Development and Testing of Biomonitoring Tools for Macroinvertebrates in Florida Streams

Leska S. Fore
Statistical Design
136 NW 40th St.
Seattle, WA 98107
leska@seanet.com

Final Report

February 2004

Prepared for:

Russel Frydenborg &
Ellen McCarron
Florida Department of Environmental Protection
2600 Blair Stone Rd.
Tallahassee, FL 32399-2400

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Florida DEP assesses the chemical, physical and biological condition of hundreds of stream sites each year. This study used measures of hydrologic condition, riparian and channel habitat condition, water chemistry, and intensity of human land use to define a gradient of human disturbance for stream sites. I evaluated the sensitivity and tolerance of over 1000 stream macroinvertebrate taxa using the human disturbance gradient (HDG). I tested for correlation between the HDG and 36 biological measures of the stream macroinvertebrate assemblage (metrics) and selected the 10 most highly correlated metrics within six categories of biological organization. The six categories (and their selected metrics) were taxonomic richness (total number of taxa, number of Trichoptera taxa, and number of Ephemeroptera taxa); feeding group (percentage filterer individuals); voltinism (long-lived taxa richness); habit (clinger taxa richness); community structure (percentage dominance of the most abundant taxon and percentage Tanytarsini midges); and sensitivity and tolerance (sensitive taxa richness and percentage very tolerant individuals).

Metrics were combined into an overall stream condition index (SCI) by transforming metric values into unit-less scores and summing the scores. The SCI was highly correlated with HDG for an independent data set (Spearman's $r = -0.81$, $p << 0.001$). SCI was independent of watershed size and geographic region (panhandle, peninsula and northeast). Across 10 years of sampling, the index showed a similar response to the HDG. SCI was somewhat higher for winter vs. summer samples (3.5%). A large portion of the variability of SCI was due to subsampling in the laboratory (49%). Confidence intervals based on estimates of SCI variance defined 3.7 categories of biological condition that the SCI could detect assuming a single sample. For two site samples, the SCI could detect five categories of biological condition.

Biological metrics were also tested for a second stream macroinvertebrate sampling protocol (BioRecon) based on sorting of invertebrates in the field and taxonomic identification in the laboratory. Of the 10 SCI metrics, six taxa richness metrics were tested using BioRecon data; all were highly correlated with both SCI and HDG. From these metrics the BioRecon index was calculated as the sum of scores for the six metrics. The BioRecon index could detect 2.5 categories of biological condition for one sample and 3.5 categories for two samples.

INTRODUCTION

Florida, along with a handful of other states, lead the nation in the development and implementation of water protection policies based on biological criteria (McCarron and Frydenborg, 1997; EPA, 2002b). Under the Clean Water Act, states are required to define designated uses for specific water bodies and develop criteria to protect them (Karr, 1991; Ransel, 1995). Over time, the emphasis has shifted from primarily chemical monitoring of pollutants to include the direct measurement of the condition of the biological assemblage (Yoder and Rankin, 1998).

Florida classifies its surface waters according to their designated uses. The five categories are ranked in order of the level of protection that they require: Class I – drinking water; Class II – shellfish propagation or harvesting (primarily for coastal waters); Class III – recreation and propagation and maintenance of a healthy, well-balanced population of fish and wildlife; Class IV – agricultural water; and Class V – navigation and industrial use. This document describes the development and testing of biological monitoring tools for macroinvertebrate stream samples for the assessment of freshwater streams that are applicable to Class I, III, IV, and V waters.

The U.S. Environmental Protection Agency (EPA) has developed numerous guidance documents to support biological assessment (EPA, 1998; Barbour et al., 1999; Bowman et al., 2000; Jackson et al., 2000; EPA, 2001; EPA, 2002a; EPA, 2003a, b, c; Fore, 2003). These documents typically emphasize multimetric indexes to measure the biological condition of surface waters (Karr 1981; Karr et al., 1986; Barbour et al., 1999). EPA's recent *Draft Report on the Environment 2003* specifically recognizes multimetric indexes as an "approach that addresses the need to measure critical multiple variables" (EPA, 2003a). Multimetric indexes are composed of biological measures, called metrics, that show a consistent response to human disturbance (Karr and Chu, 1999). Metrics are converted to unit-less scores and summed to obtain a summary index. Multimetric indexes have been developed for fish, birds, invertebrates, and algae in a variety of ecological contexts in North America, Europe and Asia (Thorne and Williams, 1997; Hughes et al., 1998; Hughes and Oberdorff, 1998; Karr and Chu, 1999; Bryce et al., 2002; Fore and Grafe, 2002; Klemm et al., 2003). Most states use multimetric indexes for biological assessment, although other techniques have been developed (Hawkins et al., 2000; EPA, 2003b, c).

Background

Barbour et al. (1996) developed multimetric indexes for Florida that have been used since the early 1990's. They tested metrics by comparing values at sites with minimal human influence (reference sites) and sites with known disturbance (test sites). The authors identified three unique geographic regions based on the presence of shared taxa and defined metric expectations for each region.

In contrast, metrics for this study were selected on the basis of their correlation with a gradient of human disturbance (Bryce et al., 1999). The sensitivity and tolerance of over 1000 taxa were also evaluated individually using the human disturbance gradient (HDG). I used confidence intervals to define the number of categories of biological condition that the indexes could reliably detect. These categories support decisions related to listing stream sites as impaired and prioritizing sites for additional evaluation as part of the TMDL (Total Maximum Daily Load) process (Karr and Yoder, 2004). These statistically robust categories may also be used to monitor degradation of excellent or outstanding waters.

From the original analysis by Barbour et al. (1996), two multimetric indexes were developed for Florida streams based on different sampling and identification protocols. The stream condition index (SCI) was derived from laboratory processing and identification of samples while the BioRecon (biological reconnaissance) index was based on field sorting and laboratory identification. This study develops new bioassessment tools for both sampling protocols. For simplicity, the names of the indexes have been retained for this new analysis and the previous indexes are referred to as the old SCI and the old BioRecon index within this document.

METHODS

Study area

Portions of Florida have been repeatedly inundated by seawater in recent geologic history; as a consequence, elevation and organism recruitment from adjacent, non-inundated areas play an important role in the present-day geographic distribution of macroinvertebrates. Florida can be divided into the northern panhandle, the southern peninsula, and a transition region known as the northeast. The middle and lower Suwannee basin provides a natural demarcation between the panhandle and peninsula, with the northeast region straddling the upper Suwannee northeast of the Cody escarpment (White, 1970).

In the panhandle, rivers typically flow from north to south and the elevation of headwater streams typically exceed 200 to 250 ft. Recruitment of stream organisms (over geologic time) from upstream, higher elevation areas (Georgia and Alabama) makes the panhandle richer in freshwater stream taxa than other parts of the state (FNAI, 1990). Vegetation communities in the panhandle generally consist of mixed pine/oak/hickory forests (*Pinus* spp., *Quercus* spp., *Carya* spp.), longleaf pine forests (*Pinus palustris*), hardwood forests with beech/magnolia climax community (*Fagus grandiflora*/*Magnolia grandiflora*), and swamp hardwood forests of cypress (*Taxodium* spp.) or tupelo (*Nyssa* spp.), interspersed by a mosaic of pine plantations, cropland (e.g., corn, soy beans, peanuts), and pasture (SWCS, 1989; Fernald and Purdum, 1992). The panhandle is less densely populated by humans than the other areas.

The peninsula has a sandy highland ridge extending down its center almost to Lake Okeechobee. The elevation of the central ridge is approximately 150 to 200 ft. Rivers west of the ridge flow into the Gulf of Mexico. Of the two large drainages east of the ridge, the Kissimmee River flows south into Lake Okeechobee and the St. Johns River flows north to the Atlantic. Streams in the peninsula tend to have lower water velocity, flowing through wetlands and pine flatwoods (White, 1958). The peninsula has fewer taxa of freshwater invertebrates due to lack of upstream recruitment sources that did not experience marine inundation. Plecoptera, for example, are common in the panhandle but extremely rare in the peninsula, probably due to a combination of poor recruitment, warmer temperatures, and lower water velocities. Peninsular vegetation communities on the ridge consist of longleaf pine/turkey oak forests (*Pinus palustris/Quercus laevis*), on flat areas are slash pine (*Pinus eliottii*)or loblolly pine (*Pinus teada*) with palmetto/gallberry understory (*Serenoa repens/Ilex glabra*), and in depressional areas are marsh/wet prairies (maidencane, pickerel weed), and hardwood wetlands of sweetbay (*Magnolia virginiana*), cypress (*Taxodium* spp.), and ash (*Fraxinus* spp.; SWCS, 1989). The dominant land use is pasture, cropland (e.g., watermelons, nursery products, tomatoes), and urban areas (Fernald and Purdum, 1992). Dense population centers are located at Tampa and Orlando.

The northeast region includes portions of the Okeefenokee Swamp, parts of the upper Suwannee drainage, the Black Creek drainage, and the Sea Island flatwoods. From the perspective of stream invertebrates, this region tends to be intermediate between the panhandle and peninsula regions. With the exception of the Black Creek basin with headwaters approximately 200 ft above sea level, other streams in this area tend to be low relief, wetland systems (Snell and Anderson, 1970). Plant communities consist of longleaf pine/turkey oak (*Pinus palustris/Quercus laevi)* on the sandy uplands, hardwood wetlands of cypress (*Taxodium* spp.), tupelo (*Nyssa* spp.), and loblolly bay (*Gordonia*

lasianthus), pine flatwoods, and marsh (FNAI, 1990). Jacksonville is the only major population center.

Barbour et al. (1996) used macroinvertebrate stream samples collected from sites with minimal human influence to evaluate the influence of geographic locations on species assemblages. They found that the panhandle, peninsula, and northeast regions, which are primarily defined according to drainage patterns, were better predictors of species assemblages than were sub-ecoregions described by Griffith et al. (1994). Regional differences translate into expectations of higher taxonomic diversity in the panhandle relative to the peninsula, with the northeast region representing a transitional area.

Site selection and data sets

The Florida Department of Environmental Protection (FDEP) has collected thousands of macroinvertebrate samples from stream sites throughout the state. Many sites have been visited repeatedly. Sites and visits were selected from this large data base using different criteria for different testing situations (Table 1). For example, when testing for metric correlation with disturbance, a range of site conditions was needed; to evaluate variability, sites with many repeat visits were needed; to evaluate laboratory subsampling, large samples were needed. Some data sets contained data from the same sites, other sites were selected to provide an independent test. For some testing situations, data from site visits were averaged by sites, for example, when one variable in the analysis had only one value for each site, such as HDG.

The relationship between human disturbance and biological condition was evaluated in several ways. To develop the human disturbance gradient (HDG), FDEP biologists selected sites from their regions that represented the broadest possible range of human influence. First, to insure that the relationships observed between human land use and biological metrics were consistent and not specific to a particular set of sites, two independent sets of sites were tested. These data were then combined to obtain 223 sites to test for metric correlation with HDG. An additional, independent set of 23 sites was used to test correlation between SCI and the HDG.

Additional data sets were compiled to estimate the variability of the biological metrics and indexes associated with repeat visits to a site, estimate the variability associated with laboratory subsampling, and evaluate the influence of field sorting vs. laboratory sorting. To evaluate SCI variability, data from 62 sites with 2–8 visits were selected. About half the sites (32) had repeat visits on the same day. Most of the other repeat visits were during different years, and a few were within the same year. Thus,

these data could not be used to estimate the influence of seasons on variability. Instead, additional data were compiled to test for seasonal (winter vs. summer) differences in SCI. To test the influence of subsampling during laboratory processing for the SCI protocol, three subsamples were selected from 59 samples collected from 54 sites (5 sites had two duplicate samples each).

To test metrics and evaluate their variability for the BioRecon protocol, two additional data sets were used. The first data set included 116 sites with macroinvertebrate samples collected according to both the BioRecon and the SCI sampling protocol. I tested BioRecon metrics for correlation with SCI and HDG. Most sites had at least two visits and several had 3–6 visits. Of the 116 sites, 53 had values calculated for HDG because they were used initially to develop the HDG. To evaluate the variability of the BioRecon metrics and index, 128 sites were selected with two to seven visits per site.

Table 1. Description of various data sets, number of sites sampled, total number of visits to all sites (varied by site) and total number of subsamples taken in the laboratory for all visits. Some sites were included in more than one data set with the exception of the 23 independent sites used to verify correlation between SCI and HDG.

Purpose	**Sites**	**Visits**	**Lab subsamples**
Test HDG consistent for 2 rounds of data			
Round 1	154	469	0
Round 2	69	160	0
Test metrics vs. HDG	223	629	0
Score metrics (same data as above but only samples with 75–175 individuals included)	176	420	0
Test sensitive & tolerant taxa vs. HDG	226	632	0
Verify SCI with independent data	23	23	0
SCI variability	62	220	0
Seasonal variation of SCI	78	578	0
Laboratory subsampling variability for SCI	54	59	177 (3 each)
Calibrate BioRecon metrics with SCI	116	165	0
Test BioRecon metrics vs. HDG	53		0
BioRecon variability	128	293	0

The data set used to test SCI and its component metrics against the human disturbance gradient was large; therefore, simple scatter plots of data against the HDG would be difficult to interpret with the many data points and overstrikes. Consequently, I used box plots to display the range of values associated with different values of HDG. In all figures, the box defined the 25th and 75th percentiles, the whiskers were the non-outlier range, and the outliers were calculated as values greater (or less) than 1.5 times the length of the distance from the 25th to the 75th percentile value. FDEP is currently modifying laboratory protocols to more often match the target of 100 individuals; therefore, several graphs include data from samples with 75-175 individuals. This range is broader than the current target of 95–110 individuals, but minimized the loss of sites from the data set. Additional analysis (not shown) indicated that the ranges of metric values were similar for samples with <125 individuals and <175 individuals. For the BioRecon testing, the number of individuals identified was not recorded. All site visits were used within a particular analysis unless specifically stated otherwise.

Quantifying human disturbance

Karr and colleagues describe five factors to summarize the ways in which humans alter and degrade rivers and streams (Karr et al., 1986; Karr et al., 2000). The five factors are flow regime, physical habitat structure, water quality, energy source, and biological interactions. First, the natural flow regime of a stream may be altered by dams, water removal or water return. In extreme cases, changes in water volume or flow timing can result in flooding or dry channels. Second, stream macroinvertebrates depend on a variety of physical habitat types for food, shelter and reproduction. Greater diversity of substrate size, types of vegetation, and bank architecture translate into a greater variety of organisms. Third, water quality may be evaluated in terms of turbidity, conductivity, nutrient concentrations, or the presence of other chemicals or metals. Fourth, energy source describes the seasonal availability, type of organic material, and size and shape of material available as food. Fifth, introduction of exotic species or disease and harvest of fish and shellfish for sport, commercial, and subsistence can alter biological interactions such as competition or predation within the natural assemblage.

Data were available to measure four of the five factors. FDEP biologists developed a hydrologic scoring system based on their knowledge of water removal, patterns of drought, and hydrographs for the sites (Table 2). To measure habitat condition, FDEP biologists routinely calculate a habitat condition index whenever stream macroinvertebrate samples are collected. The index evaluates substrate condition and availability, water velocity, habitat smothering (e.g., by sand and silt), channelization, bank stability, and the width and condition of the riparian vegetation.

Table 2. Hydrologic condition of stream site, scoring range for hydrologic index, and description of human influences. Assignment of high or low values within each condition class were assigned at the discretion of the biologist based on the extent of disturbance. These changes were associated with human disturbances, not natural events such as hurricanes or extreme droughts.

Condition	Score	Description
Excellent	1–2	Flow regime as naturally occurs (slow and fairly continual release of water after rains), few impervious surfaces; high connectivity with ground water and surface features delivering water (e.g., sandhills, wetlands; no ditches or berms)
Good	3–4	Flow regime minimally changed; some water withdrawals; some wetland drainage, some impervious surfaces, some ditching
Moderate	5–6	Flow regime moderately altered; hydrograph moderately flashy (scouring after rain events with subsequent reductions in flow), groundwater pumping evident; much wetland drainage, topographic alterations reduce natural water input; more impervious surfaces, dams/control structures change normal water delivery schedule
Poor	7–8	Flow regime highly altered; hydrograph very flashy (scouring after rain events with subsequent reductions in flow, leading to stagnant or dry conditions, related to large amounts of impervious surfaces and/or ditching); water withdrawals and impoundments or control structures severely alter flows, large amounts of impervious surfaces
Very poor	9–10	Flow regime entirely human controlled; hydrograph very flashy (scouring after rain events with subsequent reductions in flow, leading to stagnant or dry conditions, related to impervious surfaces and ditching); intensity of water withdrawals and impoundments fundamentally alter the nature of the ecosystem

FDEP biologists also routinely collect measurements of turbidity, temperature, dissolved oxygen, conductivity and nutrient concentrations. Total phosphate (TP), ammonia (NH_3), total Kjeldahl nitrogen (TKN), and nitrites/nitrates (NO_x) were measured to summarize nutrients. I selected NH_3 to summarize water quality because it was the most consistently associated with the other water quality measures and had the most complete record of data. TP had similar high correlations and complete data, but may be more closely associated with fertilizer and farming practices while NH_3 may be a more general indicator of both urbanization and agriculture.

A measure of the type and quantity of energy available to organisms living within the stream was not available. However, Brown and colleagues have developed an index to estimate the intensity of human land use based on nonrenewable energy flow (Brown et al., 1998; Brown and Vivas, *in review*). The landscape development intensity (LDI) index was calculated as the percentage area within the catchment of particular types of land use multiplied by the coefficient of energy use associated with that land use, summed over all land use types in the catchment (Table 3).

$$LDI = \sum \left(LDI_i * \%LU_i \right).$$

Where,

LDI_i = the nonrenewable energy land use for land use *i*, and

$\%LU_i$ = the percentage of land area in the catchment with land use *i*.

Brown and colleagues derived the coefficients for each land use type from actual billing records and published literature, translated reported energy use into standardized units, then averaged the values by land use type and standardized to a per unit area. Only non-renewable energies were used in calculations and included electricity, fuels, fertilizers, pesticides, and water (both public water supply and irrigation).

Land use was derived from aerial photos manipulated as layers in a GIS computer program. LDI was calculated two ways. First, a buffer area of 100 m on each side of the stream and 10 km upstream of the sampling point was used (LDI_BF). Because the definition of the buffer region was somewhat arbitrary, LDI was also calculated for the entire upstream catchment (LDI_WS) in order to determine which spatial scale was a better predictor of site condition. I selected LDI_BF to include in HDG because it was slightly more highly correlated with NH_3, the hydrologic index, and the habitat index (Table 4).

Table 3. Description of land use and the coefficient value used to calculate the LDI. Higher values indicate greater intensity of human land use (Brown and Vivas, *in review*).

Land use	**LDI value**
Natural Open water	1.00
Natural System (no human activity)	1.00
Pine Plantation	1.58
Woodland Pasture	2.02
Pasture	2.77
Recreational / Open Space (Low-intensity)	2.77
Low Intensity Pasture (with livestock)	3.41
Citrus	3.68
High Intensity Pasture (with livestock)	3.74
Row crops	4.54
Single Family Residential (Low-density)	6.79
Recreational / Open Space (High-intensity)	6.92
High Intensity Agriculture	7.00
Single Family Residential (Med-density)	7.47
Single Family Residential (High-density)	7.55
Low Intensity Highway	7.81
Low Intensity Commercial	8.00
Institutional	8.07
High Intensity Highway	8.28
Industrial	8.32
Low Intensity Multi-family residential	8.66
High intensity commercial	9.18
High Intensity Multi-family residential	9.19
Low Intensity Central Business District	9.42
High Intensity Central Business District	10.00

To define the human disturbance gradient (HDG), I converted the four measures of human disturbance to unit-less scores and summed the scores to create HDG values for each stream site. Three of the measures had values of 0, 1, or 2 indicating low, moderate or high levels of human influence. One measure, hydrologic condition, had an additional value of 4 available to indicate extreme levels of disturbance (Table 5). Thus, values for the HDG ranged from 0 (indicating minimal human disturbance) to 9 (indicating extreme disturbance). The scores for the HDG were derived from a graphical analysis of EPT taxa richness plotted against each measure of disturbance. I identified values for each disturbance measure for which the number of EPT taxa was consistently lower. For NH_3, 9 sites had no information. Most of these sites had values for other measures indicating minimal human disturbance. I assumed for these sites that NH_3 was < 0.1 mg/L.

Table 4. Correlation table of NH_3 (mg/L), hydrologic index, habitat index, LDI for the buffer, LDI for the upstream watershed, natural log of watershed size, HDG, the current SCI, and the old SCI for 223 stream sites (except NH_3 for which n=214). Repeat visits to individual sites were averaged for each site. All values shown were significant (Spearman's r, $p < 0.01$).

	NH_3	**Hydro**	**Habitat**	**LDI_BF**	**Ln(Size)**	**HDG**	**SCI**
NH_3	—	0.43	–0.45	0.39	–0.20	0.50	–0.56
Hydro	0.43	—	–0.66	0.61	–0.36	0.84	–0.71
Habitat	–0.45	–0.66	—	–0.64	0.41	–0.78	0.66
LDI_BF	0.39	0.61	–0.64	—	–0.37	0.79	–0.60
LDI_WS	0.34	0.56	–0.55	0.74	–0.28	0.70	–0.59
Ln(Size)	–0.20	–0.36	0.41	–0.37	—	–0.41	0.36
HDG	0.50	0.84	–0.78	0.79	–0.41	—	–0.76
SCI	–0.56	–0.71	0.66	–0.60	0.36	–0.76	—
Old SCI	–0.43	–0.59	0.56	–0.54	0.29	–0.69	0.86

Stream macroinvertebrate sampling

FDEP uses two protocols to collect invertebrate samples from streams. The protocols differ in the number of dip net sweeps collected, the location of sample sorting (field vs. laboratory), and the final index used for assessment. The BioRecon sampling protocol was designed to be a quick field assessment that could yield a same-day measure of stream site condition. The SCI sampling protocol was meant to be more precise. Both protocols use a D-frame dip net with a 0.25 m opening and a 600 μm-mesh net. Both protocols distribute the sweeps among the best available substrates on an approximately equal basis. Florida streams typically have a sandy substrate that supports very few invertebrates. Therefore, the habitat sampled for macroinvertebrates includes logs, roots and undercut banks, partially decayed leaf packs, aquatic vegetation, and rocks or large cobble. During SCI sampling, if four productive substrates are present (snags, roots, vegetation and leaf packs), each substrate would be sampled four times, with sand being sampled four times, for a total of 20 sweeps.

In the laboratory, the SCI sample is divided by spreading onto gridded trays. Squares are randomly selected and all the material within a square is sorted. Macroinvertebrates are identified to the lowest practical taxonomic level. Squares are selected until the target sample size of 100 individuals is reached. To avoid bias, the final square selected is always finished even if it contains more than 100 individuals. The number of taxa and the relative abundance of each taxon are derived from these data.

Table 5. Scoring rules for measures used to calculate the human disturbance gradient (HDG). HDG is the sum of the scores for each site.

Measure	**0**	**1**	**2**	**3**
NH_3	<0.1	0.1–2	>2	
Habitat index	>65	50–65	<50	
Hydrologic index	<6	6–7	8–9	10
LDI	<2	2-3.5	>3.5	

BioRecon data are collected using four dip net sweeps from a stream site. Samples are sorted for live macroinvertebrates in the field. The sample is spread in a pan and the biologist searches through all the material for every unique taxon present. Each taxon is preliminarily identified in the field and representative specimens of each taxon are preserved and returned to the laboratory for final identification under a dissecting microscope. From this field-sort protocol, the number of taxa present at a site can be measured, but their relative abundance (% composition) cannot.

Metric development and testing

Several authors have described relevant candidate metrics for stream macroinvertebrate assemblages (Barbour et al. 1996 and 1999; Kerans and Karr, 1994; and Klemm et al., 2002). For this study, 36 candidate metrics related to taxonomic richness, feeding group, voltinism, habit, community structure, and tolerance or sensitivity were evaluated for their association with HDG (Table 6). Within each category, I selected metrics with the highest correlation with HDG (Spearman's *r*). I used one-sided tests to calculate *p*-values because if a metric indicated better biological condition in response to increased human disturbance, that relationship would not be *biologically* significant even if it were statistically significant. All metrics were also evaluated graphically to ensure that they were reliable indicators of disturbance. I tolerated variability in metric values for minimally disturbed sites (low HDG), because not all sources of disturbance were known. For example, water chemistry data were unavailable for metals, pesticides or other toxics. In contrast, I considered an inconsistent response at more highly disturbed sites to be unacceptable.

Taxonomic richness. Richness was calculated as the number of unique taxa found within a particular group such as the Ephemeroptera, Plecoptera, Trichoptera, or their combination (EPT).

Feeding group. Trophic designations followed Merritt and Cummins (1996); some designations have been modified by FDEP biologists and others to reflect natural histories in Florida (Merritt et al., 1996). All feeding group metrics were calculated as the number of individuals of a particular group divided by the total number of individuals in the sample times 100 (percentage).

Table 6. Candidate metric, correlation with the human disturbance gradient (HDG), and whether the metric was included in the final SCI, BioRecon, or original SCI index. Correlation was tested for 223 sites with repeat visits to each site averaged before statistical testing. Hyperabundance was tested without averaging, n=577. All correlations shown were significant at $p < 0.01$ (Spearman's *r*, one-sided test).

Metric	HDG	SCI	BioRecon	Old SCI
Taxonomic richness				
EPT	–0.73			yes
Trichoptera	–0.66	yes	yes	
Ephemeroptera	–0.66	yes	yes	
Total taxa	–0.52	yes	yes	yes
Diptera	–0.44			
Chironomidae	–0.40			yes
Plecoptera	–0.37			
Oligochaeta	0.32			
Non-insect	—			
Feeding group				
% Filterer	–0.46	yes		yes
% Scavenger	–0.38			
% Browser/grazer	–0.30			
% Collector/gatherer	0.29			
% Predator	–0.21			
% Plant piercer	–0.18			
% Parasite	—			
% Scraper	—			
% Shredder	—			
Voltinism				
Long-lived taxa	–0.41	yes	yes	
% Long-lived	–0.35			
Habit				
Clinger taxa	–0.71	yes	yes	
% Clinger	–0.61			
Community structure				
% Dominance	0.47	yes		yes
Hyperabundance	—			
% Ephemeroptera	–0.58			
% Trichoptera	–0.55			
% Tanytarsini (Chironomidae)	–0.45	yes		

Metric	HDG	SCI	BioRecon	Old SCI
% Plecoptera	–0.37			
% Oligochaeta	0.32			
% Non-insect	0.32			
% Chironomidae	—			
% Diptera	—			yes
Sensitivity/tolerance				
Sensitive taxa	–0.75	yes	yes	
Florida index (sensitive)	–0.71			yes
% Very tolerant	0.70	yes		
Very tolerant taxa	0.61			

Voltinism. Most stream insects spend the majority of their lives as larvae underwater and emerge as adults to mate. The length of time between generations is called voltinism and may be less than a year, approximately one year, or greater than one year. FDEP biologists used published sources to identifiy taxa that require more than one year to complete their life cycles (Appendix A; Brigham et al., 1982; Thompson, 1984; Thorp, 1991; Pescador et al., 1995; Corbet, 1999; Pescador et al., 2000; Smith, 2001; Rasmussen and Pescador, 2002). These taxa require consistent flows and other favorable conditions throughout the year to complete their life cycles. Data on generation length were not available for all Florida taxa; therefore, some long-lived taxa may not be included in the current list.

Included in the list were all taxa in the Cordulegastridae and selected taxa in the families Aeshnidae, Gomphidae and Libellulidae (Odonata); the family Leuctridae and selected genera in the Perlidae and Pteronarcidae (Plecoptera); the family Corydalidae (Megaloptera); and the genera *Ceraclea*, *Pycnopsyche*, *Molanna*, and *Rhyacophila* (Trichoptera). Several Coleoptera taxa (beetles) also require more than a year to complete their life cycles, but were not included because the FDEP database does not currently distinguish between larvae (which cannot travel to find water) and adults (which can). Non-insects included in the long-lived list were all taxa in the Decapoda; the genera *Fossaria* and *Pomacea* (Gastropoda); and the families Unionidae and Corbiculidae (Pelecypoda).

Habit. Other studies have noted a strong correlation between the taxonomic richness of clingers and stream condition (Fore et al., 2001; Karr and Morishita Rossano, 2001; Mebane, 2002). Clingers have morphological and behavioral adaptations that allow

them to cling to substrate (Merritt and Cummins, 1996). Channel dredging and removal of riparian vegetation eliminates the habitat preferred by clingers.

Community structure. Percentage dominance was calculated as the number of individuals in the most abundant taxon divided by the total number of individuals in the sample. 'Hyperabundance' represented a measure of the density of stream macroinvertebrates present at a site. Sites with high organic enrichment (e.g., sewage effluent) may have a high density of invertebrates due to the increased source of food for many species. The actual number of individuals present in the total sample was not recorded; however, the number of grids in the sample tray that were searched to best approximate the target number of 100 individuals was recorded. I used the number of grid cells searched as a measure of the density of invertebrates present, or 'hyperabundance.' Other measures of community structure were based on the number of individuals within specific taxonomic groups divided by the total number of individuals.

Sensitivity and tolerance. Beck (1954) developed a list of sensitive taxa for Florida that was used to calculate a biotic index which was later modified to be the Florida index. To calculate the Florida index (which is more accurately termed a metric), very sensitive taxa are given a score of 2 and sensitive taxa a score of 1. The scores are summed to calculate this metric.

I used the HDG to evaluate the sensitivity and tolerance of 1195 taxa names. The taxa names included taxa that were not unique because the data base included synonyms for the same taxa as well as specimens identified to genus and species for the same genus. Because many sites had multiple visits and because less disturbed (reference) sites tended to have a greater number of visits, I summarized data for each taxon *by site* to avoid bias associated with greater sampling effort at less disturbed sites. I counted a taxon as present at a site if it was found for >= 50% of the site visits. I divided the 226 sites into two groups and, for simplicity, refer to these groups as 'good' and 'poor' for statistical testing. Good sites were defined as those with moderate levels of disturbance for no more than two measures (e.g., LDI or NH_3) or high disturbance for no more than one measure (e.g., habitat condition). This translated into HDG <=2 and included 141 sites. The remaining 85 sites were classified as 'poor.' Thus, the two groups might be more accurately labeled 'better' and 'worse' rather than 'good' and 'poor.' The relatively high number of sites selected to represent 'good' condition reflects the large number of reference sites in the data set.

I used contingency table analysis to determine what percentage of taxon occurrences in the good vs. poor range of HDG values represented a statistically significant difference (chi-square test, Yates continuity correction). As for any test, the

number of sampling points increases the power of the test, such that less extreme difference will be statistically significant. For example, if a taxon occurred at 10 sites, all of the occurrences (100%) would have to be in the good range of the HDG to be statistically significant ($p < 0.05$). In contrast, if a taxon occurred at 38 sites, it would only need to occur at 30 of the good sites (79%) and would still be statistically significant. For this reason, I chose not to use a strictly statistical criterion for selecting sensitive taxa. Instead, for more common taxa I only selected those with 87% or more of their occurrences in good sites. Using this criterion of 87%, if a taxon occurred at 30 sites, 27 or more of the sites would have to be in the good range of the HDG. The *p*-value associated with this would be $p < 0.002$. Thus, a strictly statistical criterion would have selected taxa that *prefer* good conditions rather than taxa that *require* good conditions. Thus, I used a statistical rule to define the minimum number of occurrences needed for statistical significance but made the standards for selection more strict ($p << 0.05$) for taxa that were more common.

Defining a statistical rule for selection of tolerant taxa was not possible because tolerant taxa can also "tolerate" good conditions. Instead, very tolerant taxa were defined as those found in at least 10 sites and for which >= 50% of their occurrences were in the poor range of the HDG. In some cases, multiple species within a genus failed to meet the criteria for inclusion as sensitive (or very tolerant) because they were not present at enough sites. If several species within a genus (or family) were consistently found at good (or poor) sites, I combined the species into the parent genus (or family) and evaluated the genus (or family).

For the BioRecon data, I only tested metrics used in the SCI in order to make the indexes as similar as possible. Relative abundance metrics could not be calculated for these data; therefore, only taxa richness metrics were evaluated. BioRecon metrics were tested for correlation with SCI and HDG.

Index development and testing

Two indexes were defined by selecting metrics that showed the most consistent response to HDG. One index was developed for the SCI protocol (20 dip net samples and laboratory sorting of macroinvertebrates) and another for the BioRecon protocol (4 dip net samples and field sorting).

For SCI, metrics were transformed into unit-less scores by determining the 95th percentile values for each metric and then dividing each metric value by its range and multiplying by 10. Metrics that declined with disturbance were scored from 0–10 and metrics that increased with disturbance were scored from 10–0 so that the best biological condition was consistently scored as a 10. For skewed metrics, the natural log was calculated before dividing by the metric range. This process was done separately for the northeast, peninsula and panhandle regions. I tested for differences in metric expectation associated with different regions by comparing the regression lines for each metric against HDG and adjusted metric scoring as necessary to insure similar responses for all metrics in the three regions.

For the BioRecon index, I scored metrics to match the SCI metric scores as closely as possible. Using histograms to evaluate the overlap of the BioRecon and SCI metrics measured at the same sites, I adjusted metric scoring for the BioRecon metrics as needed. In order to distinguish between the BioRecon and SCI, BioRecon metrics were scored from 0–1 (rather than 0–10). In this way, the index values provide a quick clue as to which index is being reported.

After correlation with disturbance was confirmed for both indexes, I evaluated different sources of variability, either due to natural sources (e.g., watershed size or season) or the sampling method (e.g., laboratory subsampling). I used correlation to test for dependence of SCI on watershed size. To test the influence of season on SCI, 78 sites with summer and winter samples were compared using a paired *t*-test. Summer was defined as May through October and winter as November through April. Within each season, data for repeat visits to each site were averaged.

To evaluate other sources of variance, such as site differences, year differences, and repeat visits, I used an ANOVA model to estimate variance components. For SCI and BioRecon I estimated variance associated with site, year, site x year interaction, and error. Error was defined as variance associated with repeat visits to the same site within a year. For SCI data, most of the repeat visits within a year occurred on the same day rather

than at different times during the year. For BioRecon, most of the repeat visits were on different days within the same or a different year; very few were same day.

Error estimates from ANOVA (mean squared error) were used to calculate the 90% confidence interval for both indexes and the number of categories of biological condition that each index can reliably detect (Zar, 1984). Confidence limits were calculated as:

$$\text{SCI} \pm \left(\sqrt{\frac{s^2}{n}} * 1.645 \right),$$

where s^2 = error variance estimated from ANOVA (mean squared error), and
n = number of samples taken from the site.

To calculate the number of categories each index could detect I divided its range (0–100 for SCI and 0–10 for BioRecon) by the confidence interval. Other statistical approaches could be used to calculate the number of categories of biological condition that the indexes can reliably detect. For example, a 90% confidence interval could also be constructed based on the *difference* of SCI values for two sites. This approach would yield a smaller confidence interval because instead of n=1 in the equation above, n would equal 2, because two SCI values would be involved in calculating the difference. Currently there are no standards for which statistical model to use; therefore, I used confidence intervals around SCI for a single site because they are conceptually simple and somewhat conservative from a statistical point of view. In other words, this approach is less likely to be controversial (Johnson, 1999; Hoenig and Heisey, 2001).

I also used an ANOVA model to estimate variability associated with laboratory subsampling. For this model, data from 59 sites with three subsamples each were used to estimate site variance and error variance. For this model, error was defined as the differences associated with subsamples from the same site visit.

Statistical model assumptions

The method described here for calculating confidence intervals assumes that the multimetric indexes are normally distributed. Other studies have evaluated the statistical properties of multimetric indexes and found that multimetric indexes for fish and invertebrates are unimodal and symmetric (Fore et al., 1994; Doberstein et al., 2000; Blocksom et al., 2003); thus, an underlying normal distribution is an appropriate assumption for multimetric indexes. A separate and additional consideration is whether to use *z*-values from the normal distribution to calculate 90% confidence limits or values

from Student's *t* distribution. For small sample sizes, the *t*-distribution should be used for calculating confidence limits because variance may be underestimated; however for large sample sizes (df > 30) the two distributions converge. I used data from 61 sites to estimate variance; therefore, *z*-values were appropriate.

The two data sets used to estimate variance for the SCI and BioRecon index were unbalanced, that is, the number of visits to each site varied for each site. In contrast, the data set used to evaluate laboratory subsampling was balanced with three subsamples for each site visit. ANOVA assumes a balanced design but is fairly robust to violations of this assumption (Milliken and Johnson, 1992). The primary consequence of an unbalanced design is related to significance testing of model effects against the error term, which was not the purpose of this analysis. Rather, the goal for these ANOVAs was to estimate an error to use in the confidence interval calculations and to evaluate the relative contributions of year and error to the metrics and indexes. For estimation of the error term, the large sample sizes mitigate any potential inaccuracy associated with unbalanced cells. Samples sizes greater than about 30 are considered large and this analysis had 61 sites for SCI and 128 for BioRecon.

RESULTS

Human disturbance gradient

The human disturbance gradient (HDG) was a consistent predictor of biological condition for two independent rounds of testing. When metrics were regressed against HDG for the two rounds of data, the slopes and intercepts of the regression lines were nearly identical indicating that measures included in the HDG yielded a reliable measure of human disturbance (Figure 1).

Correlations among the four measures of human disturbance used to define the HDG were high, indicating good agreement among the different measures (see Table 4). All the individual measures, however, were more highly correlated with the HDG, suggesting that the HDG was a better overall measure of general human disturbance. The SCI showed a similar pattern and was more highly correlated with the HDG than the individual measures of human disturbance.

Metric selection

On average, 30 insect taxa were found in each sample for the 629 site visits used to test metrics. Diptera dominated the samples with 23 taxa counted on average; of these taxa, about half were in the family Chironomidae. Of the remaining groups, three oligochaetes, two Trichoptera, one Ephemeroptera and one Plecoptera taxa were found on average.

For many metrics, the change in values from least to most disturbed sites was dramatic. I compared the 20 best sites (HDG = 0 and the lowest values for LDI_WS) to the 18 worst sites (HDG > 6) and found total taxa richness went from 33 to 19 on average. Number of Ephemeroptera taxa declined from two to zero; Trichoptera taxa from four to zero; clinger taxa from eight to zero; and sensitive taxa from nine to zero. The relative abundance of filterers declined from 22% to 3% while the relative abundance of tolerant individuals increased from 5% to 69%.

In general, I selected at least one metric from each of the categories below according to which metric was most highly correlated with HDG. If two metrics were calculated from the same set of taxa (redundant metrics), I only included one metric.

Taxonomic richness. EPT taxa richness was most highly correlated with HDG followed by Trichoptera and Ephemeroptera taxa richness (see Table 6). Both taxonomic groups disappeared from samples with high levels of disturbance (Figure 2). Correlation

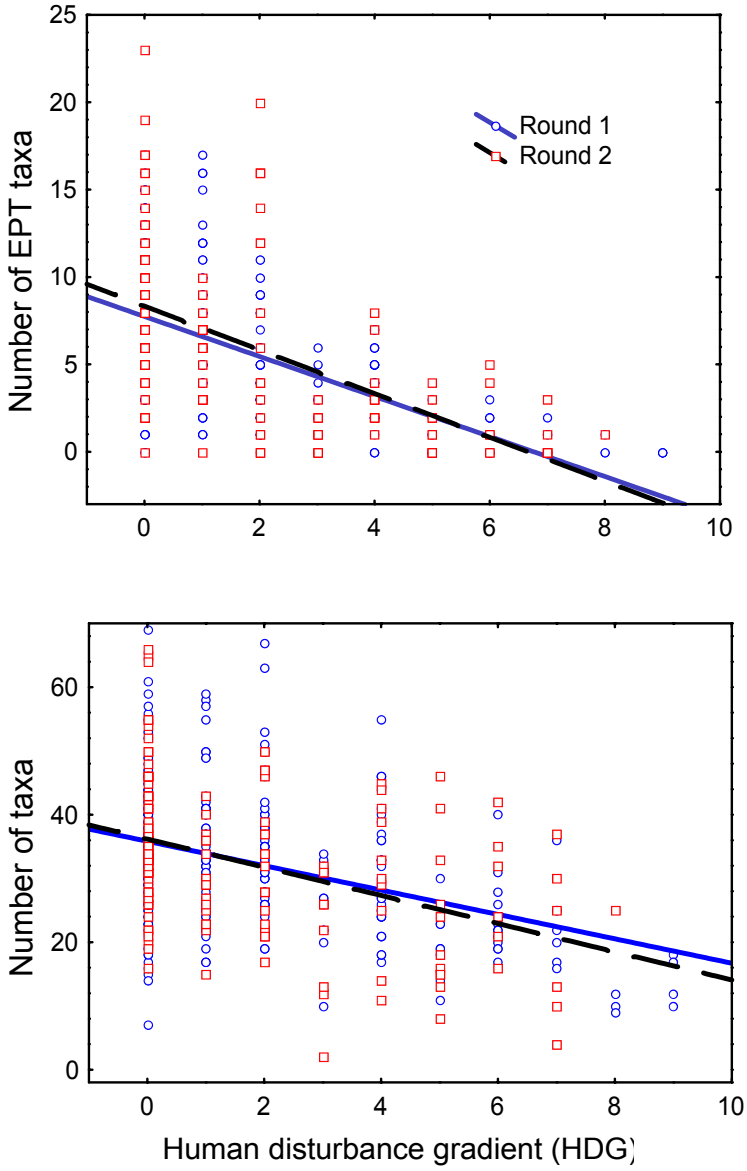

Figure 1. Number of EPT taxa and total number of taxa declined as human disturbance increased. Nearly identical regression lines for two independent sets of sites (rounds 1 and 2) indicated a consistent relationship between HDG and biological condition.

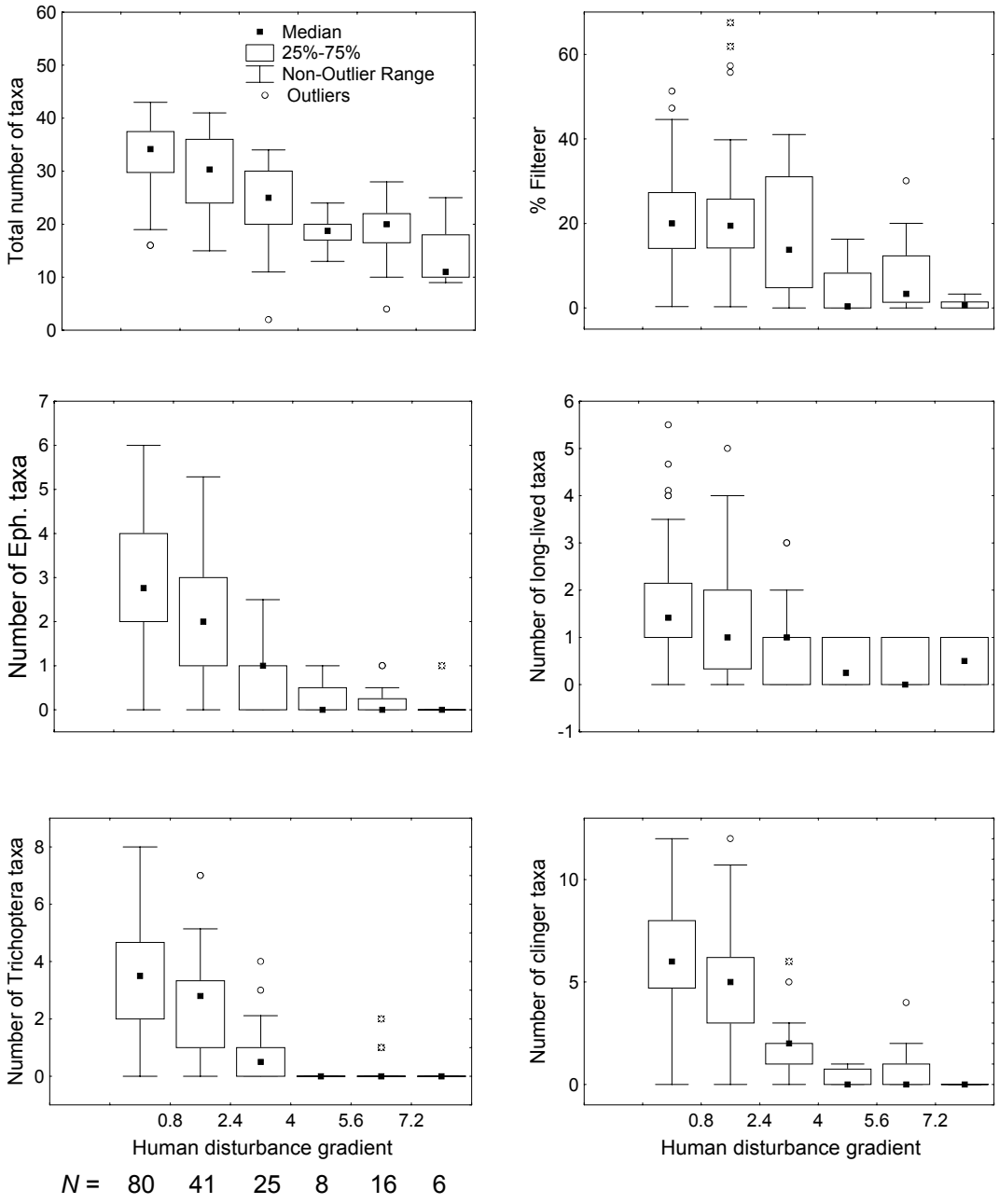

Figure 2. Six of the ten metrics included in the SCI plotted against the human disturbance gradient. All metrics declined with an increase in human disturbance. Metric values were averaged for each of the 176 sites. Samples with greater than 175 or less than 75 individuals were excluded before averaging. Number of sites within each box are shown (*N*). Plotted values are less than or equal to values on x-axis.

between HDG and Plecoptera taxa richness was somewhat lower, primarily because very few plecopterans were found in samples from the peninsula region. I selected Ephemeroptera and Trichoptera taxa richness as separate metrics because they have the potential to respond to different types of disturbance. I did not select Plecoptera taxa richness as a metric because it could only be applied in two of the three regions. Total taxa richness had the next highest correlation and was selected as an indicator of the variety of taxa that a stream site could support.

Feeding group. Relative abundance of filterers (percentage filterer individuals) had the highest correlation with HDG and the most consistent relationship with HDG when graphed. Other feeding groups failed to show the consistent decline with disturbance that percentage filterer did. I also tested this metric without *Cheumatopsyche* (a tolerant genus of Trichoptera); but the changes in correlation were very small and so all taxa were retained (Spearman's $r = -0.46$ with *Cheumatopsyche* and –0.47 without).

Voltinism. Correct information about generation times was not easily obtained for all Florida taxa; therefore, I only tested long-lived taxa richness and percentage long-lived individuals. Long-lived taxa included semi-voltine insects and non-insects that require greater than one year to complete their life cycles. Long-lived taxa richness was more highly correlated with HDG than was relative abundance of long-lived individuals; consequently, long-lived taxa richness was included in the SCI.

Habit. Clinger metrics measured either as taxa richness or as percentage of individuals were among the most highly correlated metrics with HDG. I selected clinger taxa richness for inclusion in the SCI.

Community structure. Percentage dominance increased with HDG and was included in the SCI (Figure 3); hyperabundance failed to correlate with HDG. Although percentage Ephemeroptera and Trichoptera individuals (relative abundance) were both strongly correlated with HDG, they were calculated from the same information used to calculate taxa richness of those two groups. Because the taxa richness metrics were more highly correlated with HDG, relative abundance of Ephemeroptera and Trichoptera were not included in the index. Correlation for the relative abundance of Tanytarsini midges was not quite as high as the other composition metrics (–0.45), but was included in SCI as the best available measure of the chironomid assemblage.

Sensitivity and Tolerance. Number of taxa selected as sensitive equaled ~12% of the taxa tested; and the number of very tolerant taxa was ~10% of the taxa tested. Many sensitive taxa belonged to the Ephemeroptera or Trichoptera; several chironomids were also included (Appendix B). All the Plecoptera were included though most of the genera

and species in this group were too rare to test. Considered as a group, out of 191 occurrences of Plecoptera taxa, only one was found in the 'poor' section of the HDG. All Plecoptera taxa showed a strong tendency to be found in the best of the good (minimally disturbed) sites. Many of the sensitive taxa selected by this study were previously noted by Beck (1954) and used in the Florida index; but several taxa were new to the list. Many of the taxa included on Beck's list were too uncommon to test with this data set.

The very tolerant list included one odonate (*Argia sedula*) and several chironomid taxa (*Larsia spp.*, *Cricotopus bicinctus,* and *Polypedilum illinoiense* grp.) previously noted as sensitive and included in the Florida index (Appendix C). The very tolerant list was dominated by non-insects and chironomids along with several odonate taxa. One species of tubificid worm, *Limnodrilus hoffmeisteri*, was included as very tolerant although it was present at 47% of the 'poor' sites which failed to meet the criterion of >=50% of poor sites. *L. hoffmeisteri* was included as very tolerant because it was very common at all sites. It occurred at 70 out of 85 poor sites, and tended to increase in relative abundance with disturbance.

The total number of taxa with 10 or more occurrences was 168. When a large number of tests are conducted questions arise regarding the probability of type I error. Detection of a statistically significant result will occur with the approximate frequency of *p*-value. For a *p*-value of 0.05 and 168 tests, approximately eight taxa may be selected by chance alone (0.05 * 168 = 8.4). The number of sensitive taxa falsely selected from this analysis is likely much lower because the criterion for selection was more strict than statistical significance. In other words, the *p*-values associated with 87% occurrence of more common taxa, e.g., taxa present at 20 or more sites, were much smaller.

BioRecon metrics. The six taxa richness metrics calculated from BioRecon data were significantly correlated with HDG and highly correlated with SCI (*r* ranged from 0.60–0.73; Table 7). The four SCI metrics calculated as percentages could not be tested because relative abundance was not recorded for BioRecon data. For this reason, I tested the very tolerant metric in terms of *taxa richness* rather than as a percentage. Its correlation was poor with HDG and non-significant with SCI and therefore was not included in the BioRecon index.

Multimetric index construction and evaluation

Ten metrics were selected for inclusion in the SCI and six were selected for the BioRecon index. Most of the SCI metrics required regional scoring (Table 8). In general, Taxa richness tended to be higher in the panhandle region than in the peninsula or northeast. For the SCI, the highest correlation between metrics occurred between Trichoptera and

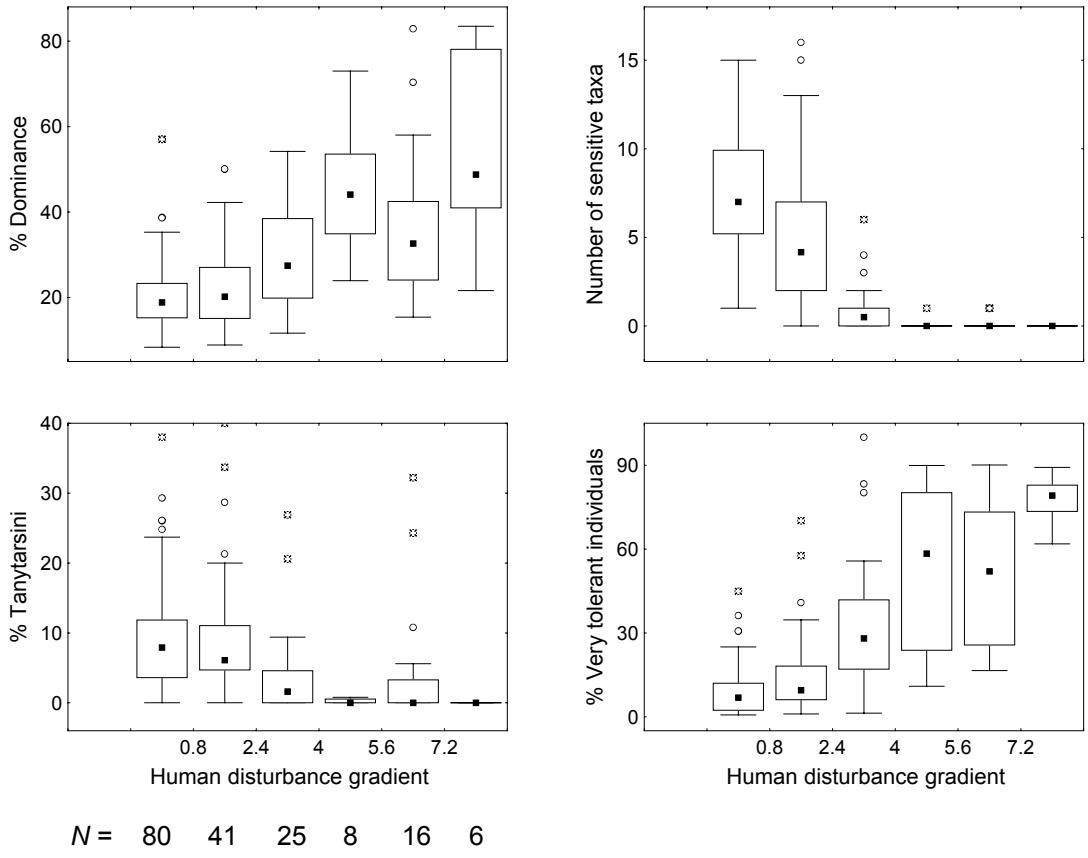

Figure 3. Four of the ten metrics included in the SCI plotted against the human disturbance gradient. Percent dominance and very tolerant individuals increased with disturbance; sensitive and Tanytarsini midge taxa declined. Metric values were averaged for each of the 176 sites. Samples with greater than 175 or less than 75 individuals were excluded before averaging. Number of sites within each box are shown (*N*).

clinger taxa (Spearman's $r = 0.77$) and clinger and sensitive taxa ($r = 0.84$). To determine whether these metrics were redundant, I evaluated the taxa included in each metric. Of the clinger taxa, approximately 41% were also sensitive taxa and 33% were also Trichoptera taxa. Thus, I concluded statistical correlation was due to an underlying association with human disturbance rather than redundant information.

Two metrics were somewhat skewed, that is, their distributions had long 'tails' with sparse values at one end of their distribution. For these metrics, I calculated the natural log of metric values before dividing by the range of the metric values. In this way, metric scores were more evenly spread between 0–10. Two SCI metrics increased with disturbance (% dominance and % very tolerant individuals), all others declined with disturbance.

BioRecon metric scoring rules differed somewhat from SCI metric scoring. Taxa richness tended to be lower in two regions while Ephemeroptera taxa richness was much higher in the panhandle (Table 9). More long-lived taxa tended to be found in all three regions for BioRecon samples.

Table 7. Correlation of BioRecon metrics and the old and new BioRecon index with HDG and SCI. Metrics included in the BioRecon (and SCI) index are noted (*). All correlations shown were statistically significant (Spearman's r, $p < 0.01$, one-sided test, n=116 for SCI, n=53 for HDG). Data from multiple visits were averaged for each site.

Metric/Index	**HDG**	**SCI**
BioRecon Index	-0.58	0.75
Total taxa *	-0.36	0.53
Ephemeroptera taxa *	-0.47	0.62
Trichoptera taxa *	-0.65	0.68
Long-lived taxa *	-0.46	0.65
Clinger taxa *	-0.48	0.67
Sensitive taxa *	-0.63	0.69
Very tolerant taxa	NS	NS
Old BioRecon Index	-0.41	0.58
EPT	-0.61	0.71
Florida Index (metric)	-0.53	0.69

Table 8. SCI metric name and range of metric values used to assign a score of 0 (indicating degraded condition) to 10 (indicating minimal disturbance) by region. Metric values higher or lower than the listed range were assigned a score of 0 or 10 as appropriate. Metric ranges listed in reverse order indicate metrics that increased (rather than declined) with disturbance. For two marked metrics (*), the natural log was calculated for the range of values before scoring.

SCI metric	Northeast	Panhandle	Peninsula
Total taxa	16–42	16–49	16–41
Ephemeroptera taxa	0–3.5	0–6	0–5
Trichoptera taxa	0–6.5	0–7	0–7
% Filterer	1–42	1–45	1–40
Long-lived taxa	0–3	0–5	0–4
Clinger taxa	0–9	0–15.5	0–8
% Dominance	54–10	43–10	54–10
% Tanytarsini *	0–26	0–26	0–26
Sensitive taxa	0–11	0–19	0–9
% Very tolerant *	78–0	36–0	59–0

Table 9. BioRecon metric name and range of metric values used to assign a score of 0 (indicating degraded condition) to 1 (indicating minimal disturbance) by region. Metric values higher or lower than the listed range were assigned a score of 0 or 10 as appropriate. Bolded text indicates metric ranges that differed from those for SCI.

BioRecon metric	Northeast	Panhandle	Peninsula
Total taxa	**14–37**	16–49	**11–36**
Ephemeroptera taxa	0–3.5	**0–12**	0–5
Trichoptera taxa	0–6.5	0–7	0–7
Long-lived taxa	**0–6**	**0–10**	**0–7**
Clinger taxa	**0–7**	0–15.5	0–8
Sensitive taxa	0–11	0–19	0–9

SCI was highly correlated with HDG with minimal overlap between values for extremely disturbed and minimally disturbed sites (Figure 4). The BioRecon index was also highly correlated with HDG and was more responsive across the range of human disturbance than the original BioRecon index (Figure 5).

Regional scoring at the metric level resulted in an SCI that showed a similar response to human disturbance within each region. The regression lines for SCI vs. HDG for the northeast and peninsula regions were nearly identical; the line for the panhandle region diverged somewhat from the other two regions for more degraded sites (Figure 6; Appendix D). The divergence was associated with two sites in the panhandle with SCI values that were higher than expected. I did not alter the metric scoring rules because several other panhandle sites did obtain the lowest SCI values for high levels of disturbance.

I calculated the natural log of the area upstream of the sampling site to test for correlation between SCI and watershed size. Watershed size had an underlying correlation with HDG (see Table 4) because the most disturbed sites tended to have small watersheds. To avoid this confounding association, I divided the sites into three groups (low, medium and high disturbance) before testing for correlation between SCI and watershed size. The low disturbance group of sites had HDG values equal to zero. I defined medium disturbance as HDG values from 2–5 and high disturbance as HDG values >5. SCI was not significantly associated with watershed size for any of the three levels of disturbance (Figure 7; Spearman's $r = –0.01, –0.005, 0.12$, $p > 0.5$).

The relationship between SCI and HDG was consistent across years as well. When site visits were divided into two-year segments for 10 years of sampling (1992-2001) and SCI regressed against HDG, the five regression lines were nearly identical (Figure 8). This result indicates that the pattern of association between SCI and HDG was independent of year.

To verify that the SCI was correlated with HDG, I used an independent set of 23 sites not used in any of the previous analyses. Correlation was high (Spearman's $r = –0.81$, $p << 0.001$; Figure 9). Two sites with >300 individuals in the samples had somewhat higher SCI values than predicted by HDG, possibly due to inflation of taxa richness measures associated with large samples.

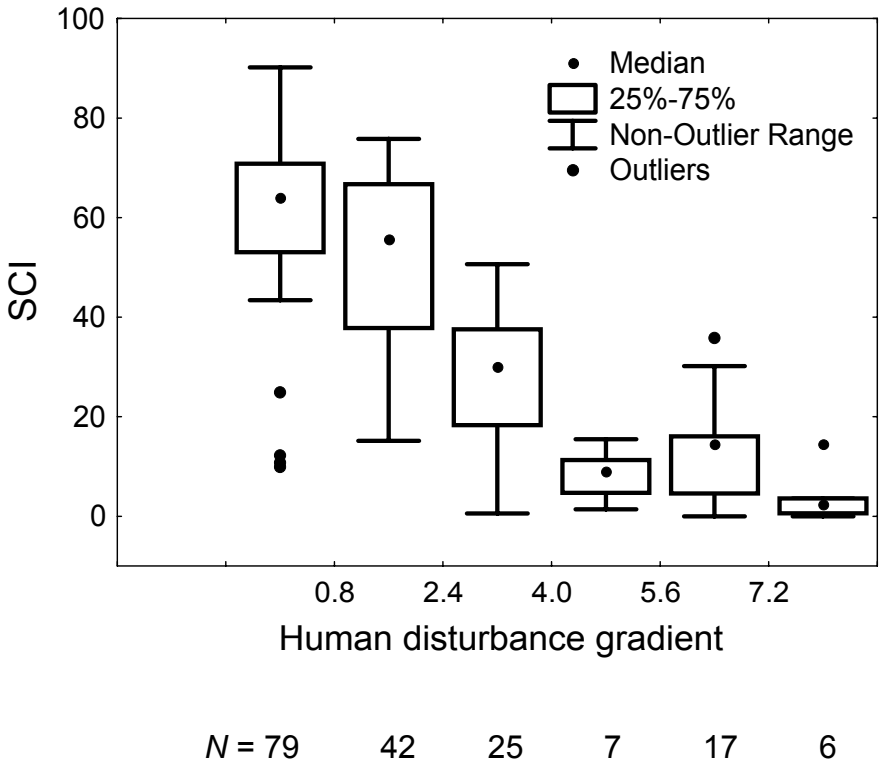

Figure 4. SCI declined as human disturbance increased. Index values were averaged for each of the 176 sites. Only samples with 75-175 individuals are shown. Number of sites in each box are shown (*N*).

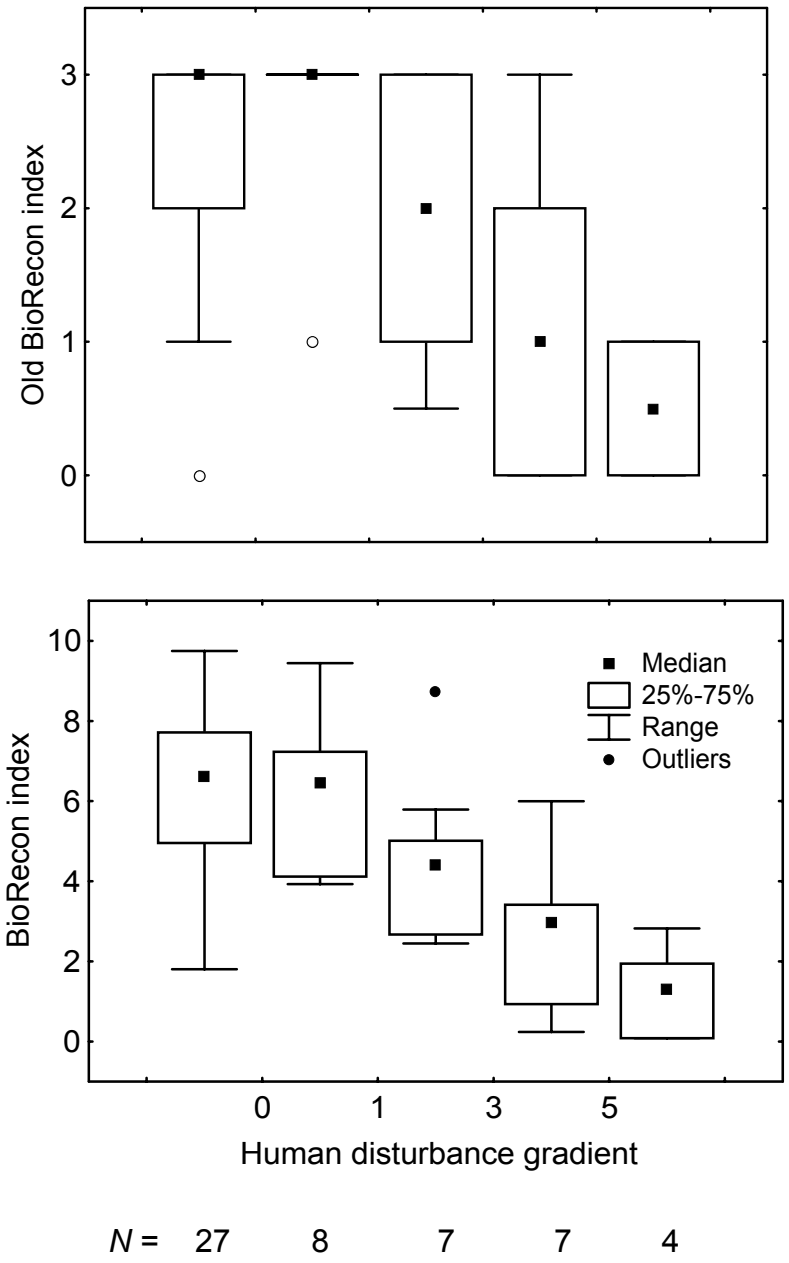

Figure 5. Original BioRecon index based on three metrics (total taxa, EPT taxa, and Florida index; upper panel) and current BioRecon index based on six metrics (lower panel). Both indexes declined as human disturbance increased, but the 6-metric index has a broader range of response to disturbance. Values in boxes are <= value on x-axis.

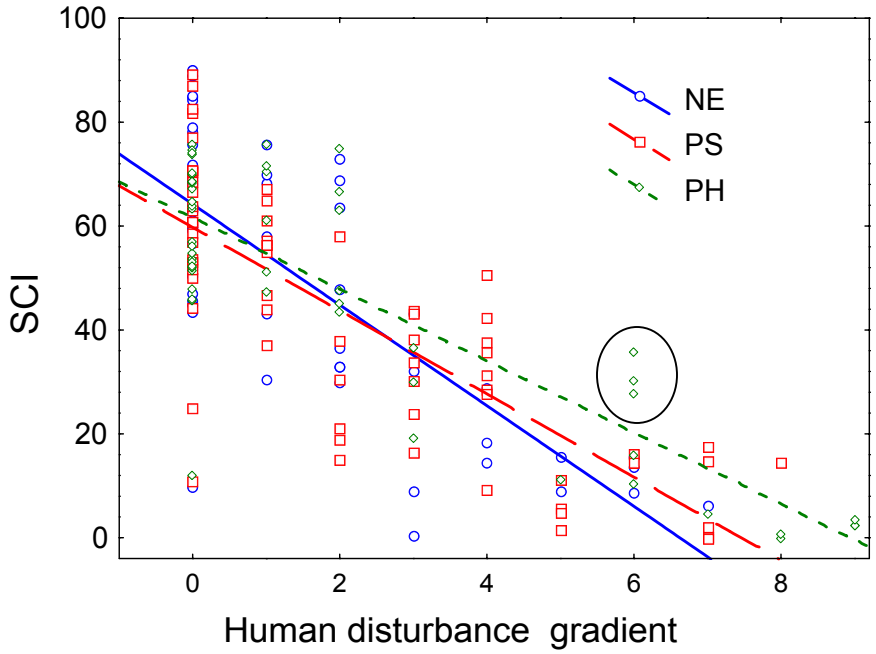

Figure 6. SCI declined as human disturbance increased. Regression lines for the peninsula and northeast had nearly identical slopes indicating a similar index response across the gradient. Three visits to two sites with higher values than predicted (circled) increased the slope for panhandle sites somewhat. Index values were averaged for each of the 176 sites. Only samples with 75-175 individuals are shown and were included in the regression. $N = 58$ (NE), 66 (PS), and 52 (PH).

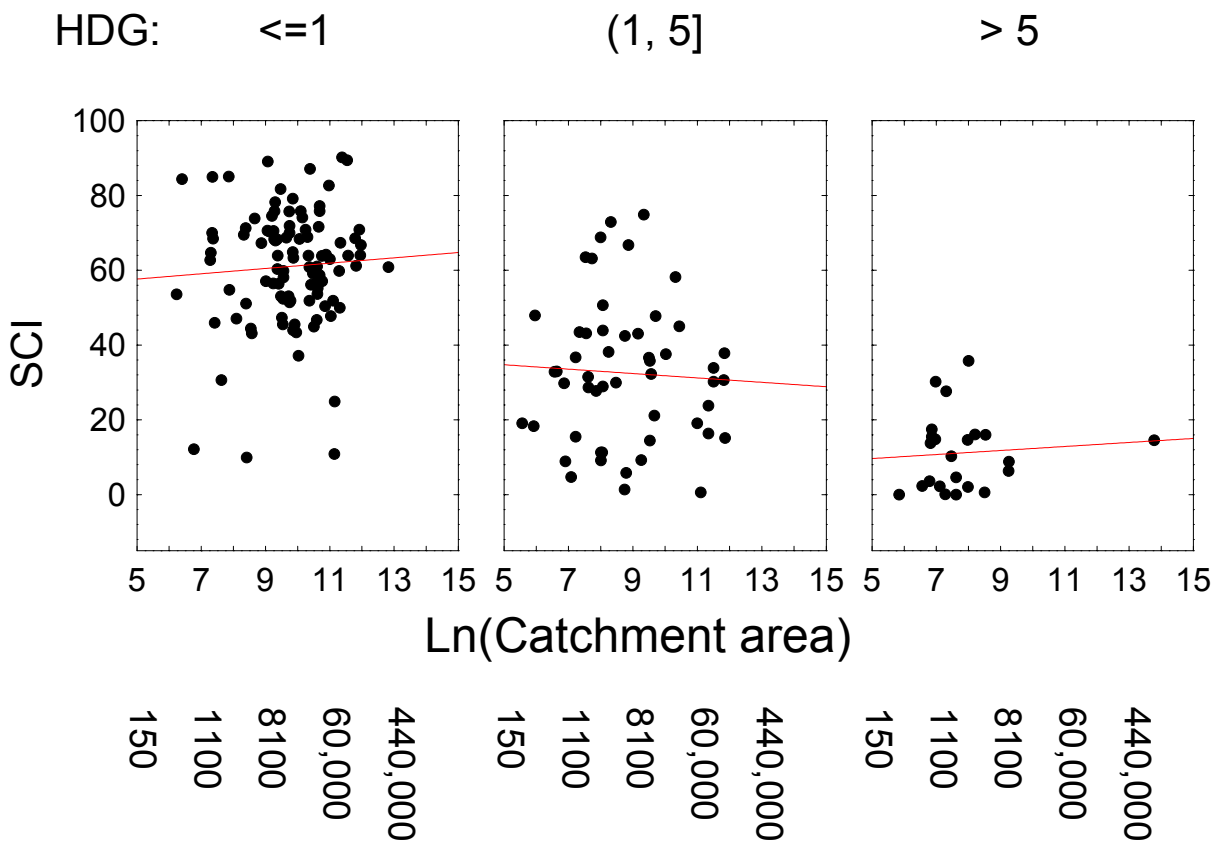

Figure 7. SCI was not significantly associated with watershed size (catchment area) when correlation was tested separately for low, moderate, and high levels of human disturbance (HDG). Highly disturbed sites tended to be small in catchment area and this relationship created a spurious correlation between SCI and catchment size. Shown are average SCI values for 176 sites. Actual catchment area shown below figures (in acres).

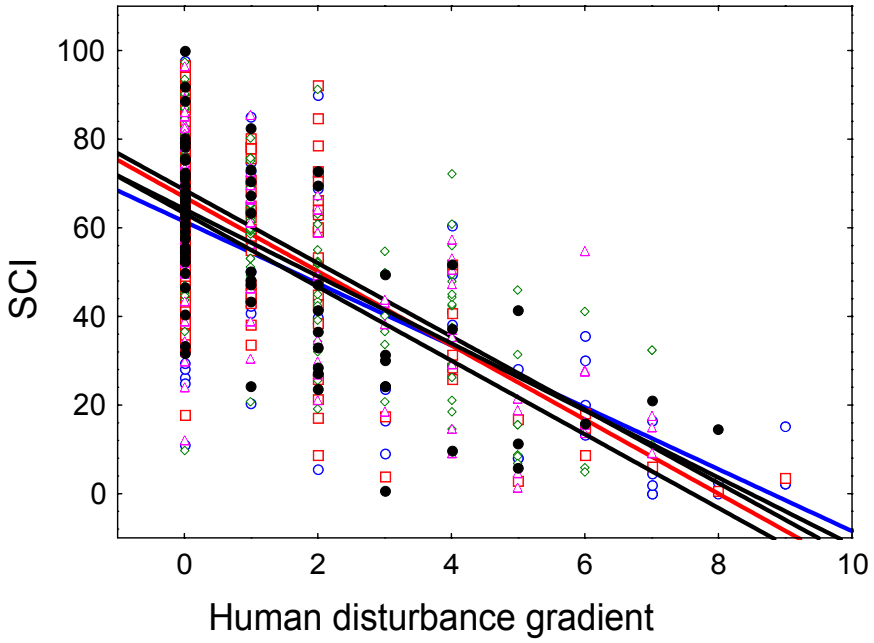

Figure 8. SCI showed a similar response to human disturbance for different years. Shown are five regression lines representing data from 1992-3, 1994-5, 1996-7, 1998-9, and 2000-1. The nearly identical slope lines indicate a consistent relationship between SCI and human disturbance for different time periods.

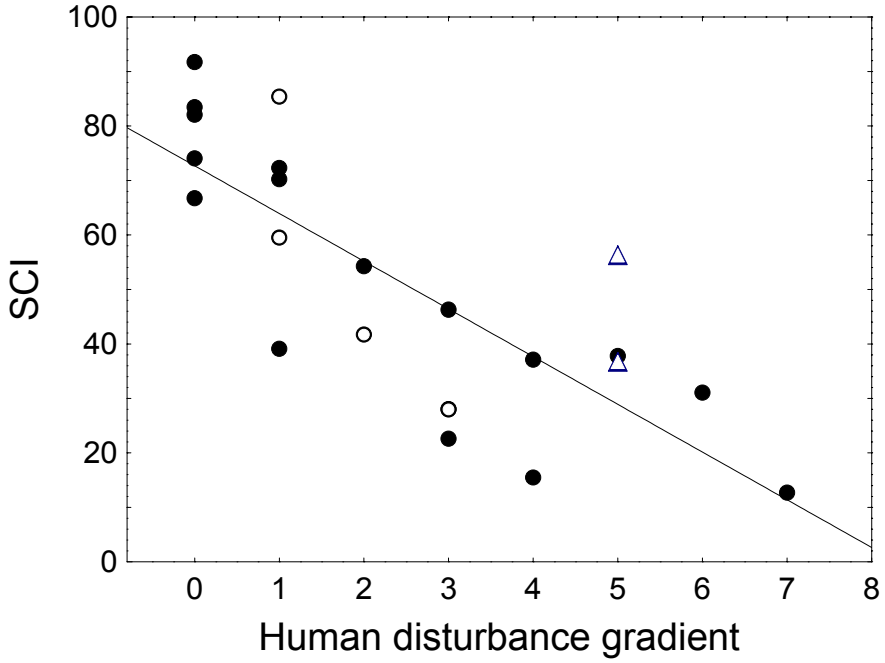

Figure 9. Correlation between SCI and the human disturbance gradient was high for an independent test of 23 sites not included in any previous analyses (Spearman's r = -0.81, $p << 0.01$). Four sites were missing data for NH3 (open circles) and two sites had >300 individuals (open triangles).

Statistical precision of indexes

The target number of individuals for laboratory identification for the SCI is 95–115. For the metric testing data set, only 27% of the 629 site visits met this target. Most samples had well over 100 individuals. For these data, 32% of the samples had >175 individuals, with some samples ranging up to hundreds of individuals. At the other end of the scale, only 36 samples (6%) had less than 90 individuals in the entire sample.

When I compared small sample sizes with standard (or large) sample sizes from the same sites, samples with <75 individuals contributed a disproportionately large amount of variability to the repeat visits (Figure 10). SCI values were typically much lower for these small samples. For this reason, I excluded these 17 samples from the analysis of variability which reduced the number of sites from 62 to 61.

Repeat visits to the same sites were used to evaluate sources of variability for SCI and BioRecon. Sources of variance included differences associated with sites, years, visits within years, season (summer vs. winter), and laboratory subsampling.

From the ANOVA analysis, the variance associated with sites was greatest (304.9, 71%), variance associated with year was relatively small (7.7, 2%), and variance components associated with site x year interaction and error were similar (54.6, 12% and 66.6, 15%; Figure 11; see Appendix D). For this model, error was defined as repeat visits within years, about half of which were on the same day.

I used the estimate of error variance from the ANOVA to estimate the 90% confidence limits around SCI as +/– 13.4 points, or a 27-point spread. Graphic representation of this range around the observed values for repeat visits captured all visits to the least disturbed sites, that is, sites with SCI values >73, but failed to include all the visits to sites with lower SCI values (see Figure 10). These sites graphically illustrate the site x year variance, that is, certain site-year combinations that changed in unique ways. The fact that undisturbed sites did not show this pattern while disturbed sites did, suggests that SCI was measuring real changes in biological condition rather than ‘noise’ associated with natural variability.

Using the length of the 90% confidence interval (27 points) to divide the range of the SCI (0–100) yielded 3.7 categories of biological condition. Calculating the confidence interval for two site visits yielded a confidence limit of +/– 9.5, or a length of 19 points. This confidence length translated into approximately five categories of biological condition based on the average SCI value for two site visits. Current FDEP protocol specifies that the two most recent visits be used for assessment. I calculated

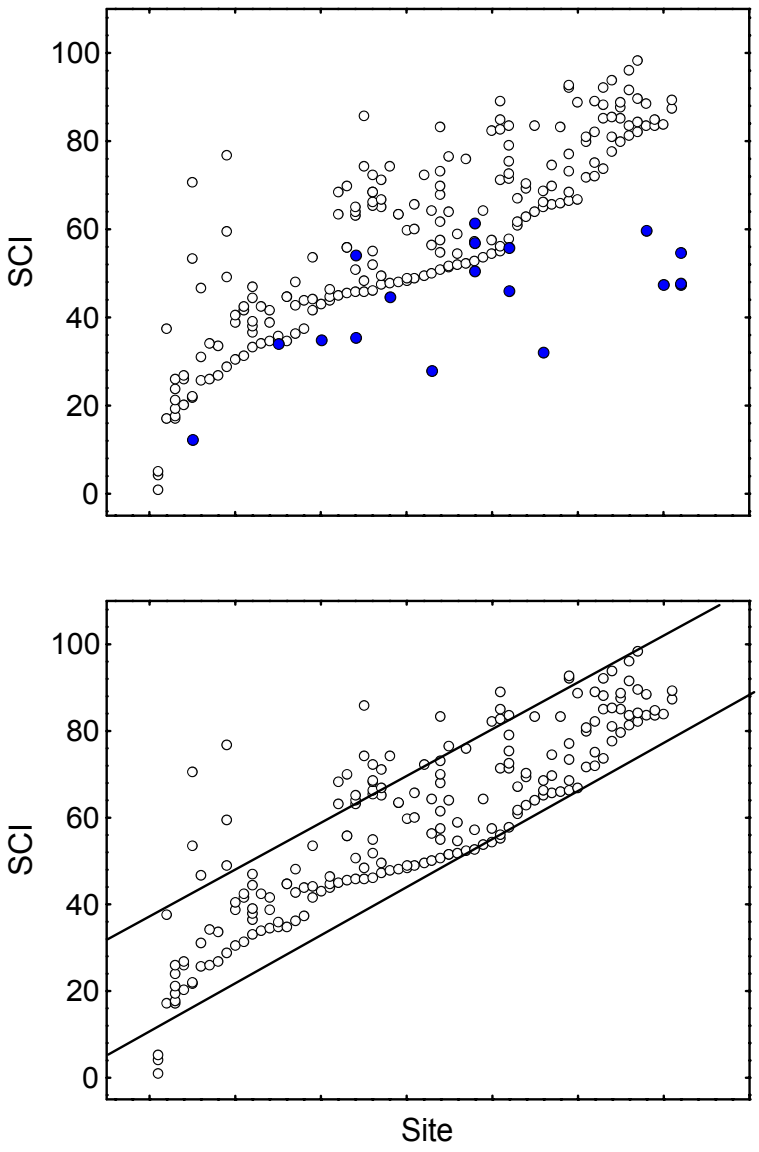

Figure 10. SCI values for repeat visits to 62 sites are shown. Sites were first sorted according to SCI value. Individual site names are not shown, but all SCI values for each site are plotted at a single point on the x-axis. Thus, vertical lines of points represent the range of SCI values at each site. Samples with < 75 individuals are indicated by solid points (upper panel). Lines (lower panel) represent the +/- 15 point confidence interval for the SCI when small samples were eliminated.

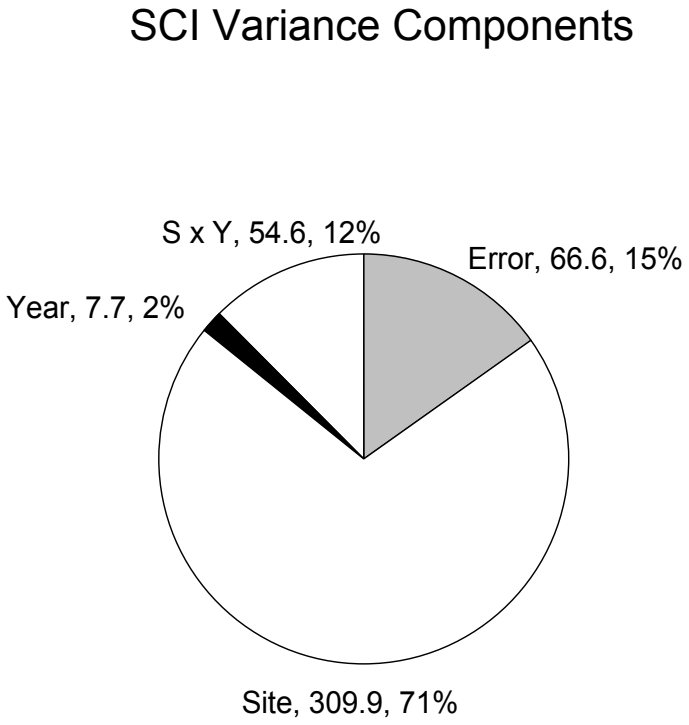

Figure 11. Estimate of variance components and their relative percentage for SCI. The largest percentage of variance was associated with site differences. Replicates collected within a single year (primarily on the same day) accounted for approximately 15% of the total observed variance. Year differences accounted for 4% of the variance. The interaction of site and year accounted for 16% of the total variance.

average metric values associated with these SCI intervals to summarize the biological condition within each category (Table 10). I used these average values to create narrative descriptions for the numeric categories of SCI (Table 11).

Variance components calculated for the SCI component metrics showed similar patterns to the SCI. Site differences and site x year variance typically contributed the largest percentages to the overall variance (Figure 12). The exception was long-lived taxa richness that had a large error component (with error defined as variance associated with repeat within year visits – as above). The least variable metrics were Ephemeroptera, clinger and sensitive taxa richness, and % very tolerant. Overall, the SCI was less variable than most of its component metrics.

For the seasonal comparison, sites were chosen to represent locations with minimal human disturbance in order to eliminate this potentially confounding source of variability from the analysis. Unfortunately, over half the SCI values for these site visits were <70 indicating moderate disturbance. Thus, the seasonal data set represented a broad range of conditions with the exception of the highly degraded sites (SCI <20) which were not included. SCI values were on average 3.5 points higher in winter than summer (paired *t*-test, $p = 0.049$). The observed difference was greater in the northeast (6.1 points) than the panhandle (1.6 points) or peninsula (3.6 points; Figure 13). Using the same data, I evaluated metrics to determine which were contributing to the seasonal differences in SCI. Five metrics had values in the winter indicating better biological condition; these were Ephemeroptera taxa richness, % filterers, % Tanytarsini, sensitive taxa richness, and % very tolerant. Long-lived taxa richness had higher values in the summer, opposite of SCI. Although significant, most metric differences were rather small compared to the overall range of the metric (Table 12). The paired *t*-test is a very powerful test and can detect very small differences because it compares each site to itself. The exception was % Tanytarsini. With a difference of 4.6% compared to a possible range of 0–26%, it had the largest relative change from summer to winter. I did not modify metric scoring to correct for seasonal differences because most metric differences were very small and the difference between SCI summer and winter values was small (3.5.%) relative to the site differences through time.

Table 10. Average metric values for ranges of SCI values corresponding to good, fair, poor and very poor condition. Note that metric values have not been corrected for different expectations associated with different regions.

Metric	Good [73-100]	Fair [46-73)	Poor [19-46)	Very Poor [0–19)
Total taxa	41	35	29	18
Ephemeroptera taxa	4	3	1	0
Trichoptera taxa	6	4	1	0
% Filterer	27	22	14	5
Long-lived taxa	3	2	1	0
Clinger taxa	10	6	2	1
% Dominance	17	21	30	46
%Tanytarsini	12	10	7	1
Sensitive taxa	10	6	2	0
% Very tolerant	4	9	24	60

Table 11. Category names, ranges of values for SCI, and example descriptions of biological conditions typically found for that category. Range of SCI values represent 90% confidence intervals for one or two samples. Square brackets indicate a value is included in the range; round brackets indicate a listed value is not included. Narrative metric descriptions are not used to score metrics, rather they describe values associated with a range of index values.

SCI category	SCI range	Description
1 sample		
Good	[73–100]	Similar to natural conditions, up to 10% loss of taxa expected
Fair	[46–73)	Significantly different from natural conditions; 20–30% loss of Ephemeroptera, Trichoptera and long-lived taxa; 40% loss of clinger and sensitive taxa; percentage of very tolerant individuals doubles
Poor	[19–46)	Very different from natural conditions; 30% loss of total taxa; Ephemeroptera, Trichoptera, long-lived, clinger and sensitive taxa uncommon or rare; Filterer and Tanytarsini individuals decline by half; 25% of individuals are very tolerant
Very poor	[0–19)	Extremely degraded; 50% loss of expected taxa; Ephemeroptera, Trichoptera, long-lived, clinger, and sensitive taxa missing or rare; 60% of individuals are very tolerant
2 samples		
Excellent	[81–100]	Proportion and abundance of taxa similar to natural conditions; minimal loss of taxa
Good	[62–81)	Similar to natural conditions with up to 10% loss of taxa; 25% loss of Ephemeroptera, Trichoptera, clinger and sensitive taxa
Fair	[43–62)	25% loss of total taxa; 50% loss of Ephemeroptera, Trichoptera, clinger, and sensitive taxa; 33% loss of long-lived taxa
Poor	[24–43)	High percentage of individuals present belong to very tolerant taxa; only tolerant Ephemeroptera, Trichoptera, and clinger taxa present; one sensitive or long-lived taxon may be present
Very poor	[0–24)	Extremely degraded; 50% loss of expected taxa; Ephemeroptera, Trichoptera, long-lived, clinger, and sensitive taxa missing or rare; 60% of individuals are very tolerant

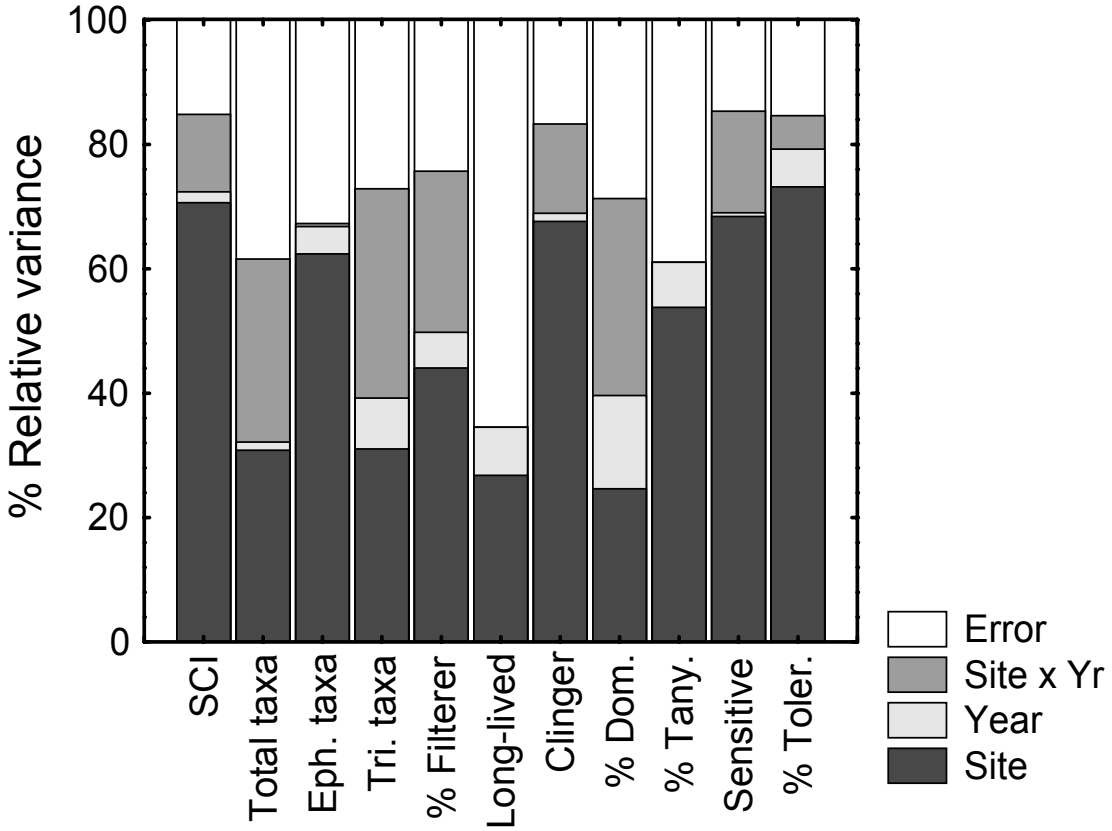

Figure 12. Variance components for SCI and its metrics. For the index and several of the metrics, variance associated with site differences represented the largest component of variance. Error variance associated with repeat visits within years (primarily on the same day) was highest for long-lived taxa richness. The least variable metrics were Ephemeroptera and sensitive taxa richness, and percent very tolerant individuals.

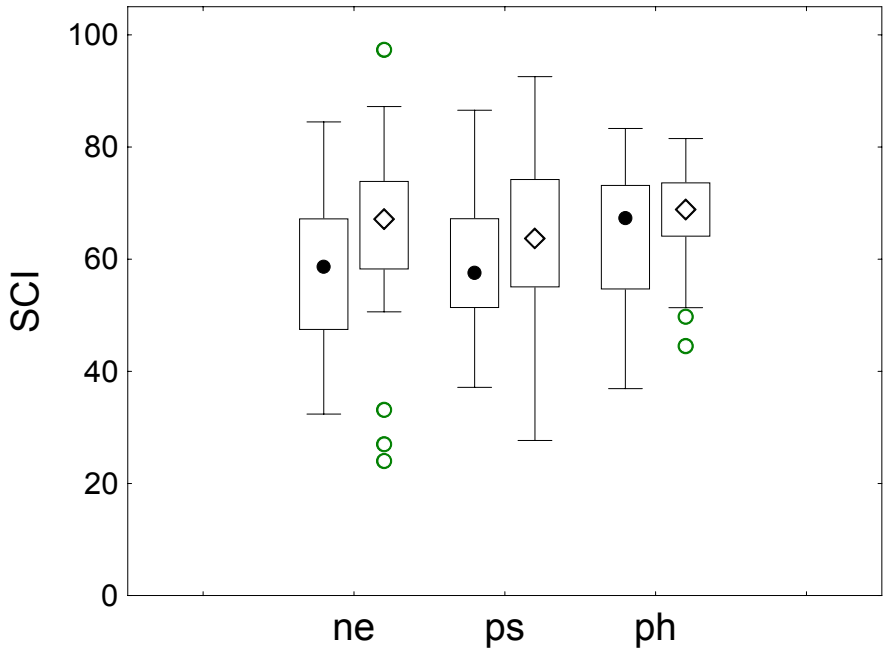

Figure 13. SCI values by region (northeast, peninsula, and panhandle) and season. Summer samples (circles in center of box) tended to be lower than winter samples (diamond centers) in the northeast and peninsula regions. *N* = 23 (NE), 34 (PS) and 31 (PH).

To evaluate the influence of laboratory subsampling on SCI, I used data that included multiple subsamples from the same site visit. I estimated variance using ANOVA with stream site-visit (n = 59) as a single factor and the three subsamples from each visit to estimate the error variance. Error variance for this model was 32.86 which represents about half the error variance calculated above when error was defined as differences associated with repeat visits to the same site (32.86/66.6 * 100 = 49%). In other words, of the variance associated with repeat visits to the same site, about half the variability is due to laboratory subsampling. However, because the variance estimates were from different data sets, the comparison should be cautiously applied.

The BioRecon index was calculated as the sum of the six scored metrics; therefore, the sum of the scores ranged from 0 to 6. I transformed the index to a range of 0-10 by dividing the sum by six. The data set for this ANOVA model was unbalanced with many site-year combinations missing; therefore, the site x year interaction component of variance could not be estimated. Components of variance were similar to those for the SCI in that the largest component of variance was associated with site differences and year variance was again small (Figure 14). Error variance for this model was 1.46 with error derived from repeat visits to the same sites. A 90% confidence interval for the BioRecon index yielded a limit of +/-2 or a length of 4.0 points. This translated into 2.5 categories of biological condition that BioRecon could reliably detect (Table 13). For two visits to the same site, 3.5 categories could be detected.

Table 12. Results for seasonal comparison of SCI and its component metrics. Differences between summer and winter samples (negative value indicates higher winter values). Significant differences noted (* $p < 0.05$, paired *t*-test).

Index/metric	Difference
SCI	–3.5 *
Total taxa	0.1
Ephemeroptera taxa	–0.4 *
Trichoptera taxa	0.3
% Filterer	–5.2 *
Long-lived taxa	0.6 *
Clinger taxa	–0.2
% Dominance	–1.9
% Tanytarsini	–4.6 *
Sensitive taxa	–0.9 *
% Very tolerant	2.0 *

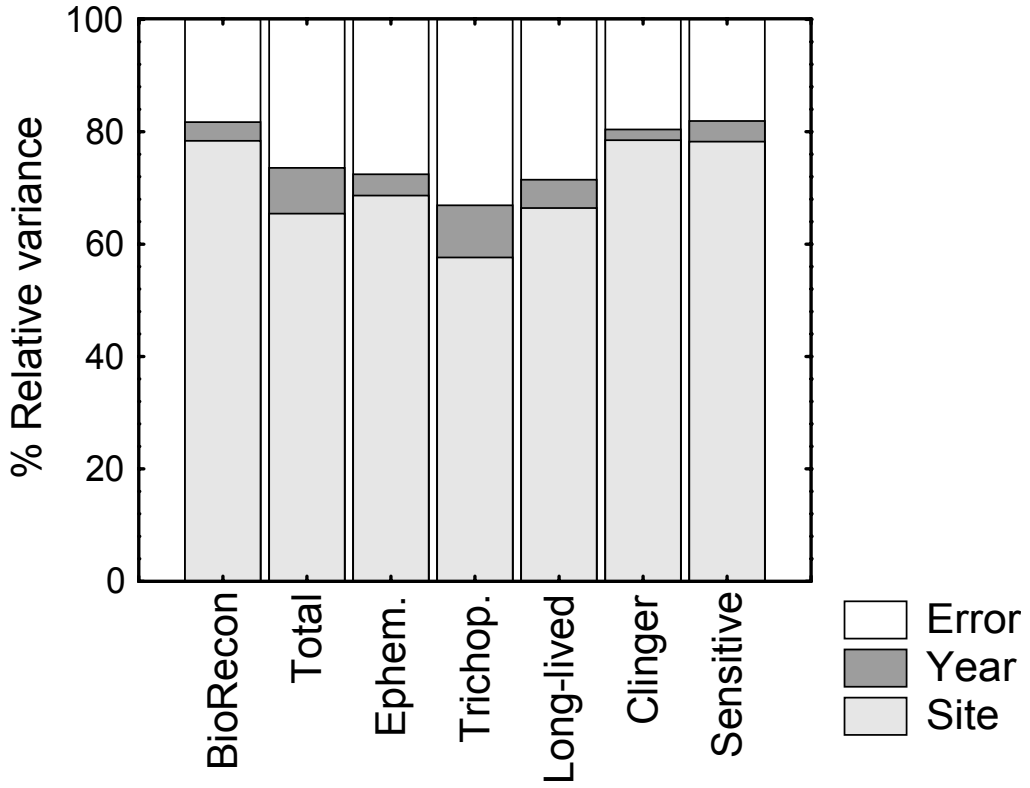

Figure 14. Variance components for BioRecon and its component taxa richness metrics. Variance associated with site differences represented the largest component for metrics and index. Error variance associated with repeat visits within years was similar across metrics ranging from 20-30% of the total. Variability associated with a specific year was relatively small for the index and metrics.

Table 13. Categorical descriptions and range of index values for the BioRecon index. Range of values represent 90% confidence intervals for one or two samples; variance estimated from repeat visits to 128 stream sites. Square brackets indicate a value is included in the range; round brackets indicate a value is not included.

BioRecon	**Index range**
1 sample	
Pass	[6–10]
Fail	[0–6)
2 samples	
Good	[7–10]
Fair	[4–7)
Poor	[0–4)

These categories for BioRecon may be somewhat conservative, particularly for sites with good biological condition, because the variability estimated from repeat visits was much higher for moderately disturbed sites than for sites with minimal disturbance (Figure 15). If variance were estimated using only minimally disturbed reference sites, the error variance would be smaller and the categories corresponding to 'passing' condition (1 sample) or 'good' condition (2 samples) would be more narrow, that is, more restrictive. To determine whether high variability was associated with natural variability or changes in human activities, I selected the sites with the most variable BioRecon values through time and asked regional biologists to note any change in human activities at the site (Table 14). Four of the sites improved where roads had been paved to eliminate sediment run off. Other changes were associated with pesticide spraying, elimination of run off from a waste water treatment plant, road closure, and the end of fertilization upstream. From this anecdotal analysis suggest that much of the BioRecon variability was associated with real changes in human disturbance rather than natural variability.

Variance results for the BioRecon components metrics were similar to those for the SCI metrics with site differences contributing the largest component of the total variance (see Figure 14). In general, the metric variance due to nuisance sources of variability (year or error) was smaller for three of the BioRecon metrics than for the same SCI metrics. Lower error variance for total taxa, Trichoptera, and long-lived taxa richness suggests that the BioRecon protocol may provide more reliable estimates of these metrics. Direct comparisons of results from different data sets should be cautiously applied, particularly for relative variance estimates, because they depend on the magnitude of the site differences.

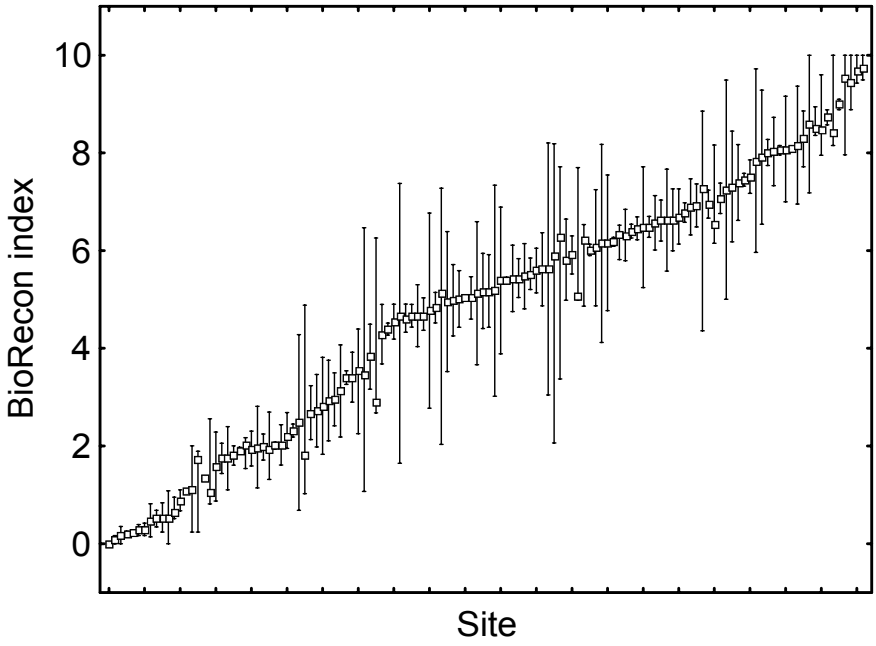

Figure 15. Range of BioRecon index values for repeat visits to 128 sites. Site names are not shown; however, each vertical line represents a single site with the minimum and maximum index values plotted as the endpoints. Note that high and low index values were less variable for repeat visits through time.

Table 14. STORET site code, region, station nickname, observed change in BioRecon values through time and possible reason for change. Listed sites represent locations with the highest variability observed for BioRecon values.

STORET	Region	Station	Change	Reason
20010521	ps	JIMREF	Decline	Unknown
20010525	ps	LILHAW@40	Decline	Unknown
20030341	ne	CECFLD7	Improve	Unknown
32030024	ph	SFBEARREF	Improve	Intermittent pesticide spraying
33020064	ph	PBRNCSTLR	Improve	319 NPS restoration project to pave a road and stop sediment
33020065	ph	BRSTWRKB	Improve	319 NPS restoration project to pave a road and stop sediment
33020067	ph	CNOECPBRNR	Improve	319 NPS restoration project to pave a road and stop sediment
33020082	ph	SANHOLTST	Improve	319 NPS restoration project to pave a road and stop sediment
33030039	ph	TRKLNHDCM	Improve	Partnership with State Forest and TNC closed logging road and eliminated sediment
33030102	ph	TURKEYCR	Improve	Summer aerial pesticide spraying
33040016	ph	WILLIAMTST	Improve	Runoff from WWTP spray field eliminated
33040017	ph	DDFLCWC189	Improve	Ended fertilization upstream of sites in State Forest
24030013	ps	HILSTP4REF	Improve	Unknown
22020062	ph	OKLREF	Variable	Unknown
27010050	ps	MOSESUS1	Variable	Unknown
27010583	ps	LITLTOMOKA	Variable	Unknown

DISCUSSION

The SCI and BioRecon indexes along with their component metrics were developed to assess the biological condition of stream sites using samples from the stream macroinvertebrate assemblage. These assessment tools will be used primarily to define biological criteria for freshwater streams and to evaluate the effectiveness of best management practices (Vowell, 2001). Three results support the use of the SCI for this purpose. First, the component metrics showed a strong and consistent response to an independent measure of human disturbance (HDG). SCI was also highly correlated with HDG using an independent data set for verification. Second, the SCI was independent of watershed size and geographic region. This result means that the SCI can be used across the state to compare stream sites. Third, the influence of seasonal and annual differences on SCI was relatively small. Index values were about 3.5% higher for winter samples and about 2% of the total variance of SCI was associated with year of sampling. Although low, these sources of variance should continue to be evaluated with more complete sampling designs because the influence of both could potentially be eliminated by adjusting index values according to season or year.

Human disturbance gradient

The human disturbance gradient simplified several aspects of metric and index testing. First, it provided a more stringent test of metric association with disturbance than a comparison of reference vs. test sites. Percentage Diptera individuals serves as an example of this. This metric was included in the old SCI because it increased at disturbed sites; however, across a gradient of disturbance it failed to consistently respond. Second, HDG was independent of biological condition because no measures of biological condition were included. Therefore, metric expectations could be defined for equivalent levels of disturbance within each region. As a consequence, it was not necessary to correct the multimetric index for regional differences. Third, objective selection criteria based on association with the HDG were used to define individual taxa as sensitive or very tolerant. Finally, the HDG resolved a spurious correlation between watershed size and SCI. During initial analysis, the SCI appeared to be correlated with watershed size because the most disturbed sites happened to be very small. After controlling for HDG, the correlation between watershed size and SCI was eliminated.

The HDG can also be used to describe the intensity of disturbance associated with specific biological losses as measured by SCI and its component metrics. EPA is currently developing guidelines for defining categories of human disturbance and

matching them with their associated biological condition; this effort is generally referred to as “Tiered aquatic life uses.” Values for HDG and SCI will be used to define and describe these categories for Florida.

Biological indicators

Multimetric indexes strive to integrate measures from a diverse set of biological categories for two reasons. First, monitoring different aspects of the biological assemblage improves the likelihood of detecting changes associated with different types of disturbance. Second, the potential exists to define metric signatures, that is, associations between specific metrics (or suites of metrics) that correspond to specific human activities (Norton et al., 2000; Yoder and DeShon, 2002). Metric signatures are particularly relevant to the TMDL process that allocates responsibility for degraded stream condition among the various human activities in the watershed (Karr and Yoder, *in press*; EPA, 2000).

Though not specific to any particular group, total taxa richness represents a general measure of the biological complexity found at a site. This metric is one of the most widely applied in biomonitoring programs because of its consistent decline with human disturbance for stream invertebrates (Kerans and Karr, 1994; Fore et al., 1996; Karr and Chu, 1999; Klemm et al., 2002) as well as fish (Hughes et al, 1998), terrestrial invertebrates (Kimberling et al., 2001), and birds (Bryce et al., 2002). Ephemeroptera and Trichoptera taxa richness have also been widely applied, though are often combined with Plecoptera into a single (EPT) metric. Splitting these taxonomic groups apart provides the opportunity for metric signatures associated with different types of disturbance. For example, Ephemeroptera taxa are known to be particularly sensitive to metals and will disappear before other taxa (Clements, et al., 2000; Fore, 2002). On the other hand, an increase in Ephemeroptera taxa may indicate increased nutrients (Miltner and Rankin, 1998). Filterers are expected to decline in response to disturbance because of the increase in sediment and silt that can damage or clog nets. Long-lived taxa are expected to decline as human disturbance alters the natural flow regime because these taxa require water in the channel year-round. Pollution events of short duration may also eliminate these taxa while other taxa may colonize from unaffected sites.

Clinger taxa have morphological and behavioral adaptations that allow them to cling in fast water. Human development near stream sites in Florida often translates into eroding sand that can smother habitat and eliminate these taxa. Other studies have found these taxa to be sensitive to disturbance associated with mining and urbanization (Fore et al., 2001; Mebane, 2002). Percentage dominance of the most abundant taxon increases

with disturbance as the natural taxonomic diversity declines and very tolerant taxa dominate samples. This metric represented a measure of the overall structure of the assemblage and has been associated with disturbance in several regions (Klemm et al., 2002). The Tanytarsini midges are used as indicators by Ohio because of their general sensitivity to human disturbance (DeShon, 1995). Three Tanytarsini midge taxa were also selected for inclusion in the list of sensitive taxa for this study.

Several of the sensitive taxa agreed with Beck's (1954) designations used to calculate the Florida index; however, a few reversals were noted for taxa that satisfied the criteria for very tolerant. The primary difference between the sensitive taxa list and the Florida index list was the exclusion of many rare taxa that could not be tested. The list of sensitive taxa is now shorter; nonetheless, the objective criteria used to select these taxa make these metrics easier to justify in a regulatory context.

The set of metrics selected for Florida represent a convergence with similar studies and programs in other states (Karr, 1998). Many of the Florida metrics also had a strong correlation with disturbance in Colorado, Idaho, Washington, Tennessee, and Japan where they responded to a diverse set of human activities including timber harvest (Fore et al., 1996), recreation (Karr, 1998), urbanization (Fore et al. 2001; Karr and Morishita Rossano, 2001; Morley and Karr, 2002), agriculture (Kerans and Karr, 1994), and mining (Mebane, 2001; Fore, 2002; Mebane, 2002). The emergence of a core set of metrics across a variety of geographic contexts further supports the use of these metrics as biological indicators.

Sources of variance

I evaluated five sources of nuisance variance for the SCI and two for the BioRecon index. For SCI, I estimated variance associated with laboratory subsampling, repeat visits within a year (error), different seasons, different years, and site x year interaction. For the BioRecon I estimated variance due to repeat visits within a year (error) and different years.

One of the largest contributors to the differences observed for SCI values collected from the same site within the same year, was the variability associated with laboratory subsampling. Subsampling represented approximately 49% of the variance due to repeat visits to a site within the same year. Unfortunately, the two variance estimates were derived from different data sets; thus, this percentage should be interpreted with caution. Nonetheless, subsampling in the laboratory seems to be a significant source of variability. Furthermore, the subsampling method for this small study was more carefully controlled than for routine sampling in that the target of 100 individuals per sample was

more frequently attained. Thus, the subsampling analysis did not include any additional variability associated with variable sample sizes; consequently, variance due to subsampling may actually be higher for routine samples.

Site x year interaction represents the variability associated with some unique site-year combinations. This source of variance represented index values from sites that increased (or decreased) from one year to the next in a different or more extreme manner than did other sites during those same years. These differences could be due to natural variability, e.g. the effect of a dry year on some sites but not others, or could be due to actual changes in biological condition associated with human influence. These two sources could not be distinguished using the current data because all sources of human disturbance were not quantified. For example, agriculture is a dominant land use in Florida. Pesticide treatments represent a potentially significant impact to stream invertebrates but could not be evaluated for this study. Nonetheless, the higher variability in both SCI and BioRecon index values observed for sites with moderate index values compared to sites with high index values, suggests that both indexes may be measuring real changes associated with human land use rather than nuisance variance. Other studies have found a similar pattern of increased variance associated with greater human disturbance (Karr et al., 1987; Steedman, 1988). For BioRecon, index variability for several sites was explained by changes in human land use. Road paving to eliminate erosion, elimination of toxic run off and discontinuation of fertilization were all associated with improved BioRecon values.

For SCI, winter index values were significantly higher than summer values collected from the same sites; however, the difference was small (3.5%). I did not adjust metric scores to eliminate this difference individual metric differences were somewhat small and the increase was due to a cumulative effect across several metrics. The seasonal difference was also relatively small compared to the variance due to other sources such as repeat visits within a year or laboratory subsampling. The influence of seasonality awaits a better data set that compares index values by season for sites with little or no change in human disturbance.

For the BioRecon index, the variance associated with repeat visits within years was somewhat higher than the variance observed for the SCI. This is largely because the site x year interaction term could not be estimated (due to incomplete data) and this variance was included in the error term. In addition, the BioRecon data included mostly samples from different days within the year rather than the predominantly same-day samples used for the SCI. As a consequence the variance estimate for the BioRecon index may be somewhat high and the precision of the index underestimated.

The higher variability noted for the BioRecon index may be due to human disturbance changing at the site, different levels of sampling effort applied by field biologists, the smaller number of metrics in the index (six vs. 10 for SCI), or some combination. In the field, biologists examine the sample until all unique taxa have been identified. They also bring back examples of each taxon for laboratory verification. Biologists may vary in both the amount of time they spend searching for unique taxa and in the number of taxa returned to the lab. The protocol specifies that a minimum of five examples of each unique taxon be returned for verification. If multiple examples of each taxon are returned, the probability of identifying additional taxa increases as a function of the number of individuals returned. This is because many field identifications can only be made to the genus level whereas in the lab more taxa can be identified to the species level.

CONCLUSIONS

Correlation between biological metrics and the independently derived human disturbance gradient provided strong evidence in support of the relationship between human influence and biological condition. The high correlation between SCI and HDG was verified using an independent set of data. In addition, the pattern of biological response to human disturbance was repeated across all three regions in Florida, indicating that the multimetric index was independent of geographic context. Finally, the index was consistent across years, with a nearly identical relationship between SCI and HDG for 10 years of data.

To document every source of human disturbance is prohibitively expensive, consequently the human disturbance gradient was inevitably incomplete. I had no information on metals, pesticides, dredging, or other pollutants. Aerial pesticide application is a frequent practice on agricultural lands and can be very damaging to stream invertebrates as pesticides are concentrated in rain runoff (Hutchens, et al., 1998). This lack of information may likely explain why low SCI values were observed for sites with no known source of disturbance. In contrast, sites with known disturbance never had high values for SCI. This result supports the idea that much of the lack of fit between disturbance and biological condition is largely due to measurement error associated with quantifying human disturbance rather than variability of biological measures (Bryce et al., 1999; Karr and Morishita Rossano, 2001; Fore, 2003). In short, multimetric indexes accurately synthesize the influence of human activities over time while measures of human disturbance are at best approximations of the complex ways that human activities influence the environment.

The SCI had good statistical precision and could detect 3.7 categories of biological condition based on a single sample and five categories based on two samples. The BioRecon was not as precise but had adequate precision to detect 2.5 categories of biological condition based on a single sample and 3.5 categories for two samples.

Strong correlation with an independent measure of disturbance, good statistical precision, immunity to regional differences or watershed size, and minimal sensitivity to seasonal or yearly differences all contribute to the usefulness of these indexes for biological assessment of Florida’s streams.

RECOMMENDATIONS

- *Support categories of biological condition with statistical criteria.*

Confidence intervals were used to define the number of categories of biological condition that SCI and BioRecon can detect for a given level of confidence. Broader categories can be used to define impairment, but smaller index ranges should not unless a different statistical model or test is used. The metric values associated with index categories can be translated into narrative descriptions of the numeric standards; for example, "a 40% loss of sensitive taxa is typically associated with a decline in SCI values from good to fair."

- *Modify the SCI subsampling method to ensure that the number of individuals in each sample is close to the target of 100 .*

FDEP is currently modifying laboratory subsampling procedures to more reliably match the target of 100 individuals per sample. Additional subsampling of the original material collected from the stream will help eliminate samples with hundreds of individuals. Precautions are also being implemented to eliminate small sample. Results from this analysis suggest that subsamples with less than 75 individuals should probably be disqualified from SCI analysis.

- *Standardize the BioRecon sampling effort.*

The precision of this index could likely be improved by ensuring that the same level of effort is applied to each sample in the field. In addition, the number of individuals returned to the laboratory for verification should be consistent for each sample.

- *Sample a small number of sites over several years and during specifically defined seasons to evaluate natural variability associated with season and year.*

Variability of SCI due to season and year, though small, represents a source of variability that could be eliminated. The data for this study were insufficient to fully evaluate the influence of seasonal differences. Better estimates could be made with a sampling design that visited each site during every season and year. Definitions for season should be made based on climatic factors (e.g., rainfall, temperature) most relevant to invertebrate life histories. In addition, sites should be selected for which the level of human disturbance is constant through time. The best design would also include laboratory subsampling as part of the design to estimate this component of variance using information from the same sites.

- *Evaluate whether the increase in precision typically associated with larger sample sizes (greater number of individuals) is worth the cost.*

Studies in other regions report lower error variance when a greater number of individuals are identified and/or replicate samples are collected for each site. For SCI, the error variance associated with repeat visits to the same site was 15% of the total and for the BioRecon index it was 18% of the total. Fore et al. (2001) report error variance of 10% for an approximate 500-count sample. A larger sample (more individuals) would likely yield a more precise assessment, but also requires additional laboratory time for analysis. Replicate samples can also be used to increase precision and would also require additional field and laboratory analysis time. A cost/benefit analysis represents a formal process for evaluating the relative advantages of each approach.

- *Identify additional long-lived taxa.*

Long-lived taxa richness was the most variable metric in SCI, probably due to low numbers of taxa. If additional taxa could be added to this list, its statistical precision would likely improve.

ACKNOWLEDGMENTS

Florida DEP biologists provided comments and guidance throughout the development and testing of the stream macroinvertebrate indexes. R. Frydenborg provided data sets, developed the hydrologic scoring index, resolved taxonomic questions, and provided biogeographic interpretation of metrics. E. McCarron provided administrative support and a regulatory perspective. E. Livingston provided administrative and contract support. Reviewers from outside FDEP included M. Barbour, J. Gerritsen, J. Karr, T. Olsen, M. Pescador, and T. Thom who provided many helpful comments. Additional comments from F. Butera, J. Espy, S. Levings, D. Miller, D. Ray, and the FDEP Biology section improved many aspects of this document and analysis. Data retrievals were performed by T. Deck, S. Gerardi, and D. Miller. D. Cobb calculated land cover for stream sites. J. Espy developed data sets and resolved taxonomic questions. Discussions with J. Espy, S. Levings and D. Whiting improved the method for selecting sensitive taxa. D. Denson, D. Evans, J. Richardson, B. Rutter, and R. Wisseman (Aquatic Biology Assoc.) provided definitions for long-lived taxa. M. Heyn provided additional clinger information. Lab subsampling was done by A. Jordan, N. Moise, C. Oaks, B. Richardson, B. Shinner, and F. Whiddon. Biologists throughout Florida collected and identified stream macroinvertebrate samples, including L. Banks, F. Butera, G. Butts, T. Deck, D. Denson, L. Donalan, J. Espy, K. Espy, T. Frick, R. Frydenborg, S. Gerardi, M. Heyn, L. Homann, J. Jackson, P. Kane, S. Kent, C. Kovach, L. Miller, P. Morgan, P. O'Conner, E. Pluchino, D. Ray, J. Richardson, B. Rutter, E. Springer, C. Swanson, M. Szafraniec, M. Thompson, F. Walton, and A. Wheeler. This project and the preparation of this report were funded in part by a Section 319 Nonpoint Source Management Program grant from the U. S. Environmental Protection Agency (U. S. EPA) through a contract with the Watershed Management Program of the Florida Department of Environmental Protection. The total cost of the project was $64,800 of which $64,800, or 100%, was provided by the U. S. EPA.

REFERENCES

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.

Barbour M. T., Gerritsen J., Snyder B. D. and Stribling J. B. 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, DC.

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.

Blocksom, K. A. 2003. A performance comparison of metric scoring methods for a multimetric index for Mid-Atlantic Highlands streams. Environmental Management, *in press*.

Bowman, M. L., J. Gerritsen, G. R. Gibson, Jr., and B. D. Snyder. 2000. Estuarine and coastal marine waters: Bioassessment and biocriteria technical guidance. EPA-822-B-00-024. Office of Water, U. S. Environmental Protection Agency, Washington, DC.

Brigham, A. R., W. Brigham, and A. Gnilka. (Eds.) 1982. Aquatic Insects of North and South Carolina. Midwest Aquatic Enterprises, Mahomet, IL. 4.1-4.100.

Brown, M. T., N. Parker, and A. Foley. 1998. Spatial Modeling of Landscape Development Intensity and Water Quality in the St. Marks River Watershed. DEP Contract #GW138. Center for Wetlands, University of Florida, Gainesville.

Brown, M. T. and B. Vivas. *In review.* A landscape development intensity index. Environmental Monitoring and Assessment.

Bryce, S. A., R. M. Hughes, and P. R. Kaufmann. 2002. Development of a bird integrity index: using bird assemblages as indicators of riparian condition. Environmental Management 30:294-310.

Bryce, S. A., D. P. Larsen, R. M. Hughes, and P. R. Kaufmann. 1999. Assessing relative risks to aquatic ecosystems: A mid-Appalachian case study, Journal of the American Water Resources Association 35(1): 23-36.

Clements, W. H., D. M. Carlisle, J. M. Lazorchak, and P. C. Johnson. 2000. Heavy metals structure benthic communities in Colorado mountain streams. Ecological Applications 10:626-638.

Corbet, P. 1999. Dragonflies: Behavior and Ecology of Odonata. Cornell University Press. Ithaca, New York.

DeShon, J. E. 1995. Development and application of the invertebrate community index (ICI). In Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making (Eds. W. S. Davis and T. P. Simon), pp. 217-244. CRC Press, Boca Raton, FL.

Doberstein, C. P., J. R. Karr, L. L. Conquest. 2000. The effect of fixed-count subsampling on macroinvertebrate biomonitoring in small streams. Freshwater Biology 44:355-371.

Environmental Protection Agency (EPA). 1998. Lake and Reservoir Bioassessment and Biocriteria: Technical Guidance Document. EPA 841-B-98-007. U. S. EPA, Washington, DC. www.epa.gov/owow/monitoring/tech/lakes.html

Environmental Protection Agency (EPA). 2000. Stressor Identification Guidance Document. EPA-822-B-00-025. U. S. EPA, Office of Water, Office of Research and Development, Washington, DC.

Environmental Protection Agency (EPA). 2001. National Coastal Condition Report. EPA-620/R-01/005. Office of Research and Development/Office of Water, U. S. EPA, Washington, DC. www.epa.gov/owow/oceans/NCCR/index.

Environmental Protection Agency (EPA). 2002a. Methods for Evaluating Wetland Condition: Developing Metrics and Indexes of Biological Integrity. EPA-822-R-02-016.Office of Water, U. S. EPA, Washington DC.

Environmental Protection Agency (EPA). 2002b. Summary of Biological Assessment Programs and Biocriteria Development for States, Tribes, Territories, and Interstate Commissions: Streams and Wadeable Rivers. EPA-822-R-02-048. U.S. Environmental Protection Agency, Office of Environmental Information and Office of Water, Washington, DC.

Environmental Protection Agency (EPA). 2003a. Draft Report on the Environment 2003. EPA-260-R-02-006. Office of Environmental Information and Office of Research and Development, U. S. EPA, Washington, DC. www.epa.gov/indicators.

Environmental Protection Agency (EPA). 2003b. Biological Indicators of Watershed Health. www.epa.gov/bioindicators.

Environmental Protection Agency (EPA). 2003c. Guidance for 2004 Assessment, Listing and Reporting Requirements Pursuant to Sections 303(d) and 305(b) of the Clean Water Act. Memorandum. Office of Water, U. S. EPA, Washington DC. www.epa.gov/owow/tmdl/tmdl0103.

Fore L. S. 2002. Biological assessment of mining disturbance on stream invertebrates in mineralized areas of Colorado. In: Biological Response Signatures: Indicator Patterns Using Aquatic Communities (Ed T. P. Simon), pp. 445-480. CRC Press LLC, Boca Raton, FL.

Fore, L. S. 2003. Developing Biological Indicators: Lessons Learned from Mid-Atlantic Streams. EPA 903/R-003/003. U.S. Environmental Protections Agency, Office of Environmental Information and Mid-Atlantic Integrated Assessment Program, Region 3, Ft. Meade, MD. www.epa.gov/bioindicators/

Fore, L. S. and C. Grafe. 2002. Using diatoms to assess the biological condition of large rivers in Idaho (USA). Freshwater Biology 47:2015-2037.

Fore L. S., Karr J. R. and Conquest L. L. 1994. Statistical properties of an index of biotic integrity used to evaluate water resources. Canadian Journal of Fisheries and Aquatic Sciences 51:212-231.

Fore, L. S., J. R. Karr, and R. W. Wisseman. 1996. Assessing invertebrate response to human activities: evaluating alternative approaches. Journal of the North American Benthological Society 15:212-231.

Fore L. S., Paulsen K., K. O'Laughlin. 2001. Assessing the performance of volunteers in monitoring streams. Freshwater Biology 46:109-123

Fernald, E. A. and E. D. Purdum (Eds.) 1992. Atlas of Florida. University Press of Florida. 280 pp.

Florida Natural Areas Inventory (FNAI). 1990. Guide to the Natural Communities of Florida. Florida Natural Areas Inventory and Florida Department of Natural Resources. 111 pp.

Griffith, G. E., J. M. Omernik, C. M. Rohm, and S. M. Pierson. 1994. Florida regionalization project. U. S. EPA. Environmental Research Laboratory, Corvallis, OR. 83 pp.

Hawkins, C. P., R. H. Norris, J. N. Hogue, and J. W. Feminella. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecological Applications:10:1456–1477.

Hoenig, J. M. and D. M. Heisey. 2001. The abuse of power: the pervasive fallacy of power calculations for data analysis. The American Statistician 55:19-24.

Hughes, R. M., P. R. Kaufmann, A. T. Herlihy, T. M. Kincaid, L. Reynolds, and D. P. Larsen. 1998. A process for developing and evaluating indices of fish assemblage integrity. Canadian Journal of Fisheries and Aquatic Sciences 55:1618-1631.

Hughes, R. M. and T. Oberdorff. 1998. Application of IBI concepts and metrics to waters outside the United States. In: Assessing the Sustainability and Biological Integrity of Water Resources Using Fish Communities (Ed T. P. Simon) pp. 79-93. CRC Press, Boca Raton, FL.

Hutchens, J. J. Jr., K. Chung, and J. B. Wallace. 1998. Temporal variability of stream macroinvertebrate abundance and biomass following pesticide disturbance. Journal of the North American Benthological Society 17:518-534.

Jackson, L. E., J. C. Kurtz, and W. S. Fisher (Eds.). 2000. Evaluation Guidelines for Ecological Indicators. EPA/620/R-99/005. U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC. 107 pp.

Johnson, D. H. 1999. The insignificance of statistical significance testing. Journal of Wildlife Management 63:763-772.

Karr, J. R. 1981. Assessment of biotic integrity using fish communities. Fisheries 6(6):21-27.

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

Karr, J. R. 1998. Rivers as sentinels: using the biology of rivers to guide landscape management. In: River Ecology and Management: Lessons from the Pacific Coastal Ecosysten (Eds. R. J. Naiman & R. E. Bilby), pp. 502-528. Springer, NY.

Karr J. R., Allan J. D. and Benke A. C. 2000. River conservation in the United States and Canada. In: Global Perspectives on River Conservation: Science, Policy, Practice (Eds P. J. Boon, B. R. Davies and G. E. Petts), pp. 3-39. J. Wiley, Chichester, UK.

Karr J. R. and Chu E. W. 1999. Restoring Life in Running Waters: Better Biological Monitoring. Island Press, Washington, DC.

Karr J. R., Fausch K. D., Angermeier P. L., Yant P. R. and Schlosser I. J. 1986. Assessment of biological integrity in running water: a method and its rationale. Illinois Natural History Survey Special Publication Number 5, Champaign, IL.

Karr., J. R. and E. Morishita Rossano. 2001. Applying public health lessons to protect river health. Ecology and Civil Engineering 4:3-18.

Karr, J. R., P. R. Yant, and K. D. Fausch. 1987. Spatial and temporal variability of the index of biotic integrity in three Midwestern streams. Transactions of the American Fisheries Society 116:1-11.

Karr, J. R. and C. O. Yoder. 2004. Biological assessment and criteria improve TMDL decision making. Journal of Environmental Engineering, *in press.*

Kerans, B. L., and J. R. Karr. 1994. A benthic index of biotic integrity (B-IBI) for rivers of the Tennessee River Valley. Ecological Applications 4:768-785.

Kimberling, D. N., J. R. Karr, and L. S. Fore. 2001. Measuring human disturbance using terrestrial invertebrates in the shrub-steppe of eastern Washington (USA). Ecological Indicators 1:63-81.

Klemm, D. J., K. A. Blocksom, F. A. Fulk, A. T. Herlihy, R. M. Hughes, P. R. Kaufmann, D. V. Peck, J. L. Stoddard, W. T. Thoeny, M. B. Griffith, and W. S. Davis. 2003. Development and evaluation of a macroinvertebrate biotic integrity index (MBII) for regionally assessing Mid-Atlantic Highlands streams. Environmental Management 31(5):656-669.

Klemm, D. J., K. A. Blocksom, W. T. Thoeny, F. A. Fulk, A. T. Herlihy, P. R. Kaufmann, S. M. Cormier. 2002. Using macroinvertebrates as indicators of ecological conditions for streams

in the Mid-Atlantic Highlands region. Environmental Monitoring and Assessment 78:169-212.

McCarron, E. and R. Frydenborg. 1997. The Florida bioassessment program: an agent for change. Human and Ecological Risk Assessment 3:967-977.

Mebane, C. A. 2001. Testing bioassessment metrics: macroinvertebrate, sculpin, and salmonid responses to stream habitat sediment and metals. Environmental Monitoring and Assessment 67:293-322.

Mebane, C. A. 2002. Effects of metals on freshwater macroinvertebrates: a review and case study of the correspondence of multimetric index, toxicity testing, and copper concentrations in sediment and water. In: Biological Response Signatures: Indicator Patterns Using Aquatic Communities (Ed T.P. Simon), pp. 287-312. CRC Press LLC, Boca Raton, FL.

Merritt R. W. and K. W. Cummins. (Eds.). 1996. An Introduction to the Aquatic Insects of North America, 3rd ed. Kendall/Hunt Publishing Company, Dubuque, IA.

Merritt, R. W., J. R. Wallace, M. J. Higgins, M. K. Alexander, M. B. Berg, W. T. Morgan, K. W. Cummins, and B. Vandeneeden. 1996. Procedures for the functional analysis of invertebrate communities of the Kissimmee river floodplain ecosystem. Florida Scientist 59:216-274.

Milliken, G. A. and D. E. Johnson. 1992. Analysis of Messy Data. Volume 1: Designed Experiments. Chapman and Hall. New York, NY.

Miltner, R. J. and E. T. Rankin. 1998. Primary nutrients and the biotic integrity of rivers and streams. Freshwater Biology 40:145-158.

Morley, S. A. and J. R. Karr. 2002. Assessing the biological health of urban streams: tools for restoration and conservation. Conservation Biology 16:1498-1509.

Norton, S. B., S. M. Cormier, M. Smith, R. Christian Jones. 2000. Can biological assessments discriminate among types of stress? A case study from the Eastern Corn Belt Plains ecoregion. Environmental Toxicology and Chemistry 19:1113-1119.

Pescador, M., Rasmussen, A., Harris, S. 1995. Identification Manual for the Caddisfly (Trichoptera) Larvae of Florida. Florida Department of Environmental Protection.

Pescador, M., Rasmussen, A., Richard, B. 2000. A Guide to the Stoneflies (Plecoptera) of Florida. Florida Department of Environmental Protection.

Ransel K. P. 1995. The sleeping giant awakes: PUD No. 1 of Jefferson County v. Washington Department of Ecology. Environmental Law 25:255-283.

Rasmussen, A. and M. Pescador. 2002. A Guide to the Megaloptera and Aquatic Neuroptera of Florida. Florida Department of Environmental Protection.

Soil and Water Conservation Society (SWCS). 1989. Twenty-six ecological communities of Florida. Florida Soil and Water Conservation Society. 146 pp.

Smith, D. 2001. Pennak's Freshwater Invertebrates of the United States. John Wiley and Sons. New York, NY.

Snell, L. J. and W. Anderson. 1970. Water resources of Northeast Florida. Florida Bureau of Geology. Invest. No. 54. 77 pp.

Steedman, R. J. 1988. Modification and assessment of an index of biotic integrity to quantify stream quality in southern Ontario. Canadian Journal of Fisheries and Aquatic Sciences 45:492-501.

Thompson, F. 1984. Freshwater Snails of Florida, A Manual for Identification. University Presses of FL, Gainesville, FL.

Thorne R. St. J. and Williams W. P. 1997. The response of benthic macroinvertebrates to pollution in developing countries: a multimetric system of bioassessment. Freshwater Biology 37: 671-686.

Thorp, J., and A. Covich (Eds.). 1991. Ecology and Classification of North American Freshwater Invertebrates.

Vowell, J. L. 2001. Using stream bioassessment to monitor best management practice effectiveness. Forest Ecology and Management 143:237-244.

White, W. A. 1958. Some geomorphic features of Central Peninsular Florida. Florida Bureau of Geology. Geological Bulletin 41. 92 pp.

White, W. A. 1970. The geomorphology of the Florida peninsula. Florida Bureau of Geology. Geological Bulletin 51. 164 pp.

Yoder, C. O. and J. E. DeShon. 2002. Using biological response signatures within a framework of multiple indicators to assess and diagnose causes and sources of impairments to aquatic assemblages in selected Ohio rivers and streams. In: Biological Response Signatures: Indicator Patterns Using Aquatic Communities (Ed T.P. Simon), pp. 23-82. CRC Press LLC, Boca Raton, FL.

Yoder, C. O. and E. T. Rankin. 1998. The role of biological indicators in a state water quality management process. Environmental Monitoring and Assessment 51:61-88.

Zar J.H. 1984. Biostatistical Analysis, 2nd ed. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Appendix A. Long-lived taxa

Taxonomic order, family, genus or species name of taxa used to calculate long-lived taxa richness. Taxonomic synonyms present in the current Florida data base are also shown. Long-lived (semi-voltine) taxa require more than one year to complete their life cycles.

Order	Family	Genus	Species (Long-lived)	Synonyms
Decapoda			All families, genera and species	
Basommatophora	Lymnaeidae	*Fossaria*	All species	
Mesogastropoda	Ampulariidae	*Pomacea*	All species	
Veneroida	Corbiculiidae	*Corbicula*	*Corbicula fluminea*	*Corbicula manilensis*
Unionoida	Unionidae		All genera and species	
Odonata	Aeshnidae	*Basiaeschna*	*Basiaeschna janata*	
		Boyeria	All species	
	Cordulegastridae		All species	
	Gomphidae	*Gomphus*	All species	*Gomphurus*
		Haegenius	*Hagenius brevistylus*	
		Progomphus	All species	
	Libellulidae	*Macromia*	All species	
		Somatochlora	All species	
		Tetragoneuria	All species	*Epitheca*
	Petaluridae	*Tachopteryx*	*Tachopteryx thoreyi*	
Plecoptera	Pteronarcidae	*Pteronarcys*	*Pteronarcys dorsata*	
	Leuctridae	*Leuctra*	All species	
	Peltoperlidae	*Tallaperla*	*Tallaperla cornelia*	
	Perlidae	*Acroneuria*	All species	
		Agnetina	*Agnetina annulipes*	
		Eccoptura	*Eccoptura xanthenes*	*Acroneuria xanthenes*
		Neoperla	All species	
		Paragnetina	All species	*Banksiana*
				Banksiella
	Perlodidae	*Clioperla*	*Clioperla clio*	*Isoperla clio*

Order	Family	Genus	Species (Long-lived)	Synonyms
		Perlinella	*Perlinella drymo*	*Atoperla ephyre*
			*Perlinella ephyre/zwick*i	
		Isoperla	*Isoperla orata*	
Megaloptera	Corydalidae		All genera and species	
Trichoptera	Brachycentridae	*Brachycentrus*	*Brachycentrus americanus*	*Oligoplectrum amercanum*
		Micrasema	All species	
	Calamoceratidae	*Heteroplectron*	*Heteroplectron americanum*	
	Hydropsychidae	*Diplectrona*	*Diplectrona modesta*	
		Macrostemum	*Macrostemum carolina*	*Macronema carolina*
				Macronemum carolina
	Lepidostomidae	*Lepidostoma*	All species	*Alepomyia, Alepomyiodes, Arcadopsyche, Atomyia, Jenortha, Mormomyia, Neuropsyche, Nosopus, Notiopsyche, Olemira, Oligopsyche, Phanopsyche, Pristosilo, Athripsodes*
	Leptoceridae	*Ceraclea*	All species	
	Limnephilidae	*Ironoquia*	*Ironoquia punctatissisma*	
		Pycnopsyche	All species	
	Molannidae	*Molanna*	All species	
	Phryganeidae	*Banksiola*	*Banksiola concatenata*	*Neuronia concatenata*
		Ptilostomis	*Ptilostomis postica*	*Neuronia postica*
		Agrypnia	*Agrypnia vestita*	
	Psychomyiidae	*Lype*	*Lype diversa*	
	Rhyacophilidae	*Rhyacophila*	All species	
	Uenoidae	*Neophylax*	All species	

Appendix B. Sensitive taxa

Taxonomic order, family, genus or species name; taxonomic synonyms in the current Florida data base; whether the taxon was included in the Florida index (2=very sensitive; 1=sensitive); and species name of Florida index taxa are listed for taxa identified as sensitive to human disturbance.

Order	Family	Genus	Species (sensitive)	Synonyms	Florida index	
Acariformes	Lebertiidae	*Lebertia*	All species			
Isopoda	Asellidae	*Caecidotea*	*Caecidotea*	*Asellus*	1	*Caecidotea* spp.
Amphipoda	Crangonyctidae	*Crangonyx*	*Crangonyx*			
Ephemeroptera	Baetidae	*Acerpenna*	*Acerpenna pygmaea*	*Baetis pygmaeus*		
	Ephemerellidae		All genera and species			
	Heptageniidae		All genera and species		1	*Stenacron interpunctatum* *Stenonema mexicanum* *integrum*
					2	*Stenonema exiguum* *Stenonema smithae*
	Leptophlebiidae		All genera and species			
	Leptohyphidae		All genera and species			
Odonata	Libellulidae	*Macromia*	All species		2	*Macromia* spp.
Plecoptera			All families, genera and species		2	All Plecoptera
Trichoptera	Hydropsychidae	*Hydropsyche*	All species		2	*Hydropsyche* spp.
	Leptoceridae	*Triaenodes*	All species			
	Philopotamidae	*Chimarra*	*Chimarra*		2	*Chimarra* spp.
	Psychomyiidae	*Lype*	*Lype diversa*			
Coleoptera	Elmidae	*Ancyronyx*	*Ancyronyx variegatus*			
		Gonielmis	*Gonielmis dietrichi*			
		Macronychus	*Macronychus glabratus*			

Order	Family	Genus	Species (sensitive)	Synonyms	Florida index	
Diptera	Chironomidae	*Microtendipes*	All species			
		Parametriocnemus	All species			
		Polypedilum	*Polypedilum aviceps*			
		Rheocricotopus	All species		2	*Rheocricotopus robacki*
		Stempellinella	All species			
		Tanytarsus	*Tanytarsus sp. d epler*			
		Tanytarsus	*Tanytarsus sp. m epler*			
		Tribelos	*Tribelos jucundum*			
	Empididae	*Hemerodromia*	*Hemerodromia*			
	Simuliidae		All genera and species		2	*Simulium* spp.

Appendix C. Very tolerant taxa

Taxonomic order, family, genus or species name; taxonomic synonyms in the current Florida data base; whether the taxon was included in the Florida index (2=very sensitive; 1=sensitive); and species name of Florida index taxa are listed for taxa identified as very tolerant to human disturbance.

Order	Family	Genus	Species (very tolerant)	Synonyms	Florida index	
Hoplonemertea	Tetrastemmatidae	*Prostoma*	*Prostoma rubrum*			
Haplotaxida	Naididae	*Bratislavia*	*Bratislavia unidentata*			
		Dero	All species			
		Nais	*Nais communis* complex	*Nais communis*		
				Nais elinguis		
				Nais pardalis		
				Nais variabilis		
		Pristina	*Pristina synclites*			
	Tubificidae	*Haber*	*Haber speciosus*			
		Limnodrilus	All species	*Camptodrilus californicus*		
				Limnodrilus aurantiacus		
Lumbriculida	Lumbriculidae	*Lumbriculus*	All species			
Rhynchobdellida	Glossiphoniidae		All genera and species			
Basommatophora	Ancylidae	*Hebetancylus*	*Hebetancylus excentricus*			
		Laevapex	All species			
	Physidae	*Physella*	All species			
	Planorbidae		All genera and species			
Mesogastropoda	Hydrobiidae	*Pyrgophorus*	*Pyrgophorus platyrachis*	*Pyrogophorus platyrachis*		
	Thiaridae	*Melanoides*	All species			
Odonata	Libellulidae	*Pachydiplax*	*Pachydiplax longipennis*			
	Coenagrionidae	*Argia*	*Argia sedula*		2	*Argia sedula*

Order	Family	Genus	Species (very tolerant)	Synonyms		Florida index
		Enallagma	*Enallagma cardenium*	*Enallagma coecum*		
		Ischnura	All species	*Anomalagrion*		
Heteroptera	Mesoveliidae		All genera and species			
Lepidoptera	Pyralidae		All genera and species			
Coleoptera	Haliplidae	*Peltodytes*	All species			
Diptera	Chironomidae	*Larsia*	All species		1	*Larsia lurida*
					1	*Larsia decolorata*
		Chironomus	All species	*Tendipes* *Chaetolabis*		
		Cladopelma	*Cladopelma*			
		Cricotopus	All species		2	*Cricotopus bicinctus*
		Cryptochironomus	All species			
		Dicrotendipes	*Dicrotendipes modestus*			
			Dicrotendipes neomodestus			
		Glyptotendipes	All species			
		Goeldichironomus	All species			
		Kiefferulus	*Kiefferulus*			
		Parachironomus	All species			
		Polypedilum	*Polypedilum beckae*	*Asheum beckae* *Pedionomus beckae*		
			Polypedilum illinoiense grp.	*Polypedilum illnoense*	1	*Polypedilum illinoiense grp.*
			Polypedilum tritum	*Chironomus tritum*		
		Tanypus	All species	*Pelopia*		
		Tanytarsus	*Tanytarsus sp. f epler*			
	Stratiomyidae	*Odontomyia*	*Odontomyia*	*Eulalia*		
	Tipulidae	*Tipula*	*Tipula*	*Nippotipula* *Playttipula* *Yamatotipula*		

Appendix D. ANOVA output and regression results for SCI and BioRecon.

Table 1. General linear model output for SCI vs. HDG by region (ne, ph, or ps). Neither region nor the interaction of HDG and region was statistically significant. Test failed to detect a difference in SCI values due to region.

Univariate Tests of Significance for AvgOfSCI6_100 (1NGM_4)					
	SS	df	MS	F	p
Intercept	403623.3	1	403623.3	1832.253	0.000000
Region	361.6	2	180.8	0.821	**0.441846**
HDG2	62750.7	1	62750.7	284.858	0.000000
Region*HDG2	1142.9	2	571.5	2.594	**0.077660**
Error	37449.0	170	220.3		

Table 2. General linear model output for SCI vs. HDG by year class (1992-3, 1994-5, 1996-7, 1998-9, and 2000-1). The interaction term was not significantly different for year class and HDG suggesting the lines were parallel (equal slopes). The year class alone, however, was significant with values for 1994-5 being slightly higher for SCI.

Univariate Tests of Significance for SCI6_100 (1NGM_4)					
	SS	df	MS	F	p
Intercept	1627060	1	1627060	5835.657	0.000000
Year class	2948	4	737	2.644	**0.032766**
HDG2	134497	1	134497	482.390	0.000000
Year class*HDG2	791	4	198	0.709	**0.586005**
Error	171749	616	279		

Table 3. ANOVA estimates of SCI variability for repeat visits to the same site within year. Error estimate used to calculate confidence interval underlined.

ANOVA for SCI (SCI6_100) (Same_day_SCI)							
	Effect	df	MS	df	MS	F	p
{1}STORET	Random	60	1211.636	39.41542	188.7279	6.420014	0.000000
{2}Year	Random	6	269.430	35.76223	170.2402	1.582648	0.180727
1*2	Random	53	134.481	82.00000	**66.5952**	2.019385	0.002054

Table 4. ANOVA estimates of SCI variability for laboratory subsampling. Error estimate used are underlined

ANOVA Results: SCI subsampling (Metrics)							
	Effect	df	MS	df	MS	F	p
{1}Sitename	Random	58	1393.986	118	**32.86312**	42.41795	0.00

Table 5. ANOVA estimates of BioRecon variability for repeat visits to the same site within year. Error estimate used to calculate confidence interval underlined.

ANOVA for BioRecon (BR3_10) (New metrics multiple_2)							
	Effect	df	MS	df	MS	F	p
{1}STORET	Random	126	15.91691	117.3771	1.696009	9.384924	0.000000
{2}year	Random	8	6.02260	156.0000	**1.458840**	4.128352	0.000171