability of IBI Attainment Given HDG Category	
P(IBI attainment, given HDG Low)	0.73
P(IBI attainment, given HDG Diffuse)	0.65
P(IBI attainment, given HDG Moderate)	0.26

5. SUMMARY AND RECOMMENDATIONS

The proposed CWIBI for New Hampshire wadeable streams showed distinct separation in the scores of reference (mean = 37.9) and test sites (mean = 19.1) in the calibration dataset indicating that the respective composition of the fish assemblages were different. The observed differences in reference and test site CWIBI scores were considered to be a reflection of locally induced human activities and the resultant impacts to the overall ecological integrity of the fish assemblage. Reference sites were dominated by coldwater specialists, especially Eastern brook trout and slimy sculpin, while test sites had higher percentages of generalist feeders and more tolerant species. Independent validation of the CWIBI from a limited number of sites (40 sites) proved encouraging as CWIBI scores showed similar separation between mean reference (30.3) and test (16.2) site scores. The observed number of reference site non-attainments (five) was of minimal concern given the small sample size (10 sites) and the significant difference in the distribution of sites above and below the threshold as compared to test sites (Table 5). As additional new sites are sampled, the CWIBI as proposed will be further validated to strengthen the recommended threshold. It is also important to note that while test sites certainly had a higher level of human disturbance than reference sites, test sites were characterized as “moderately” impacted. A group of test sites restricted to highly impacted sites would likely have shown even greater deviation from the reference condition. In addition, the negative effects of regional environmental impacts to fish communities, such as acid deposition, are difficult, if not impossible to account for, and therefore should be considered as unknown elements that may contribute to lower overall index scores for both reference and test sites.

A total of six metrics were selected to be consistent with previously developed coldwater fish assemblage IBIs for wadeable streams, especially in the State of Vermont (Langdon 2001). Lyons et al. (1996) documented the absence of several common taxonomic groups of fishes from healthy coldwater streams and proposed a five metric IBI that accounts for the lack of species richness in these harsh environments. The selected metrics represent fish feeding strategies, overall tolerance to disturbance, affiliation with coldwater environments, reproductive success, and important indicator species.

All the selected metrics proved useful in independently discriminating between reference and test sites. In particular, the percentages of generalist feeders, coldwater specialists, top carnivores, and Eastern brook trout were the most effective at detecting differences between reference and test sites. In contrast, differences between reference and test sites in the Eastern brook trout age class structure metric were less distinct. However, the concurrent presence of YOY and adults was more frequent at reference than test sites, justifying its inclusion at the current time. A moderate level of correlation between metrics was accepted, especially for the percentage of coldwater specialists, in order to meet the desired number and types of metrics. The coldwater specialists metric accounts for the presence (or lack of) of non-carnivorous species, such as slimy sculpin, a positive indicator of healthy coldwater streams. The percentage of top carnivores was included to account for streams where naturally reproducing populations of brown and rainbow trout occur (see below for further discussion).

A six metric IBI, as recommended here, differs from the 12 metric IBI for more species rich fish assemblages as originally developed by Karr (1981), but is a necessary consequence of the naturally low species diversity of coldwater fish assemblages and decreased amount of potential biological information they contain (Halliwell et al. 1999; Langdon 2001; Lyons et al. 1996). The majority of reference sites included in the development of the New Hampshire CWIBI had between two and four species present, with Eastern brook trout and slimy sculpin comprising nearly two-thirds of the individuals. Thus, while alternative metrics could be exchanged for those recommended, the inclusion of additional metrics would only add redundancy to the index rather than excluded ecological information about the structure and function of the aquatic community. Of the candidate metrics, the metrics chosen for inclusion appear to be the most descriptive of New Hampshire’s coldwater fish assemblage and most expressive of declining aquatic community structure and function.

It is also important to note that the index, as constructed, is only applicable to strict coldwater fish assemblages and not all wadeable streams capable of supporting coldwater fish species. Based on the data available these strict coldwater fish communities are species poor (two to four species) and only occur in small (less than 15 square miles) watersheds either in northern New Hampshire (greater than 43.75 degrees latitude), from the Connecticut River drainage, or at higher (less than 775 feet) elevations elsewhere throughout the state. Naturally occurring populations of coldwater fish species are supported in larger streams throughout New Hampshire. However, these streams and their associated fish assemblage, termed 'cool water' fish communities, generally consist of five or more species and are outside the limits of the CWIBI.

Naturally reproducing populations of non-native salmonids, namely brown and rainbow trout, were included in the computation of the percentage of top carnivores metric in order to reflect their utility as positive indicators of habitat and water quality (Langdon 2001). For this metric, brown and rainbow trout are to be included regardless of the occurrence of Eastern brook trout. Unlike the VT DEC CWIBI, brown and rainbow trout are not to be included in the percentage of coldwater specialists metric for the NH CWIBI as their inclusion would increase the level of redundancy with the top carnivore metric. The decision to include consideration for the presence of brown and rainbow trout only in the top carnivore metric was subjective and largely based on the relatively low frequency of their occurrence. Naturally reproducing brown and rainbow trout populations in New Hampshire streams are uncommon, but when they do occur, their inclusion into the top carnivore metric provides important information about the assemblage's condition that would otherwise be overlooked. Populations of these species are most commonly found in northern tributaries to the Connecticut River and tributaries of the Androscoggin River. The proposed CWIBI still penalizes a stream lacking Eastern brook trout through the percentage of Eastern brook trout and age class metrics. Thus, the CWIBI as proposed here represents an index focused on the exclusion of non-native species, except for a single metric where their inclusion serves as a positive indicator of aquatic community condition.

In order to further resolve questions regarding the origin of Eastern brook trout individuals, future sampling efforts should make attempts to distinguish stocked from wild fish. As noted by Halliwell et al. (1999), the presence of stocked individuals may only be temporary and reflects a decision by fishery managers rather than a stream's ability to support a natural population over the long term. In its current form, the CWIBI developed for wadeable New Hampshire streams did not distinguish between wild (i.e. naturally occurring) and stocked fish. However, based on experience, this oversight is not believed to be problematic as many of the reference sites are not commonly subjected to regular stockings and in all likelihood are supportive of naturally occurring individuals.

The exclusion of all Atlantic salmon from index development was also considered following data analysis as concerns remained in regards to the loss of potentially informative data. A review of the dataset showed that Atlantic salmon were collected at just nine of all 81 sites included in the analysis. Of these nine sites only two failed to meet the recommend the CWIBI threshold of 30. Thus, the available data indicate that the exclusion of this species had a minimal effect on CWIBI development and resulting index outcomes, especially when taken in consideration with their origin (artificial introduction).

A threshold score of 30 is recommended for the CWIBI as it reflects the 25th percentile score of established reference sites. The 25th percentile target is consistent with the threshold identification process used for the development of the benthic IBI used by DES. When applied to test sites, six of eight (75 percent) fell below the threshold. In contrast, only five of 27 (19 percent) reference sites fell below the recommended criteria. From an operational standpoint the recommended threshold requires that sites meeting or exceeding the threshold must not fall in the lowest scoring category (1.5) for more than two metrics. For sites where three of six metrics fall in the lowest scoring category the maximum overall index score would be 27 even if the remaining metrics fall in the highest scoring category (7.5) [(3 metrics x 1.5)+(3 metrics x 7.5)=27].

When comparing among HDG categories, sites with a low level of human disturbance had a 73 percent chance of threshold attainment (27 percent chance non-attainment) while sites with a moderate level of human disturbance only had a 26 percent chance of meeting or exceeding the threshold (74 percent chance non-attainment). These results indicate that the threshold strikes an appropriate balance between over or under protection of the resource. An overly protective threshold would have a high incidence of non-attainment for sites with a low level of disturbance. In contrast, an under protective threshold would have a high incidence of attainment for sites with a higher level of disturbance. The recommended threshold might, however, require adjustment over time as reference sites are better defined or additional data indicate an adjustment is warranted.

The identification of wadeable streams or stream segments where the CWIBI is applicable should follow the recommendations set forth in this report and shown in Map 1. The New Hampshire coldwater fish assemblage criteria defined herein includes wadeable tributaries to the Connecticut River and wadeable stream reaches north of 43.75 degrees latitude that have watershed areas less than 15 square miles, as well other wadeable streams throughout the state meeting this drainage requirement and at elevations greater than 775 feet. The New Hampshire coldwater fish assemblage criteria are considered to be a reasonable starting point. Of the 37 reference sites included in the analysis, Eastern brook trout were found at 35, providing a strong indication that the boundaries established by the criteria are not beyond the expected distribution of this indicator species. However, this fact does not preclude the need for future modification (i.e. expansion) of the boundaries over time as more data becomes available. The criteria also allow for the application of the CWIBI to sites outside the identified boundaries if naturally reproducing populations of Eastern brook trout or slimy sculpin are documented and there are no more than 4 naturally occurring species.

In Vermont, the VT DEC defines the basic guidelines for the identification of coldwater fish assemblages as wadeable stream tributaries to the Connecticut River (excluding the lower West River) and streams reaches in excess of 500 feet in elevation. When compared, for the data that is available, stream reaches in Vermont used in the CWIBI development process had moderately smaller mean drainage areas (Vt. = ~5 square miles; N.H. = ~11 square miles) and were at slightly lower mean elevations (Vt. = ~1,000 feet; N.H. = ~1,100 feet) (VT DEC 2004). While the interstate definitions differ somewhat in regards to where coldwater fish assemblages should naturally occur, they are not radically different, and to some extent, lend support each state's independent findings; namely that coldwater streams tend to be defined by small drainage areas and occur at higher elevations or more northerly latitudes. For both states, undisturbed coldwater streams have fish assemblages that are species poor (generally less than or equal to four species) and include Eastern brook trout as the only native carnivore, one or more insectivorous fish species (usually slimy sculpin), and one or more fish species that is a generalist feeder. The similarity among coldwater fish assemblages found in Vermont and New Hampshire also, in part, justifies DES's CWIBI development approach and the resultant selection of four of the same metrics (percent coldwater specialists, percent top carnivores, percent generalist feeders, and the Eastern brook trout age class metric).

The CWIBI proposed here will serve as a partial numeric interpretation of DES's current narrative water quality criteria relating to the biological integrity (Env-Ws 1703.19) of aquatic communities for 1st to 4th order wadeable streams meeting the definition of a coldwater fish assemblage. The index is designed to accurately and precisely describe the structural and functional status of coldwater fish assemblages through the compilation of six metrics. Other indices, such as DES' benthic IBI may be coupled with the CWIBI towards the determination of aquatic life use status for the purposes of completing federally-required water quality reports, state-level regulatory actions, and general water quality planning activities. Finally, as noted by Hughes et al. (2004), natural disturbances, unrelated to human activity, can cause temporary impacts to ecological communities, and additional investigation may be warranted before formal aquatic life use “impairment” listing.

Similar to the recommendation made by Langdon (2001), future work on the CWIBI should include an understanding of the natural variation associated with stream fish population dynamics and how these are manifested in index scores. Additionally, more rigorous collection of life stage (i.e. YOY, adult) data for Eastern brook trout should improve the discriminatory power of this metric. Finally, the CWIBI is just one of three fish indices that will ultimately be available to assess the status of wadable stream fish assemblages. The remaining two indices will be developed in a similar fashion and be used to characterize the quality of naturally “warm” and “cool” water wadable streams across the state.

6. REFERENCES

Barbour, M. T., J. Gerristen, B. D. Snyder, and J. B. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish, second edition. EPA 841-B-99-002. U. S. Environmental Protection Agency; Office of Water; Washington, D.C.

Halliwell, D. B., R. W. Langdon, R. A. Daniels, J. P. Kurtenbach, and R. A. Jacobson. 1999. Classification of freshwater fish species of the Northeastern United States for use in the development of indices of biological integrity, with regional applications. Pages 301-333 *in* T. P. Simon, editor. Assessing the sustainability and biological integrity of water resources using fish communities. CRC Press, Boca Raton, Florida.

Hughes, R. M., S. Howlin, and P. R. Kaufmann. 2004. A biointegrity index (IBI) for coldwater streams of western Oregon and Washington. Transactions of the American Fisheries Society 133:1497-1515.

Karr, J. R. 1981. Assessment of biotic integrity using fish communities. Fisheries 6:28-30.

Mundahl, N. D. and T. P. Simon. 1999. Development and application of an index of biotic integrity for coldwater streams of the upper Midwestern United States. Pages 383-411 *in* T. P. Simon, editor. Assessing the sustainability and biological integrity of water resources using fish communities. CRC Press, Boca Raton, Florida.

Langdon, R. W. 2001. A preliminary index of biological integrity for fish assemblages of small coldwater streams in Vermont. Northeastern Naturalist 8:219:232.

Leonard, P. M. and D. J. Orth. 1986. Application and testing of an index of biotic integrity in small, coolwater streams. Transactions of the American Fisheries Society 115:401-414.

Lyons, J., L. Wang, and T. D. Simonson. 1996. Development and validation of an index of biotic integrity for coldwater streams in Wisconsin. North American Journal of Fisheries Management 16:241-256.

Stribling, J. B., B. K. Jessup, and J. Gerristen. 2000. Development of biological and physical habitat criteria for Wyoming streams and their use in the TMDL process. March 2000. Tetra Tech, Inc., Owings Mills, MD. 46pp.

VT DEC. 2004. Biocriteria for fish and macroinvertebrate assemblages in Vermont wadeable streams and rivers – Development phase. Agency report. Water Quality Division, Biomonitoring and Aquatic Studies Section, Waterbury, Vermont.

APPENDIX A. Human disturbance gradient (HDG) variables, associated scoring criteria, and applicable spatial scale.

Human Disturbance Scale Scoring Category Breakdown				
Variable	**Category (density)**[1]	**Score**	**Assessment Scale**[4]	
			Biomon shed	**1-mile radius**
GWHI	0	0		x
	<2	1		
	<5	2		
	>/=5	3		
RCRA	<2	0		x
	<5	1		
	<7	2		
	>/=7	3		
Junkyards	0	0		x
	<1	1		
	<3	2		
	>/=3	3		
# dams	N	0		x
	Y, > 10 acres	1		
	Y, > 50 acres	2		
	Y, > 100 acres	3		
# waterusers	0	0		x
	<1	1		
	<3	2		
	>/=3	3		
NPDES	0	0		x
	<1	1		
	<2	2		
	>/=2	3		
Road Density	<1	0		x
	<2	1		
	<3	2		
	>/=3	3		
Impounded[2]	0	0	x	
	<.02	1		
	<.05	2		
	>/=.05	3		
NHLC Developed (%)[3]	<2	0	x	
	<4	1		
	<8	2		
	>/=8	3		
NHLC Ag (%)[3]	<2	0	x	
	<4	1		
	<8	2		
	>/=8	3		

<u>Notes:</u>

1 - # / sq mi. - GWHI, RCRA, Junkyards,Waterusers, # dams

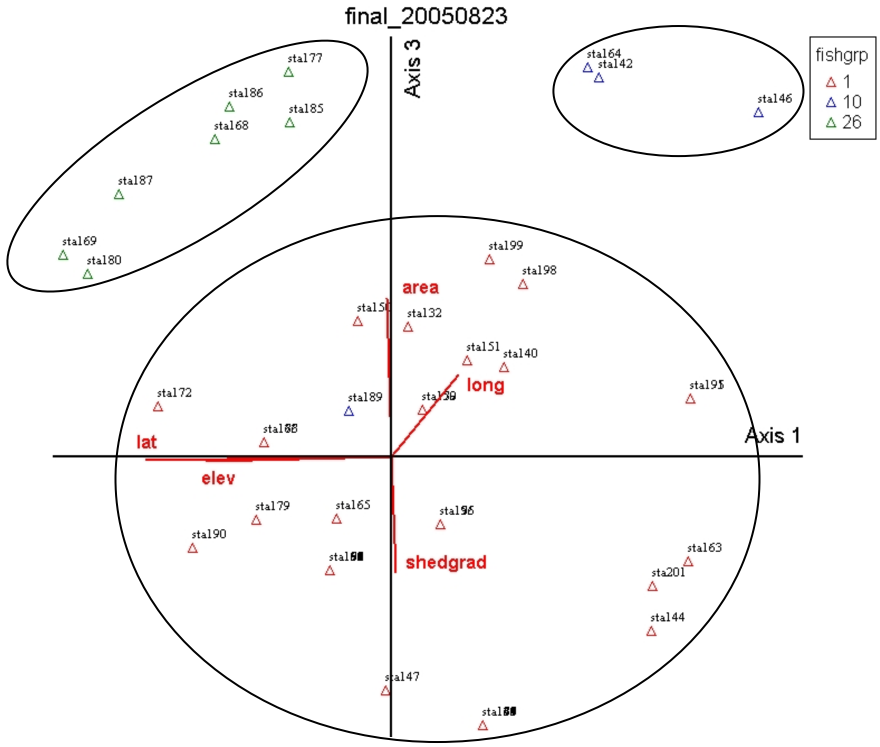
2 - Impounded - % as acres impounded / acres of watershed for "Biomon shed"; as total number acres impounded for "1-mile radius"

3 - % - of landuse type for NHLC developed & Ag

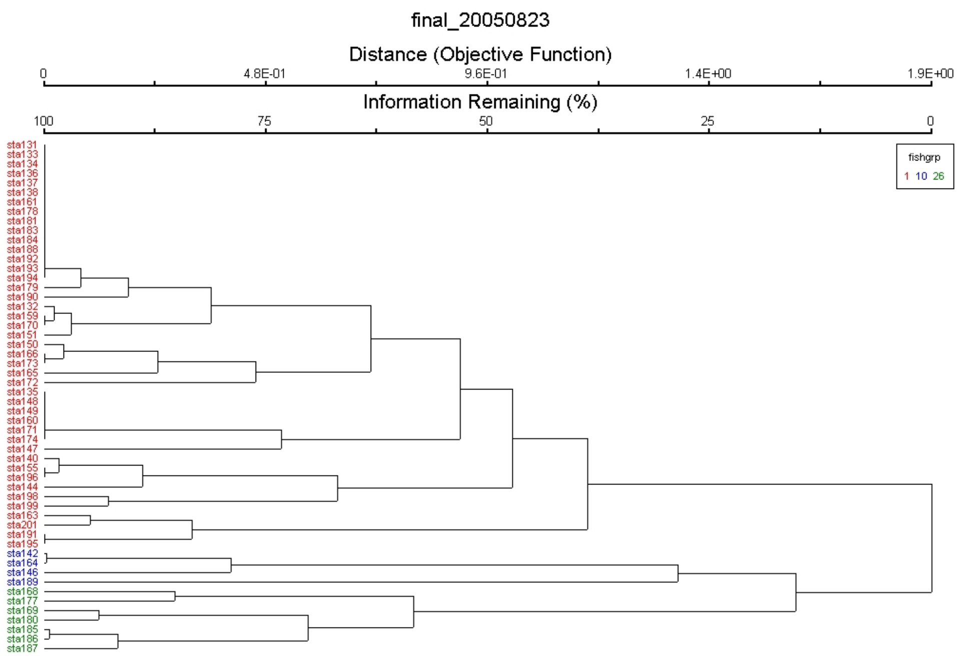
4 - x - indicates assessment scale assigned to scoring categories

APPENDIX B. Nonmetric multidimensional scaling ordination of biomonitoring reference sites (N= 54) in species space (A) with result of cluster analysis (B) overlayed onto sites. For (A) relative importance of environmental variables is indicate by length of line and affiliation of environmental variables with Axis 1 and 3 is indicated by relative nearness of line to axis. Preliminary assemblage type were as follows: Fishgrp 1 = coldwater fish assemblage; Fishgrp 10 = cool water fish assemblage; Fishgrp 26 = warm water fish assemblage.

A)



B)



APPENDIX C. Individual metric means, medians, and selected percentiles of reference and test sites from the calibration dataset.

Metric	Mean - test	Mean - reference	Meadian - test	Meadian - reference	25th percentile - test	25th percentile - reference	75th percentile - test	75th percentile - reference
CONTINUOUS METRICS								
per_CW	45.58	72.49	40.16	100.00	13.00	33.98	72.81	100.00
per_BI	22.18	23.33	19.26	15.50	1.00	0.35	30.58	37.15
per_GF	11.28	6.74	4.24	0.00	0.45	0.00	17.15	0.00
per_OI	36.09	17.89	37.84	0.00	6.73	0.00	59.04	25.08
per_TC	30.46	52.05	13.95	55.16	7.93	15.66	54.25	88.33
per_I	43.49	70.13	40.16	93.61	13.00	33.98	71.03	100.00
per_M	14.39	8.36	12.43	0.00	0.00	0.00	26.81	6.23
per_T	42.12	21.51	47.82	0.00	6.73	0.00	61.99	43.35
per_ebt	22.63	47.95	8.20	53.33	5.92	11.05	33.22	76.41
per_ss	13.04	18.28	4.21	12.26	0.00	0.35	20.81	35.97
EBT_CPUE	1.15	1.43	0.34	1.24	0.17	0.34	0.80	1.95
Total_CPUE	5.70	4.33	4.68	4.01	2.81	1.87	7.27	6.67
DISCRETE METRICS								
Num_intol_sp	1.4	1.7	1.5	2.0	1.0	1.3	2.0	2.0
EBT_age_class	1.9	2.3	1.5	3.0	1.0	1.0	3.0	3.0
Num_tol_sp	1.7	0.7	2.0	0.0	0.3	0.0	2.0	1.0