Development of Aquatic Life Use Support Attainment Thresholds for Florida's Stream Condition Index and Lake Vegetation Index

FDEP
Standards and Assessment Section
Bureau of Assessment and Restoration Support

DEP-SAS-003/11

October 2011

Contents

1 Introduction

The ability to measure whether a water body's aquatic community meets the objective of the Clean Water Act (CWA) is critical for informing decisions related to implementation of the State and Federal water quality programs. In particular, establishing biological assessment methods as part of state water quality standards can be very valuable for making designated use attainment decisions for aquatic life use support, which equates to attainment of the CWA goal regarding biological integrity/health. Given the complexities of measuring biological structure and function, it is important that the decision regarding attainment or non-attainment of the designated use be derived as a scientifically defensible, quantitatively determined threshold.

This document describes the factors necessary to consider when developing a quantitative measure of biological health, and describes the basis for the State of Florida's rationale for establishing the appropriate biological thresholds for the Stream Condition Index and Lake Vegetation Index, which in turn are used to indicate attainment of the designated use.

2 Stream Condition Index

2.1 Historical Background

The response of macroinvertebrate communities to human point source pollution began receiving attention in Florida during the late 1950's. In 1958, Bill Beck, biologist with the Florida State Board of Health, wrote a series of "Biological Letters", where he introduced the concept of using invertebrates as biological indicators, especially for demonstrating the effects of excess organic matter on streams and lakes (the saprobium index concept). What became known as "Beck's Biotic Index" was developed by sampling invertebrates at control sites located upstream of point source discharges and observing which sensitive taxa were eliminated at sites downstream of the effluent sources (Beck 1954). Concurrently, there typically was a dramatic increase in abundance of tolerant taxa, such as "bloodworms" (certain species of chironomid midges) as illustrated in **Figure 1**.

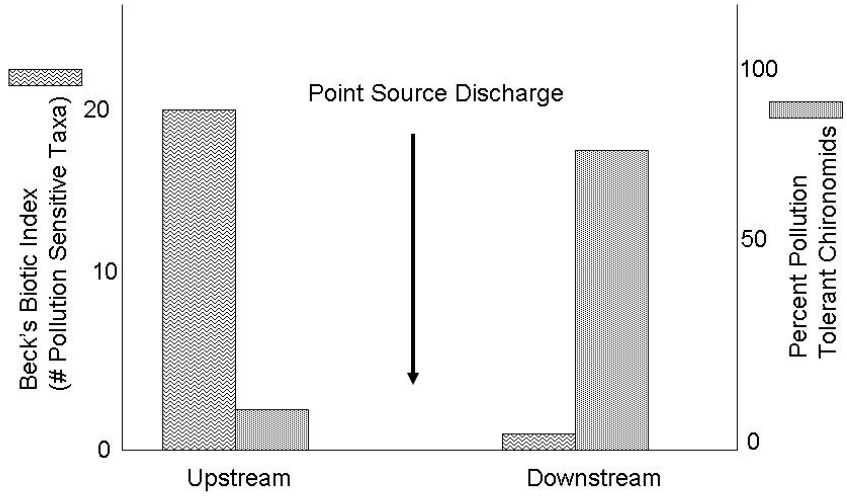

Figure 1. Typical macroinvertebrate response to organic loading associated with primary wastewater treatment typical in the 1950s and 1960s.

In the 1970s and 1980s, macroinvertebrates were routinely sampled via multi-plate artificial substrate samplers (Hester-Dendys). Hester-Dendy samplers are incubated in the receiving waters for 28 days, a minimum period of time for colonization by a representative macroinvertebrate community (**Figure 2**). The Shannon-Weaver diversity index, a biological metric derived from information theory, became a popular method to communicate complicated biological results. The Shannon-Weaver diversity index is based upon a combination of the taxa richness at a site and the equitably of the distribution of abundance of individuals. Low diversity scores represent conditions where a few organisms are very abundant, to the exclusion of other taxa. This index is specified in the Florida Administrative Code as a measure of biological integrity (Rule 62-302.530 FAC). It generally has been applied by comparing site-specific control sites to nearby test sites.

In 1992, EPA promulgated the concept of "rapid bioassessment." Regional expectations (generally eco-regions) for biological communities were established by sampling "minimally disturbed" reference sites. Metrics, defined as measures of biological health which respond in a predictable manner to human disturbance, were calculated from the raw reference site data. Next, a distribution of the reference site metric values was calculated, and a lower percentile of the reference site distribution was selected to represent the expectations for that metric in a minimally disturbed condition. A variety of metrics would then be combined into a dimensionless index. This was accomplished by assigning points to individual metrics based on their relative similarity to the reference condition, and summing the points.

Figure 2. Photo of Hester-Dendy samplers used for determining the Shannon-Weaver diversity index.

The current Stream Condition Index was built upon the 1990s concepts. The main improvement in the present index is the use of a human disturbance gradient to determine effective metrics and then determining impairment thresholds by using a combination of the Biological Condition Gradient (BCG) and reference site approaches. The BCG employs a group of experts to individually review species level data and determine the site's ecological status (see section 2.4.2). Further discussion of the present Stream Condition Index occurs below.

2.2 Bioassessment Theory

To successfully manage ecosystems, a basic understanding of the system's biological components is critical. The biota respond to a wide variety of cumulative factors, both natural and anthropogenic (**Figure 3**). As the organisms integrate these factors over time, a characteristic community structure emerges. When human actions adversely affect a system, the biological population will change, leading to an impaired or imbalanced community. For example, pollution sensitive taxa will disappear, food webs will be disturbed, taxa richness and diversity usually will decrease, and undesirable nuisance species may dominate.

To accurately determine when humans have negatively affected a biological community, one must be familiar with the structure and function of natural, or "reference" systems in a given geographical region (Griffith *et al.* 1994; **Figure 4**). FDEP has selected reference sites using a rigorous, objective method, ensuring that they are subject to no, or very minimal, human disturbance (see Section 2.4.1 below). Based on reference site community similarity, three SCI Bioregions have been established for Florida: the Panhandle, the Northeast, and the Peninsula (note that the SCI is not calibrated for Ecoregion 76, the Southern Florida Coastal Plain, where few natural streams exist). First, it is important to establish the normal or typical range of certain key measures of community health at these reference systems, often thought of as "biological integrity". Karr (1991) defined biological integrity as the ability of an aquatic ecosystem to support and maintain a balanced, adaptive community of organisms having:

- species composition,
- diversity, and
- functional organization comparable to that of natural habitats within a region.

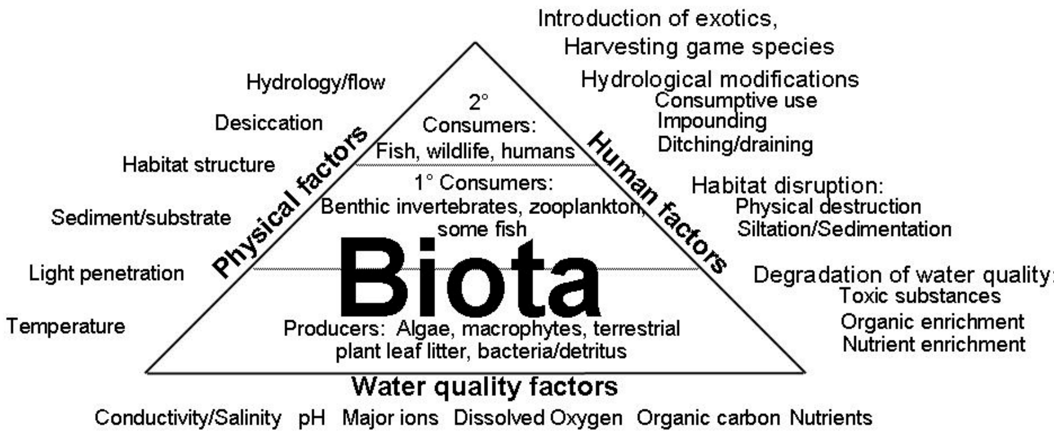

Figure 3. Many factors affect biological community composition. To conclude that human factors are primarily responsible for biological degradation, reasonable knowledge of the influence of natural factors is essential.

Measures (or attributes) of ecosystem health that respond predictably to human influence are termed **metrics**. Metrics from reference sites are compared with the same metrics from an unknown or "test" system to determine unacceptable departures from the expected condition associated with human impairment. To be scientifically defensible, the systems being compared should be similar except for potential human influences (compare streams to streams, not streams to systems with lake-like conditions). Additionally, one or more natural stressors (*e.g.*, flood, drought, low substrate diversity, periodic natural low dissolved oxygen, etc.) may affect sampling sites, even those sites with minimal disturbance from humans. These natural stressors should be reasonably understood and controlled for in the sampling design to more conclusively determine when human actions have caused biological degradation.

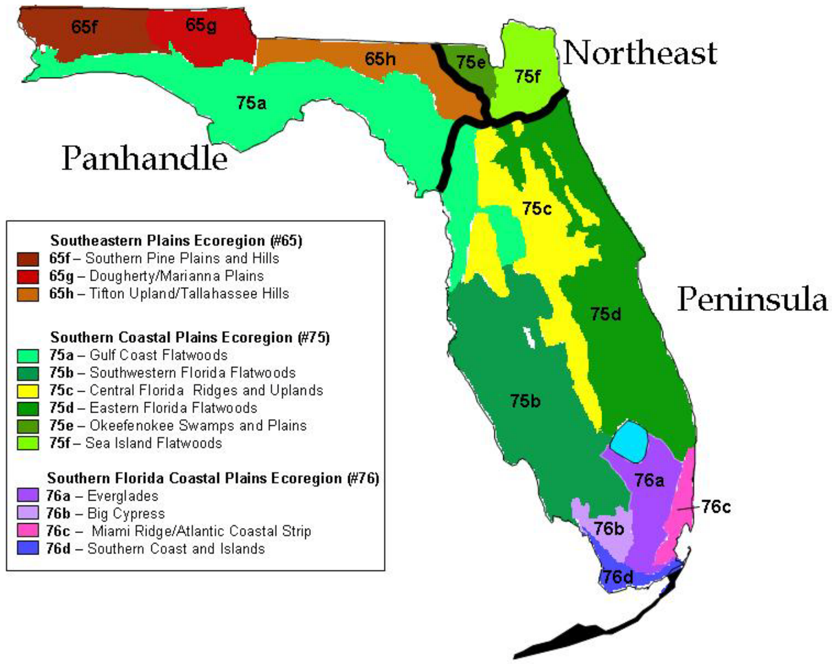

Figure 4. Sub-ecoregions of Florida, which were aggregated into 3 bioregions, based on multivariate measures of taxonomic similarity.

2.3 Development of Stream Condition Index Metrics using the Human Disturbance Gradient

DEP has utilized a Human Disturbance Gradient approach to allow for the objective selection of metrics (Fore *et al.* 2007a). The Human Disturbance Gradient is composed of four factors:

- The Landscape Development Intensity Index (Brown and Vivas 2004);
- Habitat Assessment scores (DEP SOPs);
- Hydrologic Modification Index; and
- Water column ammonia concentration.

These components, described in detail by Fore *et al.* (2007a), were converted into a dimensionless index, with low values denoting low disturbance and increasing values associated with more intense human influences. The index was subsequently used as the x-axis for testing a wide variety of biological attributes associated with the measurement of ecological integrity (**Figure 5**). **Figure 6** depicts the absolute value of correlation coefficients (Spearman's r) for a variety of biological attributes against the HDG. Once an attribute is demonstrated to respond predictably to human influence, it is termed a metric**.** The 10 selected metrics were chosen to:

- Represent as many attribute categories as possible;
- Provide meaningful and predictable assessment of human effects; and
- Avoid redundancy if several correlated metrics were providing similar information.

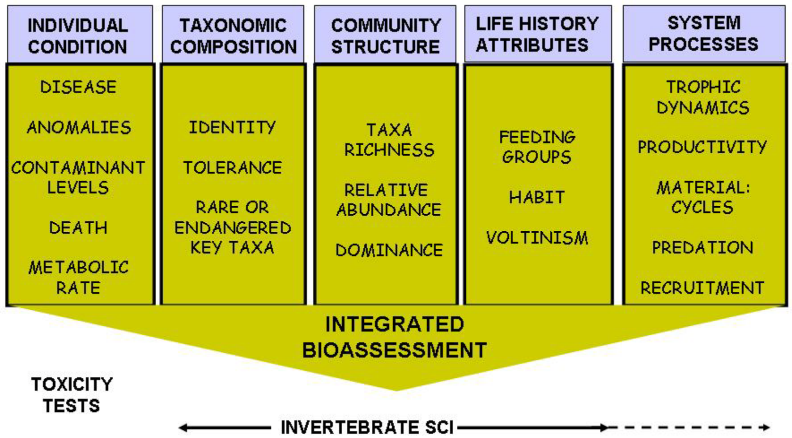

Figure 5. Major attribute categories and example metric, for determining biological integrity.

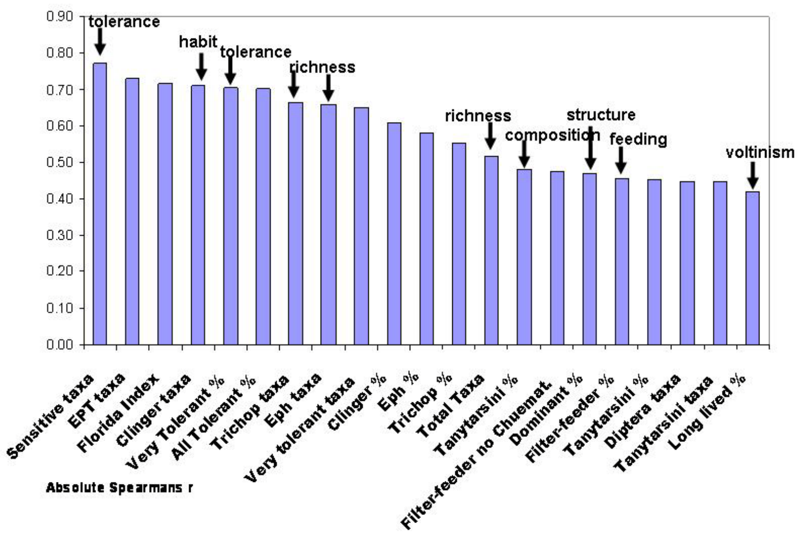

Figure 6. Correlations between various metrics and the Human Disturbance Gradient. Arrows indicate metrics selected for the SCI, and associated attribute group.

These ten metrics comprise the SCI: total number of taxa, number of mayfly (Ephemeroptera) taxa, number of caddisfly (Trichoptera) taxa, percent filter feeder individuals, number of long-lived taxa, number of clinger taxa, percent dominant taxa individuals, percent Tanytarsini midge individuals, number of sensitive taxa, and percent very tolerant taxa individuals. Metrics are converted to a score from 0-10, and then the scores are added to yield the final SCI score. The following is a brief description of the metrics, divided into several metric types as shown in **Figure 5**.

2.3.1 Taxonomic composition

2.3.1.1 Identity

Tanytarsini midges are sensitive to disturbance, so the % Tanytarsini metric was included in the SCI as the best available measure of the chironomid assemblage, which is an important group in stream invertebrate communities.

2.3.1.2 Sensitivity and Tolerance

Lists of sensitive and very tolerant macroinvertebrates were established by analyzing the responses of individual species to the HDG (Fore *et al.* 2007a). The number of taxa selected as sensitive equaled around 12% of the taxa tested, and the number of very tolerant taxa was approximately 10% of the taxa tested. Many sensitive species belonged to the Ephemeroptera, Trichoptera or Odonata; several chironomids were also included. All the Plecoptera were included as sensitive taxa. The number of sensitive taxa and the percent very tolerant individuals were highly correlated with the HDG (**Figures 7 and 8**).

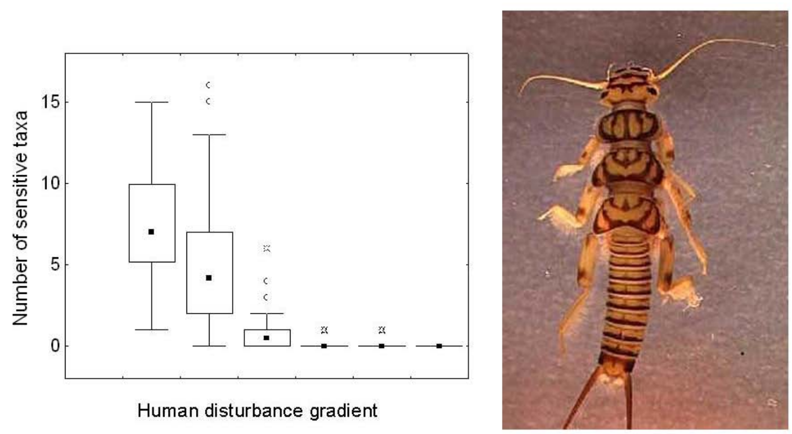

Figure 7. Response of the number of sensitive taxa metric to the HDG. The photo is of a Plecopteran (stonefly).

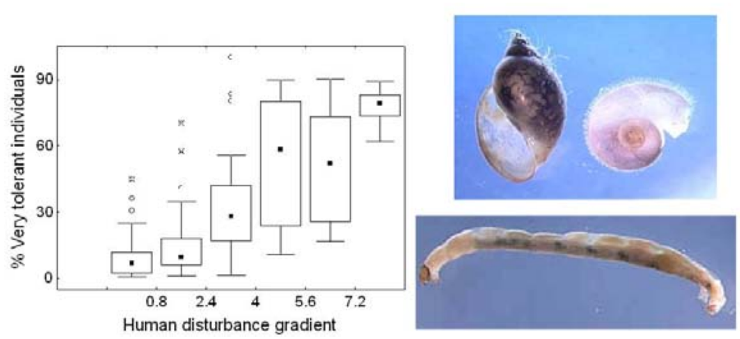

Figure 8. Response of the percent very tolerant metric to the HDG. Photos are of lunged snails and tolerant midges.

2.3.2 Community Structure

2.3.2.1 Taxa richness

Total taxa richness (the number of different types of organisms present) and the richness of the Trichoptera (caddisflies) and Ephemeroptera (mayflies) historically have been shown to decrease with human disturbance. **Figure 9** depicts the response of the number of Ephemeroptera taxa to human disturbance, which is similar to the response of the Trichoptera taxa and total taxa metrics. These three measures were chosen since each metric may respond differently, depending on the type of disturbance (*e.g.*, mayflies are more sensitive to metals, certain caddisflies may be more sensitive to flow disruption).

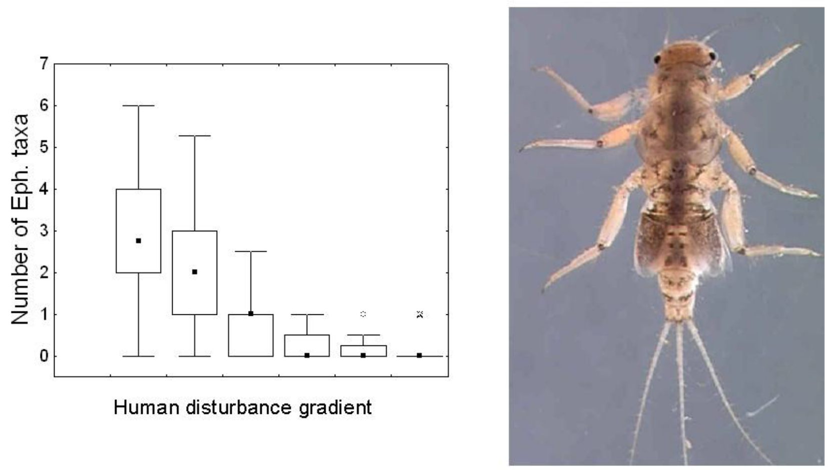

Figure 9. Response of the Ephemeroptera metric to the HDG. The photo is of Tricorythodes, a sensitive mayfly.

2.3.2.2 Dominance

Substantial shifts in proportions of major groups of organisms, compared to reference conditions, may indicate degradation. The percent dominant taxon, which increases in conditions where a few pollution tolerant organisms are very abundant, to the exclusion of other taxa, was selected as a metric.

2.3.3 Life history attributes

2.3.3.1 Feeding groups

Disruption of food webs has long been associated with human influence, especially organic pollution. Of the functional feeding group measures, the relative abundance of filterers or suspension feeders (percentage of filterer individuals) had the highest correlation and most consistent relationship with the HDG (**Figure 10**). Filter feeders extract nutrients by straining food particles from the water column. If the water flow or quality of the organic matter in the water is compromised, a reduction in filter feeders will occur.

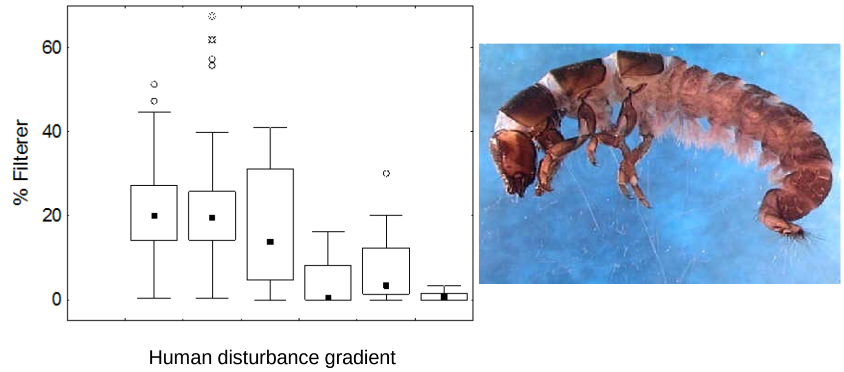

Figure 10. Response of the percent filter-feeder metric to the HDG. The photo is of a net-spinning caddisfly.

2.3.3.2 Voltinism

Voltinism refers to the number of distinct reproductive cycles for a given organism that may take place in a year. Long-lived taxa include semi-voltine insects and non-insects that require greater than one year to complete their life cycles. Long-lived taxa richness would be expected to decrease if a disturbance event (*e.g.*, sporadic illegal dumping, periodic pulses of chemicals from rain events) occurred at a site within a year of sample collection (**Figure 11**).

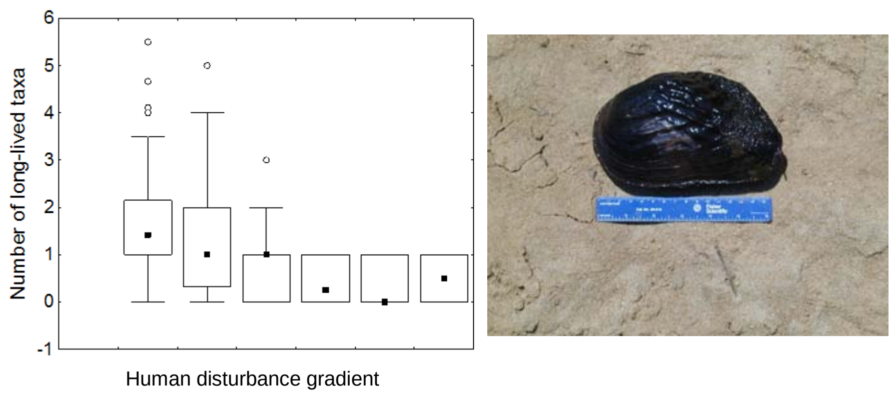

Figure 11. Response of the long-lived taxa metric to the HDG. The photo is of a mollusk, the threatened purple bank climber.

2.3.3.3 Habit

Clingers are those taxa morphologically adapted to hold onto substrates during routine flow conditions and would be expected to decline as humans alter a stream's hydrograph (*e.g,.* channelization), especially during abrasive events caused by high stormwater inputs from impervious surfaces. Clinger taxa richness was highly correlated with the HDG (**Figure 12**).

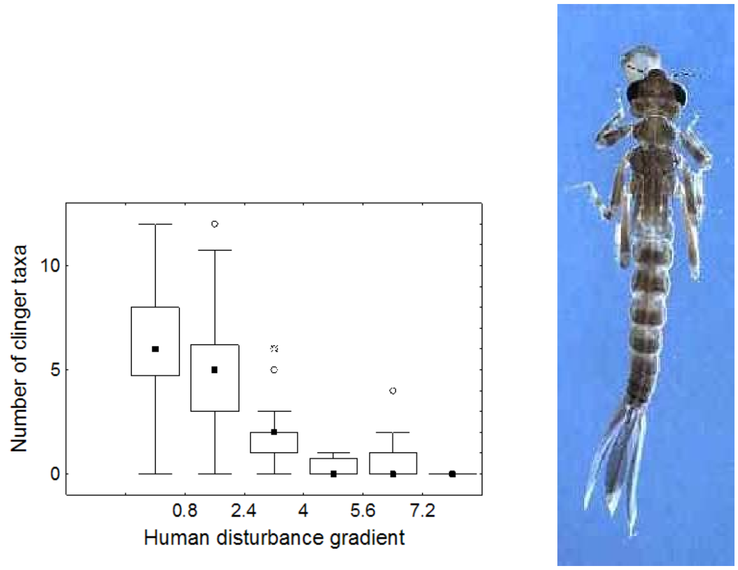

Figure 12. Response of the clinger taxa metric to the HDG. The photo is of a damselfly larvae.

2.4 Establishing Expectations for Aquatic Life Use

2.4.1 Application of the Reference Site Approach

A distinct advantage of biological assessment is that aquatic life use support may be directly measured and a minimum threshold consistent with maintaining a healthy, well-balanced aquatic community may be quantitatively established. In 2007, DEP calibrated the SCI using primarily the Biological Condition Gradient approach (secondarily the reference site approach), resulting in a minimum threshold for aquatic life use support of 34 and an exceptional threshold of 67 (see discussion below). Subsequent EPA review resulted in the recommendation that Florida use an examination of the lower distribution of reference sites as the principal line of evidence for establishing aquatic life use support thresholds, in combination with the Biological Condition Gradient approach.

In response to this request, DEP conducted statistical interval and equivalence tests with SCI data from 55 reference streams (predominantly consisting of the rigorously verified reference sites used for nutrient criteria development with additional data from the Fore *et al.* (2007a) analysis). The purpose of these tests was to determine the lower bounds of the reference site distribution of SCI scores while balancing Type I errors (falsely calling a reference site impaired) and Type II errors (failing to detect that a site is truly impaired) (**Table 1**). The examination of the two most recent visits at 55 reference streams showed that the 2.5th percentile of reference data was in the range of 35-44 points with 95% confidence. The middle of this range was 40 points, which represents a minimum threshold for aquatic life use support that balances Type I and Type II errors.

When calibrating a minimum threshold for aquatic life use support for an index, the amount of human disturbance inherent at the reference sites is a major issue. Some states select reference sites based on the "best available condition" (which may have substantial disturbance), using a Best Professional Judgment approach. Florida has employed a rigorous reference site selection approach, which objectively demonstrates the "minimally disturbed" (limited human influence) nature of Florida's reference sites. When establishing a minimum threshold for aquatic life use support using a lower distribution of reference sites, a rigorous reference site selection process provides greatly increased confidence that the reference site population is minimally disturbed, thereby significantly reducing Type II errors (*i.e.,* classifying impaired sites as healthy). This increased confidence also allows for establishing the minimum threshold for aquatic life use support at a low level of the reference site distribution to minimize Type I errors (classifying healthy sites as impaired)**.**

In the proposed threshold for the SCI, impairment will be determined by the average of two or more consecutive SCIs, so the threshold determined from the interval and equivalence tests (40, based on an average of two SCIs) is closely aligned with the assessment methods. An impairment threshold of 40 would result in approximately 2.5 % of reference sites (known to be minimally disturbed) to be deemed impaired. DEP believes that this threshold is consistent with the Clean Water Act aquatic life use support goal and complies with Florida law (Chapter 403, F.S.).

Table 1. Results of interval and equivalence tests conducted on reference sites with 2 SCI results. Shown are site mean, minimum aquatic life use support threshold based on an average of two SCIs, and range for threshold values defined at the 2.5th and 5th percentile of reference sites ($p < 0.05$; N = 55 reference sites with two averaged SCI values for each site). Reference site values from Fore *et al.* (2007a) and comprehensively nutrient benchmark verified sites.

Threshold (description)	Reference Site mean	Minimum Aquatic Life Use Support Threshold (average of two results)	Not Reference Condition (95% Confidence)	Undetermined	Reference Condition (95% Confidence)
2.5th percentile of reference	65	40	<35	35–44	>44
5th percentile of reference	65	44	<39	39–47	>47

2.4.2 Biological Condition Gradient Approach

The U.S. EPA has outlined a tiered system of aquatic life use designation, along a Biological Condition Gradient (BCG), that illustrates how ecological attributes change in response to increasing levels of human disturbance. The BCG is a conceptual model that assigns the relative health of aquatic communities into one of six categories, from natural to severely changed (**Figure 13**). It is based in fundamental ecological principles and has been extensively verified by aquatic biologists throughout the U.S.

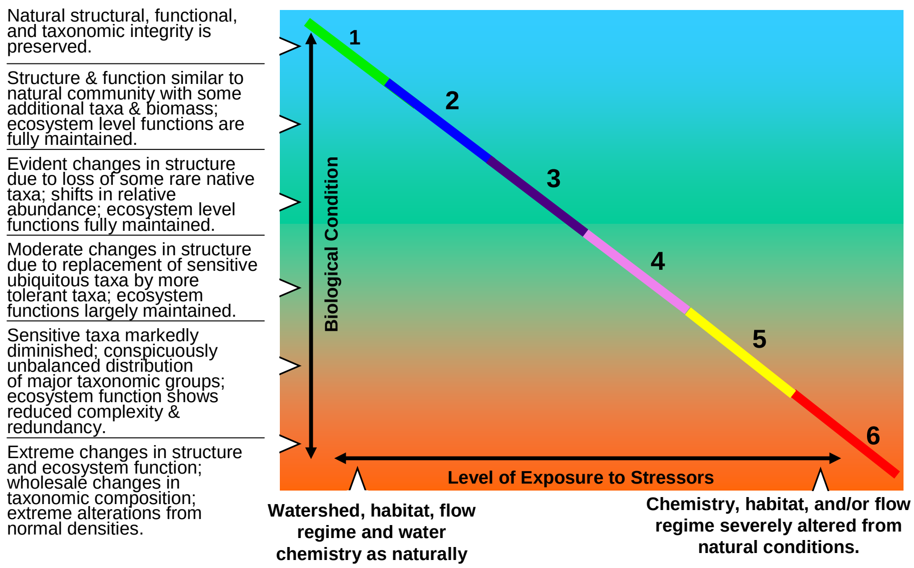

Figure 13. The Biological Condition Gradient Model (from Davies and Jackson 2006).

The BCG utilizes biological attributes of aquatic systems that respond predictably to increasing pollution and human disturbance. While these attributes are measurable, some are not routinely quantified in monitoring programs (*e.g.*, rate measurements such as productivity), but may be inferred via the community composition data (*e.g.*, abundance of taxa indicative of organic enrichment).

The biological attributes considered in the BCG are:

1. Historically documented, sensitive, long-lived or regionally endemic taxa
2. Sensitive and rare taxa
3. Sensitive but ubiquitous taxa
4. Taxa of intermediate tolerance
5. Tolerant taxa
6. Non-native taxa
7. Organism condition
8. Ecosystem functions
9. Spatial and temporal extent of detrimental effects
10. Ecosystem connectance

The gradient represented by the BCG has been divided into six levels (tiers) of condition that were defined via a consensus process (Davies and Jackson 2006) using experienced aquatic biologists from across the U.S., including Florida representatives. The six tiers are:

1) Native structural, functional, and taxonomic integrity is preserved; ecosystem function is preserved within range of natural variability;
2) Virtually all native taxa are maintained with some changes in biomass and/or abundance; ecosystem functions are fully maintained within range of natural variability;
3) Some changes in structure due to loss of some rare native taxa; shifts in relative abundance of taxa but sensitive–ubiquitous taxa are common and abundant; ecosystem functions are fully maintained through redundant attributes of the system;
4) Moderate changes in structure due to replacement of some sensitive–ubiquitous taxa by more tolerant taxa, but reproducing populations of some sensitive taxa are maintained; overall balanced distribution of all expected major groups; ecosystem functions largely maintained through redundant attributes;
5) Sensitive taxa are markedly diminished; conspicuously unbalanced distribution of major groups from the expected; organism condition shows signs of physiological stress; system function shows reduced complexity and redundancy; increased buildup or export of unused materials; and
6) Extreme changes in structure; wholesale changes in taxonomic composition; extreme alterations from normal densities and distributions; organism condition is often poor; ecosystem functions are severely altered.

The six levels described above are used to correlate biological index scores with biological condition, as part of calibrating the index. Once the correlation is well established, a determination is made as to which biological condition represents attainment of the CWA goal according to paragraph 101(a)(2) related to aquatic life use support, "protection and propagation of fish, shellfish, and wildlife."

During the development of the BCG model at National BCG Workshops, each of the break-out groups independently reported that the ecological characteristics conceptually described by tiers 1–4 corresponded to how they interpret attainment of the CWA's interim goal for protection and propagation of aquatic life (Davies and Jackson 2006). Additionally, the State of Maine has adopted a policy that aquatic communities conceptually aligned with BCG Category 4 meets the CWA's interim goal for protection and propagation of aquatic life, and this was subsequently approved by EPA.

DEP conducted a BCG exercise to calibrate scores for the SCI. Twenty-two experts examined taxa lists from 30 stream sites throughout Florida, 10 in each Ecoregion that spanned the range of SCI scores. Without any knowledge of the SCI scores, they reviewed the data and assigned each macroinvertebrate community a BCG score from 1 to 6, where 1 represents natural or native condition and 6 represents a condition severely altered in structure and function from a natural condition. Experts independently assigned a BCG score to each site, and then were able to discuss their scores and rationale, and could opt to change their scores based on arguments from other participants. At the conclusion of the workshop, DEP performed a regression analysis on the mean BCG score given to each stream against the SCI score for that site (**Figure 14**).

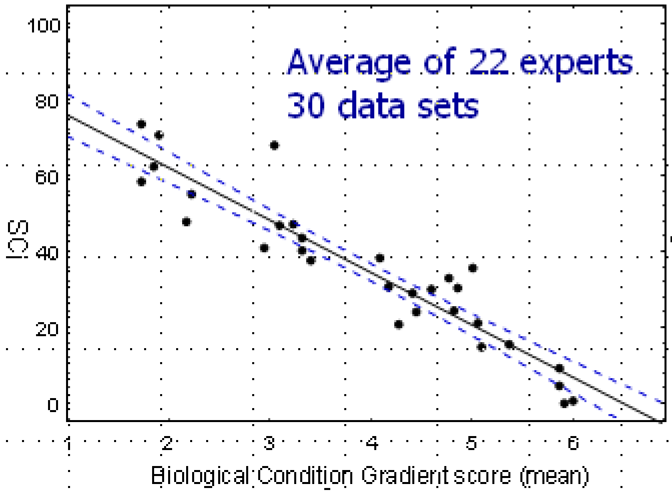

Figure 14. Regression line with 90% confidence interval showing the relationship between the mean BCG score and SCI score.

The experts also were asked to identify the lowest BCG level that still provided for the propagation and maintenance a healthy, well-balanced aquatic community (the interim goal of the Clean Water Act) and the BCG category (and higher) that represented exceptional conditions (the ultimate goal of the Clean Water Act, also referred to as "biological integrity"). Eleven of 22 participants thought SCI scores associated with category 5 should be the lowest acceptable level, while nine participants thought category 4 represented an impaired ecological condition and two experts thought that category 4 was the lowest acceptable condition. Therefore, combined, the majority identified BCG category 4 as an acceptable biological condition, representing a healthy, well balanced community, which corresponds to an SCI score of 34 (**Figure 14**). All 22 participants thought category 2 BCG scores should be considered exceptional, which corresponds to an SCI score of 64.

EPA noted the variability in the expert responses within each BCG category, and conducted an additional analysis of the BCG results to further define an acceptable aquatic life use threshold. EPA calculated a proportional odds logistic regression model (Guisan and Harrell, 2000) to better describe the relationship between a continuous variable (SCI scores) and a categorical variable (BCG categories). This model is based on the cumulative probability of a site being assigned to a given tier (*e.g.,* Tier 3) or to any higher quality tier (Tiers 1 and 2). Thus, five parallel models are fit, modeling the probability of assignment to Tiers 5 to 1, Tiers 4 to 1, Tiers 3 to 1, Tiers 2 to 1, and Tier 1 only. Once these five models are fit, the probability of assignment to any single tier can be extracted from the model results.

In **Figure 15**, the mean predictions of the proportional odds logistic regression models are plotted as solid lines. Lines are color-coded and labeled by different tiers, and each line can be interpreted as the proportion of experts who assigned samples with the indicated SCI value to a particular tier. For example, approximately 90% of experts assigned a sample with the lowest SCI score to Tier 6 (brown line), while the remaining 10% of experts assigned the sample to Tier 5 (purple line). In the figure, the solid circles represent the actual expert assignments recorded from the workshop for each SCI value. The size of the circle is proportional to the number of experts who assigned a sample to a particular tier, and the circles are color-coded by tier. There is some variability among experts in their assignment of BCG scores, but there is a clear central tendency at any given SCI score.

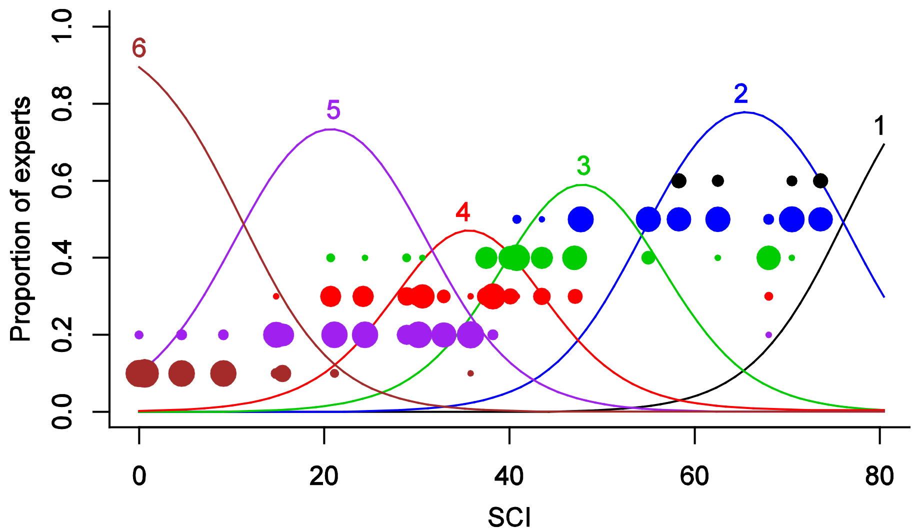

Figure 15. BCG tier assignments modeled with a proportional odds logistic regression.

EPA recommended that the threshold be set at an SCI score where there is an approximately equally low probability of assignment to Tier 5 (*i.e.*, impaired) and a low probability of assignment to Tier 2 (*i.e.*, reference conditions). The resultant threshold of 42 balances the probability of mistakenly assessing a degraded site as meeting aquatic life use goals with the probability of mistakenly assessing a reference site as impaired. This score is consistent with the impairment threshold of 40 as determined by the reference site approach.

2.4.3 Setting and Evaluating a SCI Impairment Threshold

Weighing these multiple lines of evidence, the DEP has determined that a SCI score of 40 indicates that the designated use is being met, and a score of 39 is impaired. This minimum threshold for aquatic life use support is supported by the distribution of benchmark site scores and corresponds with a BCG category midway between Tiers 3 and 4. The proportional odds analysis provides assurance that macroinvertebrate communities deemed exceptional (BCG category 2) will not be considered impaired at a threshold of 40. The DEP evaluated recent data for the individual metrics of the SCI to determine what range of macroinvertebrate attributes would be considered healthy using this impairment threshold. Since DEP conducted the SCI calibration in 2007, the state has collected approximately 700 additional SCI samples from a variety of sites, including minimally disturbed reference sites (for nutrient criteria development), sites located along a nutrient gradient, and randomly chosen sites for the status and trends network. Based upon the relationship described in **Figure 16**, the SCI values from this data set were subdivided into increments representing half-step BCG Categories, and the individual metrics associated with each half-step interval were averaged. The metric data bracketing BCG category 2 were averaged to demonstrate metric values associated with exceptional conditions. Data within the range of the minimum threshold of aquatic life use support of 40 were also averaged to provide an example of the stream condition that Florida's SCI biological criterion will protect (**Table 2**). Note that although there are moderate differences between metrics associated with exceptional biological communities and those near the range of the minimum threshold for aquatic life use support, the attributes associated with communities near the threshold are still considered to be indicative of healthy, well-balanced communities by the majority of the Florida stream experts who participated in the BCG exercise.

Table 2. Average values for metrics at an SCI score equivalent to a Biological Condition Gradient of Category 2, and average values for metrics near the minimum SCI score for aquatic life use support. Data was based upon the DEP's data collection effort since 2007 (total N = 696 SCI samples).

SCI Metric	Metric Average at BCG 2 (Exceptional)	Metric Average At Minimum Aquatic Life Use Support Threshold
Number of Total Taxa	32.0	28.7
Number of Clinger Taxa	5.6	3.3
Number of Long Lived Taxa	1.5	1.1
Percent Suspension Feeders and Filterers	22.0	15.8
Number of Sensitive Taxa	5.4	2.7
Percent Tanytarsini	13.3	9.5
Percent Very Tolerant	6.5	14.3
Number of Ephemeroptera Taxa	3.5	2.3
Number of Trichoptera Taxa	4.5	2.6
Percent Dominant	22.6	26.2
Number of Sites in Average	134	64

During the development of the Stream Condition Index, the DEP observed a strong relationship between the SCI and the Human Disturbance Gradient (Figure 16). Note the highest range of actual SCI scores were observed in the two groups of lowest human disturbance gradient sites (2 boxes at left in **Figure 16**). This wide range needs to be considered when establishing the threshold to limit the probability of falsely identifying unimpacted sites as not attaining an aquatic life use. However, the range of scores in the higher human disturbance gradient sites (expected to result in a BCG category 5-6) are low. Therefore, the risk is low (virtually non-existent for the SCI) in applying the biological assessment tool and falsely identifying impacted sites as attaining an aquatic life use.

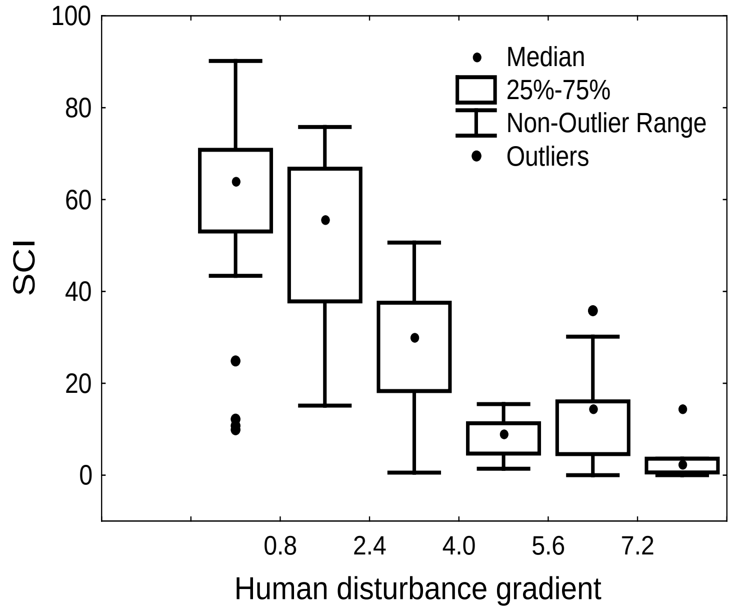

Figure 16. Relationship between the SCI (2004 data) and the Human Disturbance Gradient (from Fore *et al.*, 2007a).

This variability of the SCI scores within a given range of the human disturbance gradient is generally caused by changes in biological community relative to natural occurrences (droughts, floods, etc.), as well as the inherent limitation of the biological assessment methods.

Biological field observations can be influenced by natural conditions that may have occurred prior to the sampling event. Changes in hydrology, particularly high and low flow events that result in differential water velocities and habitat availability, will affect the biological community in a stream, potentially resulting in lower scores. The variability in low human disturbance gradient sites also reflects the fact that the biological communities in these systems are able to rapidly recover because the habitat and health of the stream is conducive to recovery. In high HDG sites, natural hydrologic events (along with human disturbance) can affect the biology, but any recovery is slow and limited due to the human disturbance impacts and lack of recruitment of organisms from surrounding areas. Therefore, in high human disturbance gradient sites, SCI scores always tend to be low, and the range of values remains small.

The other factor leading to higher variability in scores for low disturbance sites relates to sampling issues. DEP's SCI collection methods follow EPA rapid bioassessment guidance, but do not result in a complete census of all taxa present at a site. Instead, they provide a practical level of effort that can be used to distinguish healthy from impaired sites. Therefore, the sampling method is inherently conducted in a manner that may result in a high range of results where taxa are present and a low range of results where taxa are diminished. In other words, when taking a sample, it is possible to *fail to capture taxa that exist* in the water body, but it is not possible to *capture taxa that do not exist* in the water body.

In statistical terms, undisturbed sites have a higher probability of Type I error (falsely concluding that the site was impaired). Because the variability in the SCIs decreases as human disturbance increases, the disturbed sites fundamentally are subject to much lower occurrence rate of a Type II error (falsely concluding that the site was unimpaired) when compared to undisturbed sites. From a theoretical standpoint, since the error of the method used to collect representative taxa can only fail to capture and count taxa, and only 2 of the 10 metrics result in an improved SCI when specific organisms are missed, it is likely that Type I errors are of greater concern (occur more frequently) with this methodology.

2.4.4 Additional Analysis of Rigorously Verified Benchmark Site SCI Data

SCI scores from an early version of DEP's field-verified nutrient benchmark site dataset were also evaluated to determine the range and variability of biological condition found in Florida's minimally-disturbed sites. "Theoretically," these sites would be expected to have a SCI score reflective of a BCG category 2. In reality, for the reasons previously stated, there is more variability in the actual scores. This benchmark dataset consists of sites determined by experienced DEP scientists to be influenced by only very low levels of anthropogenic stressors. Additional selection criteria included a Landscape Development Intensity index score of ≤2, absence of upstream point source discharges, examination of aerial photographs, direct observations of watershed land use and hydrologic conditions during site visits, and habitat assessment. The dataset included 69 sampling events at a total of 53 stations across the state (16 stations were sampled twice during the verification process).

The mean SCI score from all 69 sampling events was 65.1, and the median was 65. The standard deviation from the mean was 15.8, and the range of scores was 80, spanning from 100 to 20. The one nutrient benchmark site that scored below the impairment threshold of 40 occurred at a Steinhatchee River site (at CR 357), which scored 20 on the SCI on August 12, 2008, after an extended period of low flow conditions (see **Figure 17**). However, when this site was subsequently re-sampled on January 14, 2009 (after a period of higher flows), it scored a 53 (see **Figure 18 and 19**). Note that another minimally disturbed Steinhatchee River site located approximately 8 miles downstream with slightly more flow (at Canal Road), scored 41 and 62 on the SCI during the same time period. Based on direct observations, the flow regime was the dominant factor for the variability in the SCI scores. DEP SOPs provide clear guidance regarding appropriate conditions during which to sample, including a minimum velocity of 0.05 m/sec. Although the Steinhatchee at CR 357 achieved this velocity and was not dry prior to sampling, the sluggish flows and less than optimal inundated habitat were responsible for the low SCI scores, not any human disturbance (the upstream basin is almost 100% forested). This is an example of the type of hydrologic conditions that occur randomly throughout the state, prompting DEP, in an attempt to minimize Type I errors, to select the lower 2.5% distribution of reference sites as the impairment threshold.

Figure 17. Steinhatchee River at CR 357 under low water level conditions, August 2008.

Figure 18. Steinhatchee River at CR 357 under more optimal water level conditions, January 2009.

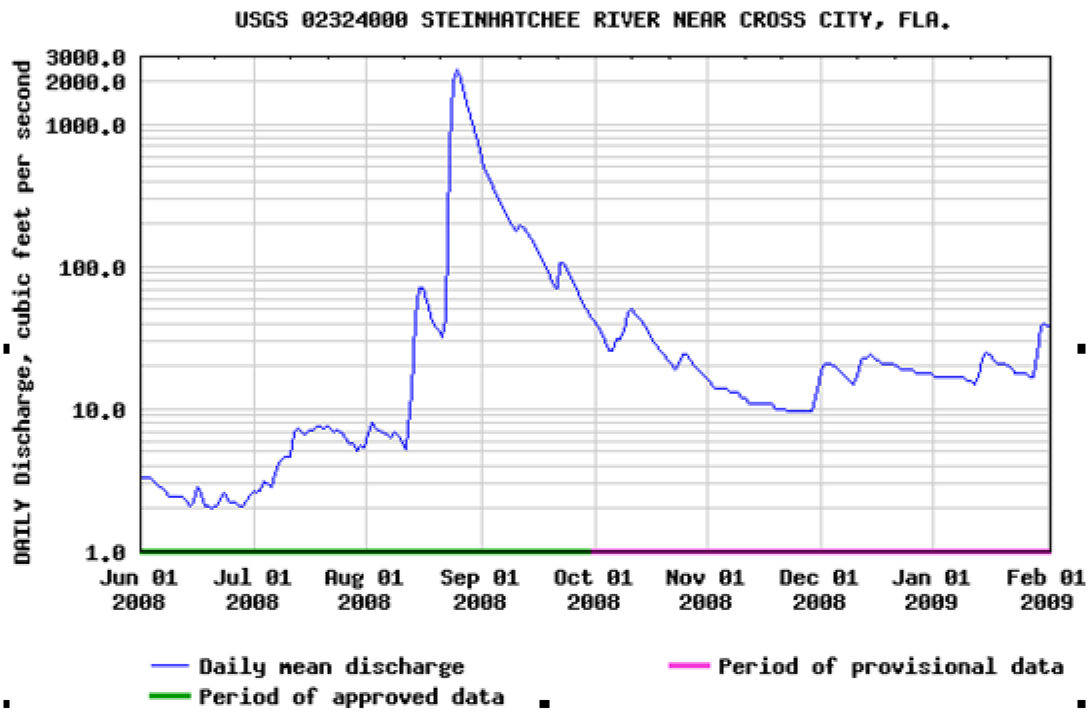

Figure 19. USGS hydrograph for the Steinhatchee River during the period of the two sampling events. The mean discharge rate for the Steinhatchee River near Cross City was 7.4 ft^3/sec on 8/12/2008 and 23 ft^3/sec on 1/14/2009.

2.4.5 Evaluation of Nutrient Benchmark Site Replicate Data: SCI

The 16 nutrient benchmark sites with replicate data were analyzed to determine the variability that can occur in SCI scores at the same sampling location within two years of each other. The benchmark sites with replicate data are shown below in **Table 3**. The mean difference in SCI scores from this sub-dataset was 17.1, with a standard deviation of 13.3. The median difference was 18. The largest difference in scores occurred at the St. Marys River at SR 2, which received SCI scores of 50 in June 2008, and 100 in November 2008.

Table 3. Minimally disturbed stream benchmark sites with replicate SCI data.

Benchmark Site	Date sampled	SCI score	Difference between replicates
Blackwater River at Highway 4	3/26/2007	56	14
	7/9/2008	70	
Cypress Branch	11/3/2008	66	3
	12/16/2008	63	
Escambia River at Highway 4	9/19/2007	57	6
	7/10/2008	51	
Manatee River at 64	5/16/2007	81	17
	12/17/2008	64	

Orange Creek upstream of Highway 21	2/26/2007	74	8
	5/1/2008	82	
Peters Creek at CR 315	5/28/2008	92	19
	10/28/2008	73	
Sopchoppy River	6/19/2008	41	23
	11/13/2008	64	
Steinhatchee River at CR 357	8/12/2008	20	33
	1/14/2009	53	
Steinhatchee River at Canal Road	8/12/2008	41	21
	1/14/2009	62	
St. Marys River at SR 2	6/18/2008	50	50
	11/12/2008	100	
Telogia Creek at CR 1641	6/10/2008	78	20
	11/20/2008	58	
Suwannee River at CR 6	10/10/2006	53	2
	12/12/2007	51	
Withlacoochee River above River Dr.	5/7/2008	44	2
	10/8/2008	42	
Withlacoochee River at Stokes Ferry	2/20/2007	68	21
	11/7/2007	47	
Yellow River at Hwy 2	5/15/2007	54	25
	7/9/2008	79	
Yon Creek at SR 12	6/13/2008	81	7
	11/20/2008	74	

Differences in SCI scores between replicates can be caused by the natural variability of environmental factors such as recent hydrologic conditions resulting in changes in habitat availability, as well variability associated with laboratory sub-sampling. Based on field observations, it was natural factors (water level and flow), not changes in human disturbance, that were the main drivers of the differences in SCI scores between replicates taken at different times. Note that sampling visits to the sites with duplicate data were not separated by more than fourteen months (most were sampled less than six months apart).

Another indication that human disturbance was not associated with this variability was that no correlation was found between Landscape Development Intensity Index score and SCI score within the entire benchmark site dataset (**Figure 20**). This is in contrast to the strong relationship between the LDI and SCI scores across the entire range of human disturbance (in **Figure 20**, the LDI exerts a prominent influence on the HDG).

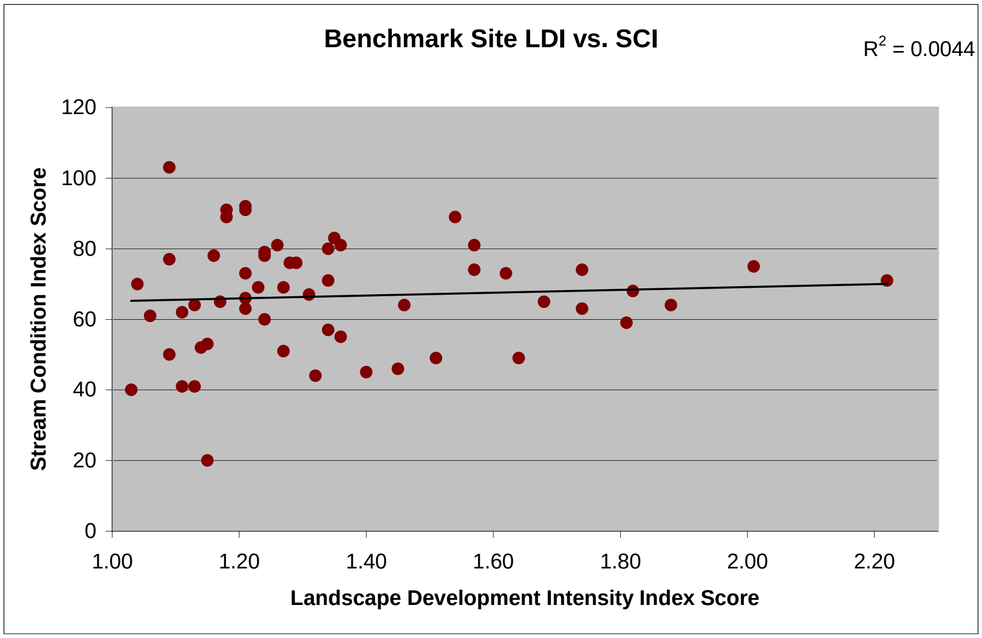

Figure 20. Minimally disturbed benchmark sites plotted against the Landscape Development Intensity Index (LDI). Direct observations indicated that the LDI reflected current land use and disturbance conditions.

2.5 SCI Training Materials, Training Requirements, and Checklists

See: http://www.dep.state.fl.us/water/bioassess/training.htm

3 Lake Vegetation Index

3.1 Development of the Lake Vegetation Index using the Human Disturbance Gradient

Similar to the development of the SCI, DEP used a Human Disturbance Gradient (HDG) approach to create the Lake Vegetation Index (LVI), a multimetric index of lake plant community biological integrity. The HDG employed in the LVI development included the following four components, as described in Fore *et al.* (2007b):

- Landscape Development Intensity Index (LDI)
- Lake Habitat Assessment (DEP SOP FT 3200)
- Water Quality Index (based on relative values for conductivity, total Kjeldahl nitrogen [TKN], nitrites/nitrates [NO_x], total phosphorus [TP], and algal growth potential [AGP])
- Lake Hydrological Connectivity (DEP SOP FT 3200)

The LVI was originally developed in 2005 and recalibrated in 2007 (Fore *et al.*, 2007b). Final component metrics that were the most highly correlated with the HDG were percent native taxa, percent invasive exotic taxa as determined by Florida Exotic Pest Plant Council (FLEPPC), the coefficient of conservatism (C of C) of dominant/codominant taxa, and percent sensitive taxa.

3.1.1 Percent Native Taxa

Native taxa are those whose natural range includes Florida. One of the results of human disturbance is the introduction of non-native taxa either purposefully or accidentally. The disturbance of native habitats can also allow non-native taxa to colonize by their own mechanisms. Southern portions of Florida are especially vulnerable to the introduction of non-native taxa due to their tropical climate. Presence and abundance of native and non-native taxa are common components of plant biological assessment tools (*e.g.,* Ervin *et al.* 2006, Fennessey *et al.* 2002, Mack 2004, Miller *et al.* 2006, Rocchio 2006).

3.1.2 Percent Invasive Exotic Taxa as determined by FLEPPC

The Florida Exotic Pest Plant Council (FLEPPC) is a non-profit, non-regulatory board of Florida scientists whose mission is to support the management of invasive exotic plants in Florida's natural areas by providing a forum for the exchange of scientific, educational, and technical information (http://www.fleppc.org). Every two years, the FLEPPC board releases a list of the most invasive exotic plants occurring in Florida. FLEPPC defines Category I invasives as "exotics that are altering native plant communities by displacing native species, changing community structures or ecological functions, or hybridizing with natives" (FLEPPC 2009). Category II invasives are defined as "exotics that have increased in abundance or frequency but have not yet altered Florida plant communities to the extent shown by Category I species."

3.1.3 Coefficient of Conservatism (C of C) of Dominant or Codominant Taxa

The coefficient of conservatism (C of C) concept is based on two factors: 1) plants have varying degrees of fidelity to specific habitats and their quality, and 2) plants have varying tolerances to disturbances and respond in varying degrees. Each plant is assigned a C of C value by a group of local botanists in a region, each of whom has extensive experience with the plants in the field. A C of C value is assigned to species on a scale of 0-10, and is based on its fidelity to a habitat relatively unaltered from what is believed to be pre-European settlement condition. A C of C of 0, therefore, is given to species that would indicate the most disturbed or ruderal conditions. A value of 10 is given to species that demonstrate a high degree of fidelity to a specific set of ecological conditions; in other words, these species would provide the highest level of confidence that the places in which they are found are the most intact remnant natural systems. The C of C concept has been used by others for various ecosystems and locations to assess the biological integrity of plant communities (Alix and Scribailo 1998, Ervin *et al.* 2006, Lopez and Fennessy 2002, Miller *et al.* 2006, Nichols *et al.* 2000, Mortellaro *et al.* 2009, Wilhelm and Masters 1995).

3.1.4 Percent Sensitive Taxa

We defined sensitive taxa as those with a C of C score greater than or equal to seven. That range of scores includes taxa that are typical of well-established communities which have sustained only minor disturbances, and taxa that exhibit high degrees of fidelity to a narrow set of ecological conditions (Andreas 1995). A plant index of biological integrity in the Midwestern United States uses C of C scores from eight to ten to indicate sensitive taxa (Rothrock *et al.* 2008).

3.2 2011 LVI Metrics Adjustments

3.2.1 Background

As described above, the LVI was originally developed in 2005 and recalibrated with additional data in 2007. The major community attributes reflected in the component metrics are the percent native taxa, the percent of taxa determined to be invasive exotic pests by the Florida Exotic Pest Plant Council (FLEPPC), and the coefficient of conservatism (C of C) of plant taxa, defined as a taxon's level of fidelity to a defined ecological niche. Since 2005, our scientific understanding of these metrics has changed in the following ways:

1. Every two years, FLEPPC updates two lists of exotic plant taxa. Category I taxa are considered to be the most problematic invasive exotic plants while Category II taxa are considered to be not as disruptive to ecosystems as Category I plants. Since the development of the LVI, several common taxa have been added to the FLEPPC list, causing a potential upward shift in the distribution of the % FLEPPC taxa metric. Also, as more taxa are added to the FLEPPC list, the correlation between the % FLEPPC and % Native taxa becomes higher, meaning the metrics are more redundant.
2. The C of C scores used for the LVI were established by panels of botanists, in six iterations, from 1999 to 2006. The first five of these scoring efforts were done for wetland bioassessment development. In February 2011, DEP held a workshop of eight botanists to review C of C scores and their relevance to use in lakes, and they recommended changing scores for 134 taxa based on their experience in seeing these taxa in various habitats and at various disturbance levels throughout the state. This workshop employed a consensus approach to revising these scores, and evaluated the scores on a holistic level from the perspective of their use in the LVI. The panel recommended 107 decreases in C of C scores and only 27 increases, so these changes would cause a downward shift in the distribution of the percent Sensitive and Dominant C of C metrics.

To make the LVI as scientifically relevant as possible (reflecting the collective knowledge of the FLEPPC panel and the participants in the C of C scoring workshop), DEP has made the following three adjustments to the LVI. These adjustments are further described in sections that follow.

1. Inclusion of the revised C of C scores that were established by the February 2011 C of C workshop participants, to reflect the current state of knowledge of Florida aquatic and wetland plants;

2. Modification of the % FLEPPC invasive exotic metric to include only FLEPPC Category I taxa, to reduce the redundancy of this metric with the % Native metric while highlighting those invasive exotic taxa that are most seriously displacing natives; and
3. Regionalization of the scoring for the % Native metric to recognize the observed differences in minimally disturbed lakes in south vs. north Florida.
4. To account for shifts in the distribution of metric data caused by these changes, we have redefined the metric scoring rules based on a probabilistic distribution of Florida LVI data.

3.2.2 Assessment of Effects of C of C Score Changes

To assess the effects of modified C of C scores on LVI results, the full dataset of LVI 2008 data (data collected since 2007, calculated using the methods of the 2007 recalibration effort) was examined. That dataset contained 629 samples from 511 stations (373 in the North LVI region, 176 in the South LVI region). The analysis showed that the modified C of C scores would cause either no change or a change of up to 2 points in 75% of samples, an increase of 3 or more points in 4% of samples, and a decrease of 3 or more points in 21% of samples (**Table 4**).

Table 4. Changes in the final LVI scores for 629 LVI2008 samples due solely to the modification of C of C scores established by the 2011 expert panel.

Change in Score due to CC	6	5	4	3	2	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
# Samples	2	4	8	13	19	49	171	137	93	65	28	24	9	4	1	2
% Samples	0%	1%	1%	2%	3%	8%	27%	22%	15%	10%	4%	4%	1%	1%	0%	0%

To evaluate whether the modified C of C scores would cause changes in the correlation of the LVI with human disturbance, the correlation between the Landscape Development Intensity index and LVI score was calculated both with original and modified C of C scores. These comparisons were made both on the sample level and station level (using station average LVI scores). This analysis showed that the correlation coefficient between the LVI and LDI was essentially equivalent (reduced by only a very small amount) when using the revised C of C scores (**Tables 5 and 6**).

Table 5. Sample level correlation between LVI and LDI for full dataset (629 samples) to determine the influence of revised CC scores.

	LDI	*OrigCC_LVI2008*	*NewCC_LVI2008*
LDI	1		
OrigCC_LVI2008	-0.48036	1	
NewCC_LVI2008	-0.45047	0.980159	1

Table 6. Station level correlation between LVI and LDI for full dataset (511 stations) to determine the influence of revised CC scores.

	LDI	*OrigCC_LVI2008-Station Average*	*NewCC_LVI2008-Station Average*
LDI	1		
OrigCC_LVI2008-Station Average	-0.47681	1	
NewCC_LVI2008-Station Average	-0.44502	0.980322	1

The C of C scores contribute information to two of the four LVI metrics, including the percent Sensitive taxa and the C of C of the dominant or co-dominant taxa. The modification of C of C scores shifted the distribution of the raw metric information for those metrics; for example, the median C of C of dominant/codominant dropped from 5.07 to 4.66 (**Table 7**). The scoring of these metrics should be adjusted to account for this change. Note that analyses of metric scores were conducted on lake sections, where each LVI sample contains four lake sections.

Table 7. Relative differences in the data distribution when the dominant C of C and percent sensitive taxa are calculated using the old and new (2011) C of C values (n=2516 lake sections).

Percentiles of section level data	**CofC_Dom_Codom**		**% Sensitive**	
	OldCC	**NewCC**	**OldCC**	**NewCC**
5th	0	0	0	0
25th	2.62	2.06	4	3.45
50th	5.07	4.66	9.52	8.33
75th	6.36	5.82	16.67	13.38
95th	7.28	7.21	30	25

3.2.3 Justification for Changing the FLEPPC metric and Regionalization of the % Native metric

The presence and dominance of invasive exotic plants is an important element of plant community integrity, reflected in several metrics of the LVI. The percent FLEPPC metric accounts for how the coverage of noxious, invasive exotic plants results in adverse changes to the structure (and often function) of the plant community, while the percent Native metric considers the nativity status of resident taxa, without regard to whether or not the non-native are invasive. In a correlation analysis of 227 probabilistically selected lakes, sampled between 2008-2010, representing 908 lake sections, the percent Native taxa in each lake section was correlated with the percent FLEPPC metric (r = 0.95), and with the percent FLEPPC Category I proposed metric (r = 0.88). Although both correlations are high, use of the FLEPPC Category I would provide for a metric that is slightly less redundant with percent Native than a metric that includes all FLEPPC taxa.

The LVI South region has a warmer and more tropical climate than the North, with very few freezes, and consequently, has more non-native plant taxa present at a low disturbance level. DEP believes that the percent FLEPPC (Category I or II) metric should not be regionally scored because these taxa are associated with major negative effects on the plant community. However, because a higher percentage of non-native taxa are found in southern lakes than northern lakes, regardless of disturbance, a regional adjustment of the percent Native metric would be appropriate.

3.2.4 Description of 2011 Calculations and Metric Scoring

The following describes the 2011 metric calculations and scoring process:

1. Include the C of C Scores as revised by the 2011 expert panel;
2. Use the FLEPPC Category I (only) instead of including both FLEPPC categories; and
3. Scale the percent Sensitive, C of C Dominant/Co-dominant, and percent Native metrics by region.

Note that regional scaling of percent Sensitive and C of C Dominant/Co-dominant was part of the 2007 calibration. Regionalizing the scoring for the percent Native metric acknowledges that differences in the expected Native taxa between north and south Florida are due to climate (lack of freezes) as well as human disturbance.

LVI data from small and large lakes, probabilistically sampled by the Integrated Water Resource Monitoring (IWRM) program from 2008-2010, were used to develop the scoring criteria for the 2011 LVI metrics. This data set contained 227 samples (149 North, 78 South). Probabilistically-derived data are ideal for this application because they include the range of conditions expected for Florida lakes, and thus, provide an unbiased distribution of data from which to calculate metric scores.

In general, higher LDI scores were more likely to be encountered in the South than in the North region (i.e., a larger proportion of sites in the South have more disturbance compared to the North). To avoid inappropriately inflating scores for southern lakes, we established metric scaling to represent approximately equal LDI levels for North and South. Consequently, the lower 20th percentile of data in the South was used to derive the regionally-scaled scores for metrics that decrease in response to human disturbance (**Table 8**).

Table 8. Revised metric scoring rules for the LVI, derived from the 5th and 95th percentiles (for statewide and the North region) or 20th and 95th percentiles (for the South region). The values were from 909 lake sections (227 samples) from DEP's IWRM probabilistic sampling program.

	LVI Region	**N (lake sections)**	**5th %ile**	**95th %ile**	**Scoring Rule**
% Sensitive Taxa	North	596	0	27.78	x/27.78
	South	312	0*	20.00	x/20
C of C Dominant Taxa	North	539	0	7.91	x/7.91
	South	283	0*	7.00	x/7
% Native Taxa	North	596	62.50	100	(x-62.5)/37.5
	South	312	66.67*	92.56	(x-66.67)/25.89
% FLEPPC1 Taxa	Both	908	0	30.00	1-(x/30)

*These represent the 20th percentile of data for the South region.

Once the new 2011 LVI scores were calculated as described above, correlation analyses were conducted to determine how the component metrics compared with each other, to the total LVI score, and with the LDI (**Table 9**).

Table 9. Results of correlation analyses between the LDI and LVI component metric information for 2008-2010 IWRM dataset (227 samples, 908 lake sections), shown separately for the entire dataset, north LVI region, and south LVI region. All correlation coefficients shown are from pairwise correlations and significant at P<0.001.

Statewide (908 lake sections)	LDI	CofC_Dom_Codom	% FLEPPC1	% Native	% Sensitive	LVI Score
LDI	1					
CofC_Dom_Codom	-0.30	1				
% FLEPPC1	0.39	-0.34	1			
% Native	-0.48	0.42	-0.89	1		
% Sensitive	-0.27	0.27	-0.19	0.23	1	
LVI Score	-0.46	0.72	-0.74	0.81	0.57	1

North Region (596 lake sections)	LDI	CofC_Dom_Codom	% FLEPPC1	% Native	% Sensitive	LVI Score
LDI	1					
CofC_Dom_Codom	-0.27	1				
% FLEPPC1	0.43	-0.38	1			
% Native	-0.48	0.44	-0.90	1		
% Sensitive	-0.21	0.27	-0.18	0.23	1	
LVI Score	-0.46	0.72	-0.77	0.83	0.56	1

South Region (312 lake sections)	LDI	CofC_Dom_Codom	% FLEPPC1	% Native	% Sensitive	LVI Score
LDI	1					
CofC_Dom_Codom	-0.27	1				
% FLEPPC1	0.19	-0.26	1			
% Native	-0.36	0.30	-0.82	1		
% Sensitive	-0.27	0.19	*	*	1	
LVI Score	-0.35	0.68	-0.64	0.72	0.51	1

*correlations were not significant (P > 0.001)

To evaluate how the 2011 scoring procedure affects LVI scores, 227 probabilistically-derived lake samples, collected from 2008-2010, were compared using both the old and new calculation methods (**Table 10**). Overall, there was a mean change of -3.7 points and median change of -3.7 points between the two calculation methods, with changes in scores ranging from -21 points to +14 points. The correlation between the LVI and LDI was similar to, but slightly lower, when comparing the new calculation method to the old one. The LVI scores calculated using the old vs. new methods were highly correlated, with a correlation coefficient of 0.97 (see **Figure 21**).

Table 10. Comparison of 227 probabilistically-derived lake samples, collected from 2008-2010, using both the old and new LVI calculation methods.

	LDI	*Revised LVI*	*Current LVI*
LDI	1		
Revised LVI	-0.50963	1	
Current LVI	-0.55351	0.96757	1

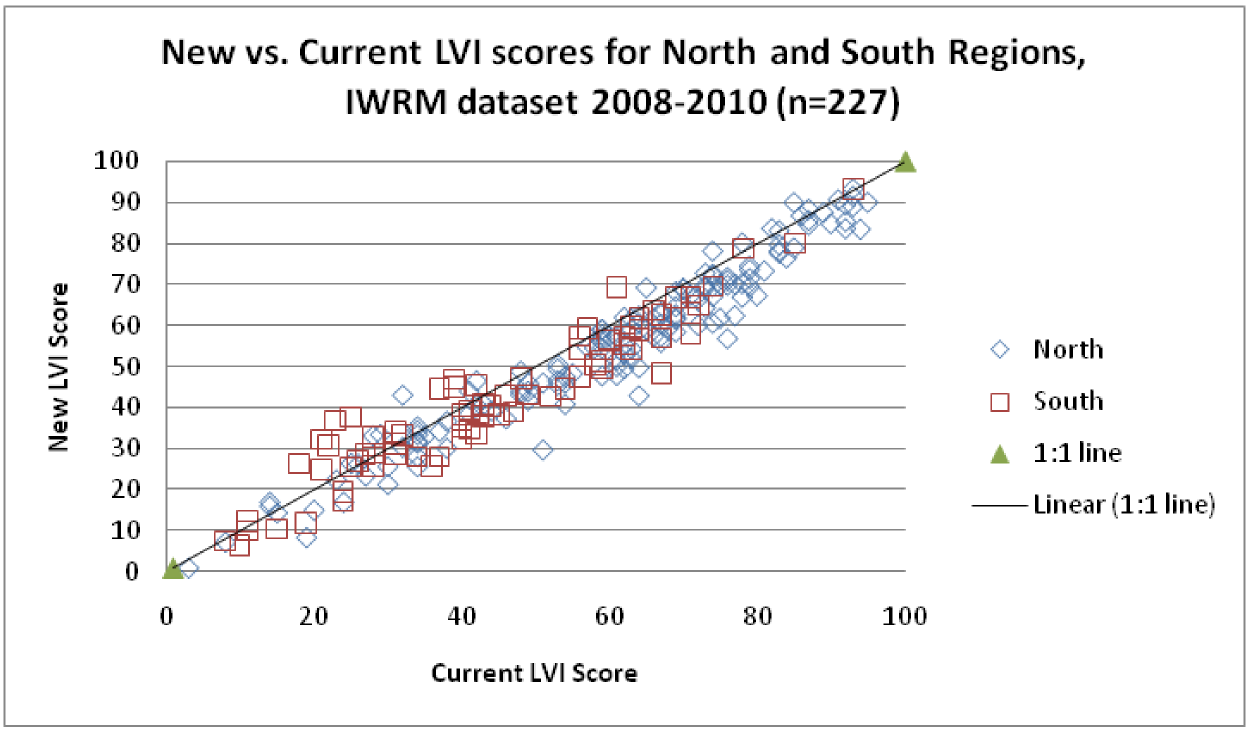

Figure 21. Current LVI score, calculated per the 2007 recalibration, vs. the new LVI score calculated as described in this document, for 227 samples collected from 2008-2010 as part of the IWRM probabilistic monitoring program.

3.3 Establishing Expectations for Aquatic Life Use – Lake Vegetation Index:

In 2007, DEP calibrated the LVI using primarily the Biological Condition Gradient approach. Subsequent EPA review resulted in the recommendation that Florida use an examination of the lower distribution of reference sites as the principal line of evidence for establishing aquatic life use support thresholds, in combination with the Biological Condition Gradient approach. DEP initially proposed a LVI threshold in 2009, and has repeated those analyses with the new calculations described in section 3.2 to evaluate that threshold.

3.3.1 LVI Benchmark Site Approach

2009 Analysis

DEP evaluated data from existing sites to identify benchmark lakes that could be used to determine the appropriate threshold for the LVI. To be considered benchmark, the watershed-scale landscape development intensity (LDI) index score had to be less than 3, and the LDI of the 100-m buffer zone around the lake had to be less than 2. DEP biologists also examined aerial photos and conducted an onsite watershed survey to ensure that there were no adverse human influences not detected by the LDI, and performed a whole-lake habitat assessment. Candidate benchmark lakes were excluded if they had a history of adverse human activity (*e.g.,* aquatic plant control, artificial fertilization) or current human activity (*e.g.,* adjacent citrus groves). Thirty benchmark lakes were identified and used in this analysis (**Table 13**).

DEP conducted statistical interval and equivalence tests with LVI data from these 30 reference lakes to determine the lower bounds of the reference site distribution. As was described for the SCI, the intent was to identify a threshold for the LVI that balanced Type I (falsely concluding that the site was impaired) and Type II (failing to detect that a site is truly impaired) errors. The analysis of the most recent LVIs at all 30 sites showed that the 2.5^{th} percentile of all reference data was statistically in the range of 33-48 points, while the analysis of the two most recent visits at 15 lakes showed that the 2.5^{th} percentile of reference data, when considering the mean of two visits, was in the range of 31-53 points (**Table 11**). The middle of this range was 46 points, representing a minimum aquatic life use support threshold that balances Type I and Type II errors. In the proposed water quality threshold for the LVI, impairment will be determined by two site visits, so the threshold of 46 is closely aligned with the assessment methods. A minimum aquatic life use support threshold of 46 would limit the percentage of reference sites that would falsely be deemed impaired to 2.5%.

Table 11. Results of interval and equivalence tests conducted in 2009 on reference sites with 2 LVI results. Shown are site mean, impairment threshold, and range for threshold values defined at the 2.5^{th} and 5^{th} percentile of reference sites ($p < 0.05$; N = 15 reference sites with two LVI values for each site).

Impairment threshold (description)	Minimum Aquatic Life Use Support Threshold (average of two results)	Not Reference Condition (95% Confidence)	Undetermined	Reference Condition (95% Confidence)
2.5^{th} percentile of reference	46	<31	31–53	>53
5^{th} percentile of reference	50	<37	37–57	>57

2011 Analysis

In 2011, LVI scores for the two most recent samples from the 30 benchmark lakes were recalculated using the procedure described above. Additional data were available for some sites, and increased the number of lakes with two or more samples from 15 to 20 lakes. It is the mean of the two most recent samples that we are most interested in because no impairment decisions would be based on only one sample. With the 2011 LVI adjustments, the 2.5th percentile of the reference site distribution shifted from 46.47 to 43.27 (**Table 12**), meaning that a score of **43** is equivalent to the former minimum acceptable threshold for aquatic life use support. Therefore, a value of **43** is the new lowest acceptable LVI score. In **Figure 21**, it is apparent that the new LVI calculation method slightly reduces the score, making it important to adopt the revised aquatic life use support threshold of **43** to prevent an increase in Type I errors (see **Table 13**).

Table 12. Comparison of the two most recent samples from the 30 reference lakes DEP used for Numeric Nutrient Criteria (NNC) development, using the 2007 and 2011 scoring procedures.

	Original Distribution (15 lakes)	New Distribution with new Scoring (20 lakes)
2.5th %ile	**46.47**	**43.27**
5th %ile	**49.36**	**45.73**

Table 13. Minimally disturbed benchmark lake sites with replicate LVI data.

	2009 Dataset		2011 Dataset	
Station	**Date**	**LVI_2008**	**Date**	**LVI_2011**
Blue Cypress Lake	6/14/2007	60.25	6/14/2007	60.20
	10/1/2008	58.75	10/1/2008	52.76
Buck Lake	-	-	6/1/2005	70.81
	-	-	10/14/2009	72.66
Charles Lake	-	-	7/11/2008	77.27
	-	-	5/3/2011	68.33
Dunford Lake	-	-	10/26/2004	80.49
	-	-	10/12/2005	83.95
Gore Lake	9/17/2003	66.11	9/17/2003	58.32
	11/13/2006	61.82	11/13/2006	54.66
Lake Annie	11/3/2005	77.83	11/3/2005	70.40
	10/8/2008	92.75	10/8/2008	96.36
Lake Ashby	6/7/2005	45.14	11/3/2005	39.48
	11/3/2005	42.03	10/26/2010	52.50
Lake Cassidy	-	-	10/6/2003	87.07
	-	-	11/1/2006	92.59
Lake Harney	10/19/2005	36.93	7/23/2008	65.15
	7/23/2008	66.75	6/9/2010	47.08

Lake Norris	10/8/2003	63.03	10/8/2003	64.02
	10/29/2008	73	10/29/2008	71.92
Lake Palestine	10/11/2005	94.25	11/8/2006	89.90
	11/8/2006	90.72	6/10/2010	76.68
Merial Lake	6/14/2005	77.52	6/14/2005	75.12
	10/26/2005	82.74	10/26/2005	89.70
Ocean Pond	10/11/2005	90.44	11/1/2006	84.32
	11/1/2006	86.89	6/10/2010	84.78
Orange Lake	-	-	8/4/2009	37.46
	-	-	10/20/2010	44.16
Otter Lake	10/13/2005	69.73	10/17/2006	70.82
	10/17/2006	72.84	7/29/2009	73.55
Rattlesnake Lake	11/10/2005	82.07	11/10/2005	91.31
	10/31/2006	70.21	10/31/2006	68.39
Russell Lake	9/30/2003	68.9	9/30/2003	50.39
	10/2/2008	74	10/2/2008	62.98
Sellers Lake	10/26/2005	78.53	9/10/2009	86.97
	10/18/2006	81.71	5/20/2010	88.41
Swift Creek Pond	10/11/2005	89.78	7/21/2008	75.36
	7/21/2008	67.75	9/1/2009	78.36
Wildcat Lake	10/25/2005	92.37	10/25/2005	92.93
	10/17/2006	90.1	10/17/2006	84.16

3.3.2 LVI Biological Condition Gradient Approach

2007 and 2009 Analyses

In a process analogous to that for the SCI BCG calibration, 20 Florida plant ecologists, botanists, and field lake managers, all with at least five years of experience, were involved in BCG calibration of the LVI. The experts examined taxa lists from 30 lakes throughout Florida that spanned the range of LVI scores (see Fore *et al.* 2007b for site information and taxa lists). Without any knowledge of the LVI scores, they reviewed the plant data and assigned each plant community a BCG score from 1 to 6, where 1 represents natural or native condition and 6 represents a condition severely altered in structure and function from a natural condition. Experts independently assigned a BCG score to each lake, and then were able to discuss their scores and rationale, and could opt to change their scores based on arguments from other participants. At the conclusion of the workshop, DEP conducted least squares regression analysis on the mean BCG score given to each lake against the LVI score for that lake (**Figure 22**).

The experts were also asked to identify the lowest BCG level that still provided for the propagation and maintenance of a healthy, well-balanced aquatic community (the interim goal of the Clean Water Act) and the BCG category (and higher) representing exceptional conditions (the ultimate goal of the Clean Water Act, also referred to as "biological integrity"). Twelve of 20 participants thought LVI scores associated with category 5 should be impaired, while 5 participants thought category 4 represented an impaired ecological condition (see **Table 14** for summary statistics). Thirteen of 20 participants thought Category 2 LVI scores should be considered exceptional, six were equally divided between Categories 1 and 2, and one expert did not provide an opinion. Although DEP originally proposed that the LVI impairment threshold be established at the BCG line of 4.6 (Fore *et al.* 2007b), DEP decided, in conjunction with EPA, to establish the LVI impairment threshold based primarily on the benchmark distribution. This analysis suggests that the pre-20011 adjustment scores of 45 and below should represent impairment, and scores of 78 and above should represent exceptional.

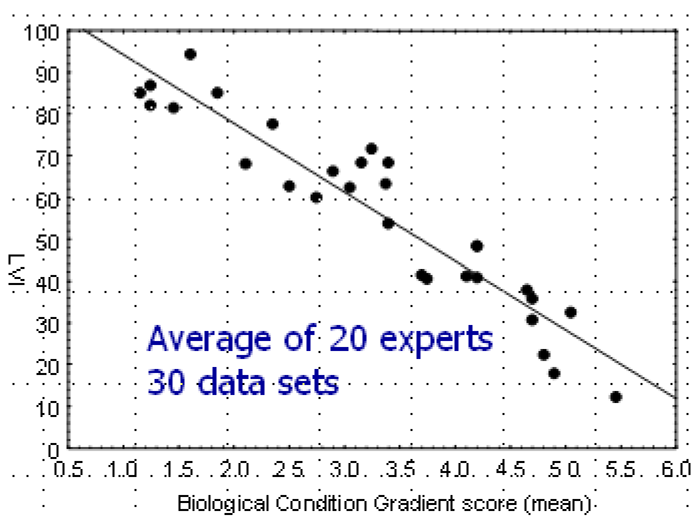

Figure 22. The Lake Vegetation Index regressed against the Biological Condition Gradient scores developed "blindly" by a panel of lake experts. These data reflect updated LVI calculations from the 2007 calibration exercise; Fore et al. (2007b) contains a previous analysis of these data.

Table 14. Biological Condition Gradient (BCG) workshop participants' judgment of which BCG categories should be considered exceptional and impaired for the LVI.

	Exceptional	**Impaired**
Mean	**2**	**4.6**
Median	**2**	**5**
Range	**1-3**	**3-6**

Results from the LVI BCG workshop were also analyzed with a proportional odds logistic regression model (Guisan and Harrell 2000) to describe the relationship between a continuous variable (LVI scores) and a categorical variable (BCG categories). See FDEP (2009) for a full report of this analysis by Lester Yuan of EPA. This model is based on the cumulative probability of a site being assigned to a given tier (*e.g.,* Tier 3) or to any higher quality tier (Tiers 1 and 2). Thus, five parallel models are fit, modeling the probability of assignment to Tiers 5 to 1, Tiers 4 to 1, Tiers 3 to 1, Tiers 2 to 1, and Tier 1 only. Once these five models are fit, the probability of assignment to any single tier can be extracted from the model results.

The mean predictions of the proportional odds logistic regression models are shown in **Figure 23**. Lines are color-coded and labeled by different tiers, and each line can be interpreted as the proportion of experts that assigned samples with the indicated LVI value to a particular tier. For example, approximately 45% of experts assigned a sample with the lowest LVI score to Tier 6 (brown line), while the remaining 55% of experts assigned the sample to Tier 5 (purple line). In the figure, the solid circles represent the actual expert assignments recorded from the workshop for each LVI value. The size of the circle is proportional to the number of experts that assigned a sample to a particular tier, and the circles are color-coded by tier. There is some variability among experts in their assignment of BCG scores, but there is a clear central tendency at any given LVI score.

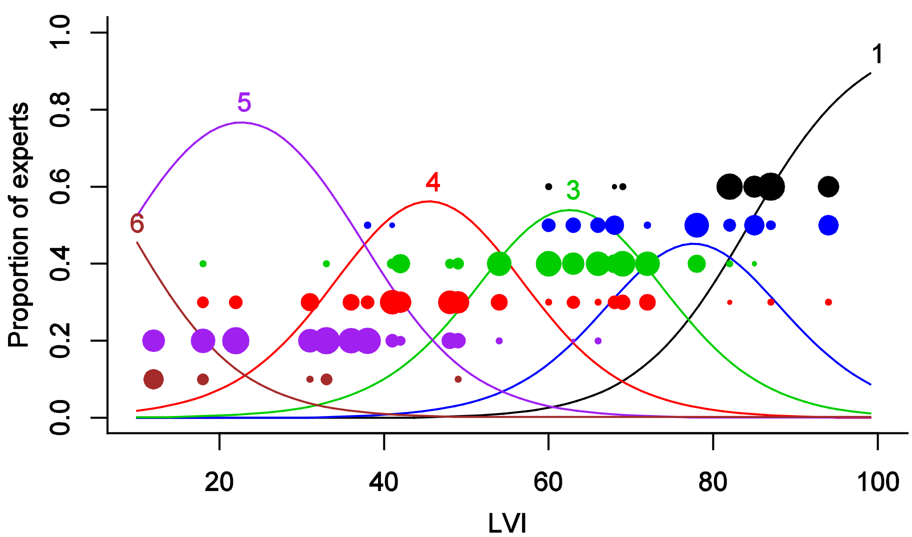

Figure 23. BCG tier assignments based on the Lake Vegetation Index.

The LVI range of approximately 50-58 corresponds with both a low probability of assignment to Tier 5 (*i.e.,* impaired) and a low probability of assignment to Tier 2 (*i.e.*, reference conditions). Thresholds selected in this range of values balance the probability of mistakenly assessing a degraded site as meeting aquatic life use goals with the probability of mistakenly assessing a reference site as impaired.

2011 Analysis

To determine the relationship between the 2011 LVI scores with the same scores used for the 2007 Biological Condition Gradient (BCG) workshop, the LVI scores used for the BCG workshop were recalculated as described above and regressed against the 2007 BCG scores (**Figure 24**). Note that this recalculation involved using the modified C of C scores, the revised FLEPPC Category I metric, and regional scoring for percent Native, with all metrics re-scaled using the 5th and 95th percentiles of the IWRM dataset. Based on EPA's interpretation of the 2007 BCG workshop, a BCG score of four, equivalent to a LVI score of **44.8**, associated with the lowest acceptable biological condition that still met the description of a healthy, well-balanced aquatic community. After the 2011 recalculations, the same BCG score (4) was now equivalent to an LVI score of **42.3**. This reduction of 2.5 points for an equivalent score means that the minimum threshold for meeting a healthy, well balanced plant community should be adjusted by 2.5 points, and that the new acceptable LVI score, based on the BCG approach, would be 42.3. This is just slightly lower than the reference site threshold of **43** points.

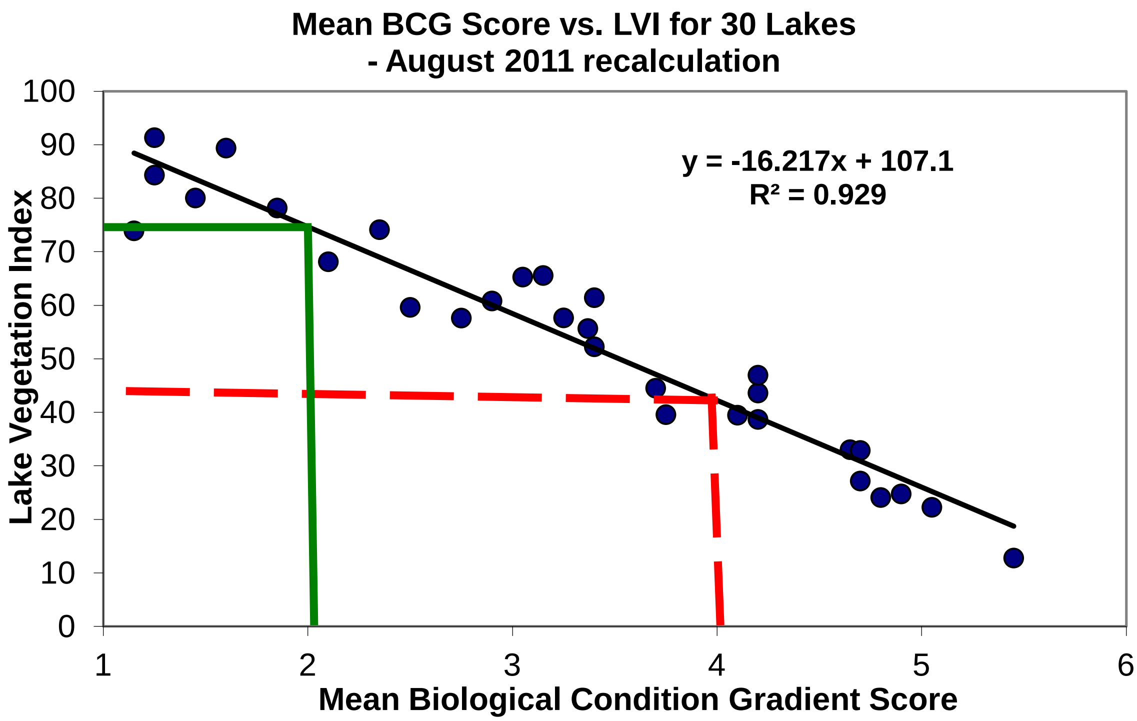

Figure 24. Least squares regression of Biological Condition Gradient vs. recalculated 2011 Lake Vegetation Index scores.

3.3.3 Setting and Evaluating a LVI Impairment Threshold

Weighing these multiple lines of evidence, the DEP has determined that an LVI score of **43** indicates that the designated use is being met, and a score of **42** is impaired. This impairment threshold is supported by the lower distribution of verified reference site scores. The proportional odds analysis provides assurance that plant communities deemed exceptional (BCG Category 2) will not be considered impaired at the adjusted threshold of 43.

4 SCI and LVI Conclusions

The DEP, in consultation with EPA, has used two lines of evidence to set thresholds for exceptional and impaired aquatic life conditions for both the Stream Condition Index and the Lake Vegetation Index. The primary method for establishing the impairment thresholds involved an examination of the lower distribution of minimally-disturbed, rigorously-verified reference site scores. The second approach included an examination of the results of expert opinion elicited through Biological Condition Gradient (BCG) workshops, primarily for the exceptional thresholds, and as a second line of evidence for the minimum threshold for aquatic life use support. For the SCI, the exceptional threshold is a score of 64 and above, while scores below 40 are considered impaired. For the LVI, the exceptional threshold is a score of 78 and above, while scores below 43 are considered impaired.

5 Literature Cited

Alix, M.S., R.W. Scribailo. 1998. Aquatic plant species diversity and floristic quality assessment of Saugany Lake, Indiana. Proceedings of the Indiana Academy of Science 107: 123-139.

Andreas, B.K. 1995. A floristic assessment system for Northern Ohio. Technical Report WRP-DE-8. February 1995. U.S. Army Corps of Engineers, Washington, D.C.

Barbour M. T., Gerritsen J., G. E. Griffith, R. Frydenborg, E. McCarron, J. S. White, and M. L. Bastian. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. Journal of the North American Benthological Society 15:185-211.

Beck, W. M. 1954. Studies in stream pollution biology. I. A simplified ecological classification of organisms. Quarterly Journal of the Florida Academy of Sciences 17(4):212-227.

Brown, M. T., and B. Vivas. 2004. Landscape development intensity index. Environmental Monitoring and Assessment 101: 289-309

Davies, S.P. and S.K. Jackson. 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic systems. Ecological Applications 16(4):1251-1266.

Ervin, G.N., B.D. Herman, J.T. Bried, D.C. Holly. 2006. Evaluating non-native species and wetland indicator status as components of wetlands floristic assessment. Wetlands 26: 1114-1129.

Fennessy, S., M. Gernes, J. Mack, and D. H. Wardrop. 2002. Methods for evaluating wetland condition. # 10 Using vegetation to assess environmental conditions in wetlands. EPA-822-R-02-020. U.S. Environmental Protection Agency. Office of Water, Washington, D.C. http://www.epa.gov/waterscience/criteria/wetlands/10Vegetation.pdf

Florida Department of Environmental Protection (FDEP) 2009. Draft Technical Support Document: Development of Numeric Nutrient Criteria for Florida Lakes and Streams. Florida Department of Environmental Protection, Standards and Assessment Section. June 2009. Available http://www.dep.state.fl.us/water/wqssp/nutrients/index.htm

Fore, L.S, R.B. Frydenborg, D. Miller, T. Frick, D. Whiting, J. Espy, and L. Wolfe. 2007a. Development and Testing of Biomonitoring Tools for Macroinvertebrates in Florida

Streams. Florida Dept. Environmental Protection. 110 pp. Available http://www.dep.state.fl.us/water/bioassess/pubs.htm

Fore, L.S., R. Frydenborg, N. Wellendorf, J. Espy, T. Frick, D. Whiting, J. Jackson, and J. Patronis. 2007b. Assessing the Biological Condition of Florida's Lakes: Development of the Lake Vegetation Index (LVI). Report to the Florida Department of Environmental Protection. 106 pp. Available http://www.dep.state.fl.us/water/bioassess/pubs.htm

Griffith, G. E., J. M. Omernik, C. M. Rohm, and S. M. Pierson. 1994. Florida regionalization project. Corvallis, Oregon: U. S. Environmental Protection Agency, Environmental Research Laboratory.

Guisan, A. and F. E. Harrell. 2000. Ordinal response regression models in ecology. Journal of Vegetation Science 11:617 – 626.

Karr, J.R. 1991. Biological integrity: A long-neglected aspect of water resource management. Ecological Applications 1:66-84.

Lopez, R.D. and M.S. Fennessy. 2002. Using the floristic quality assessment index (FQAI) in wetlands as a biological indicator of landscape condition along a gradient of human influence. Ecological Applications 12(2): 487-497.

Mack, John J. 2004. Integrated Wetland Assessment Program. Part 4: Vegetation Index of Biotic Integrity (VIBI) and Tiered Aquatic Life Uses (TALUs) for Ohio wetlands. Ohio EPA Technical Report WET/2004-4. Ohio Environmental Protection Agency, W etland Ecology Group, Division of S urface Water, Columbus, Ohio.

Miller, S.J., D.H. Wardrop, W.M. Mahaney, and R.P. Brooks. 2006. A plant-based index of biological integrity (IBI) for headwater wetlands in Pennsylvania. Ecological Indicators 6 (2):290-312.

Mortellaro, S., M. Barry, G. Gann, J. Zahina, S. Channon, C. Hilsenbeck, D. Scofield, G. Wilder, and G. Wilhelm. 2009. Coefficients of Conservatism Values and the Floristic Quality Index for the Vascular Plants of South Florida. US Fish and Wildlife Service, South Florida Ecological Services Field Office. January 2009.

Nichols, S., S. Weber, and B. Shaw. 2000. A proposed aquatic plant community biotic index for Wisconsin lakes. Environmental Management 26:491-502.

Rocchio, J. 2006. Vegetation Index of Biotic Integrity for Southern Rocky Mountain Fens, Wet Meadows, and Riparian Shrublands: Phase 1 Final Report to Colorado Department of National Resources, Denver CO.

Rothrock, P.E., T.P. Simon, P.M. Stewart. 2008. Development, calibration, and validation of a littoral zone plant index of biotic integrity (PIBI) for lacustrine wetlands. Ecological Indicators 8: 79-88.

Wilhelm, G.S. and L.A. Masters. 1995. Floristic Quality Assessment in the Chicago Region and Application Computer Programs, Morton Arboretum, Lisle, IL. 17 pp.