2014

USING INDICATORS OF BIOTIC INTEGRITY FOR ASSESSMENT OF STREAM CONDITION

Stephanie A. Ogren
Michigan Technological University

Follow this and additional works at: https://digitalcommons.mtu.edu/etds
Part of the Biology Commons, and the Ecology and Evolutionary Biology Commons

USING INDICATORS OF BIOTIC INTEGRITY FOR ASSESSMENT OF STREAM CONDITION

By

Stephanie A. Ogren

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Biological Sciences

MICHIGAN TECHNOLOGICAL UNIVERSITY

2014

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Biological Sciences

Department of Biological Sciences

Dissertation Advisor: *Dr. Casey J Huckins*

Committee Member: *Dr. Rodney A. Chimner*

Committee Member: *Dr. Amy M. Marcarelli*

Committee Member: *Dr. Eric B. Snyder*

Department Chair: *Dr. Chandrashekhar Joshi*

Table of Contents

Preface

Goals and objectives

The overarching goal of this research was to assess the characteristics of indices of biotic integrity that had previously been developed for the Upper Midwest and determine their assessment capability using local, long-term datasets. We assessed indices developed for application to stream macroinvertebrate community assemblages and in a companion project we assessed indices developed for fish community assemblage data. We then used the best performing indices to evaluate the responses of biotic communities in three sites in the Big Manistee River Watershed to determine the effects of culvert replacement on improving biotic integrity of the stream system. Objectives included:

1. Evaluation of suitability and comparability of stream assessment indices using macroinvertebrate data sets from the Northern Lakes and Forests Ecoregion (Chapter 2).
2. Comparison and application of coolwater and coldwater fish indices for the Upper Midwest (Chapter 3).
3. Culvert replacement effects on fish and macroinvertebrate index of biotic integrity scores (Chapter 4).

Format of dissertation chapters

Other than the introductory chapter (Chapter 1), all other chapters have been developed for publication in pertinent scientific journals collaboratively with review of proposal and general focus of the research completed by the entire committee. Chapter 2 has been published in the journal *Ecological Indicators* (Ogren and Huckins 2014) with contributions of data compilation completed by myself and analysis and manuscript development completed by myself and Dr. Casey Huckins. Chapter 3 has been submitted

for publication in *The North American Journal of Fisheries Management* with contributions of data compilation completed by myself and analysis and manuscript development completed by myself and Dr. Casey Huckins. Chapter 4 has been submitted for publication in *Restoration Ecology* with data compliation completed by myself and analysis and manuscript development completed by myself and Dr. Casey Huckins.

Abstract

Multiple indices of biotic integrity and biological condition gradient models have been developed and validated to assess ecological integrity in the Laurentian Great Lakes Region. With multiple groups such as Tribal, Federal, and State agencies as well as scientists and local watershed management or river-focused volunteer groups collecting data for bioassessment it is important that we determine the comparability of data and the effectiveness of indices applied to these data for assessment of natural systems. We evaluated the applicability of macroinvertebrate and fish community indices for assessing site integrity. Site quality (i.e., habitat condition) could be classified differently depending on which index was applied. This highlights the need to better understand the metrics driving index variation as well as reference conditions for effective communication and use of indices of biotic integrity in the Upper Midwest. We found the macroinvertebrate benthic community index for the Northern Lakes and Forests Ecoregion and a coldwater fish index of biotic integrity for the Upper Midwest were most appropriate for use in the Big Manistee River watershed based on replicate sampling, ability to track trends over time and overall performance. We evaluated three sites where improper road stream crossings (culverts) were improved by replacing them with modern full-span structures using the most appropriate fish and macroinvertebrate IBIs. We used a before-after-control-impact paired series analytical design and found mixed results, with evidence of improvement in biotic integrity based on macroinvertebrate indices at some of the sites while most sites indicated no response in index score. Culvert replacements are often developed based on the potential, or the perception, that they will

restore ecological integrity. As restoration practitioners, researchers and managers, we need to be transparent in our goals and objectives and monitor for those results specifically. The results of this research serve as an important model for the broader field of ecosystem restoration and support the argument that while biotic communities can respond to actions undertaken with the goal of overall restoration, practitioners should be realistic in their expectations and claims of predicted benefit, and then effectively evaluate the true impacts of the restoration activities.

Chapter 1. Overview

Introduction

One of the greatest environmental challenges of this century is to sustain natural biological structure and functional attributes of aquatic ecosystems, rivers in particular (Bernhardt et al. 2006). This goal requires that we know the condition of these dynamic systems and can assess what and how specific factors and forces are affecting them. Due to past degradation, it is often the case that river ecosystems require rehabilitation to redirect them to a desired state. Despite the value humans place on rivers and streams for drinking water, agriculture, recreation and food, human activities continue to disturb and degrade the natural structure and function of these systems (NRC 1992). During the last two decades, there has been extensive local, national and international effort to improve the quality and integrity of freshwater ecosystems (Frissell and Bayles 1996, Stanford et al. 1996, Baron et al. 2002). Rehabilitation efforts are often focused on improvement of resources that are of economic, cultural or spiritual importance (Roni et al. 2008). Unfortunately, much of this effort has proceeded without documentation of the relative successes and failures of individual activities (Ham and Pearsons 2000). Even when success is noted there is often a lack of data to identify specific results or endpoints for the management activity (Bernhardt et al. 2005). Thus, whether our goal is to protect a relatively pristine system, manage an actively used system, or improve a degraded one, it is imperative that we efficiently and accurately assess the physical and biological condition of ecosystems to guide the process. An integrated approach that evaluates the

utility and applicability of bioassessment for monitoring and assessment data analysis is necessary.

Historically, the quality of a stream was assessed by monitoring the chemical and toxicological parameters with little attention given to biotic components. Chemical and toxicological data may not provide a broad view of the system condition that is integrated over longer temporal and spatial extents. Recently, attention has focused on a more integrated approach with greater emphasis on physical and biological integrity (US EPA 2002). Monitoring programs are shifting from point source, single parameter monitoring to holistic watershed strategies that include greater focus on biological entities and processes while continuing to monitor the standard physical and chemical parameters. Analyzing the biological community as a whole and across multiple scales, may allow detection of water quality problems that could be overlooked by assessessments focused on only chemical parameters. This shift formalized the bioassessment approach and promotes the evaluation of the biological integrity of a waterbody. Karr and Dudley (1981) defined biotic integrity as "the ability to support and maintain a balanced, integrated, adaptive community of organisms having a species composition, diversity, and functional organization comparable to that of the natural habitat of the region".

With numerous national, regional and local organizations working independently on aquatic assessment programs, innovative technical approaches have been developed (Davies and Jackson 2006). However, this has also been problematic as various organizations use different methods for data collection and analysis with little effort towards standardization. This has often resulted in organizations creating their own assessment program to address management needs and objectives with little regard for

related programs and opportunities for synergistic collaboration. Variation in duration of effort, choice of index periods, and analytical techniques are all factors resulting in lack of consistent local, regional and national reporting on aquatic resource condition. Without a consistent approach it is extremely difficult to compare assessment results and communicate conditions to the public and policy makers. Biotic indices aid in the interpretation of biological assessment data. The product of an index of biotic integrity (IBI) is a single site- and time-specific numeric score that can be interpreted within a regional gradient of condition (Karr and Chu 1999). In the Upper Midwest of the United States there are numerous indices available (fish and macroinvertebrate based); however, determining the appropriate index and when to apply it is problematic. One biological data set can be interpreted differently depending on the index applied, each subsequently indicating different courses of action.

We assessed fish and macroinvertebrate indicators of biotic integrity (IBIs) over temporal and spatial scales. Issues of thermal regime, taxonomic resolution and sampling protocols were addressed. We then were able to determine the most appropriate IBI for assessment in the Big Manistee River watershed based on variability, comparability, correlation analysis and ability to track trends over time. We used these IBIs to assess the effectiveness of three culvert replacements that were completed in the watershed.

Restoration

Aquatic systems have been degraded and continue to be impaired as a result of anthropogenic actions (Dudgeon et al. 2006). The degree to which physical or ecological function is compromised or impaired will dictate the potential for a system to resist

change or improve in condition (Brierly and Fryirs 2005). River restoration has become a common management activity that is growing exponentially (Bernhardt et al. 2005). A widespread constraint of aquatic restoration is the lack of data and consistency required to identify achievable results or endpoints for the management activity (Palmer et al. 2005).

Locally, multi-agency efforts for habitat restoration have occurred within, and in close proximity to the Big Manistee River watershed in the Lower Peninsula of Michigan. Historically, little of this work was evaluated to determine the benefit to the watershed for more than aesthetic criteria. Agencies within the region have identified a need to improve evaluation, monitoring and selection of restoration projects. We applied the conclusions from the research outlined in the first chapters to identify the most appropriate indices of biotic integrity and applied them to evaluate biological responses to three culvert replacements. These types of projects are often undertaken with the expectation that they will improve biotic integrity of a system (Roni et al. 2008). While we did detect positive biological responses to the culvert replacements, this research supported the conclusion that incorporating watershed scale assessment will aid in directing ecologically significant restoration and lead to realistic, physical and biological goals that can be appropriately monitored.

Literature Cited

Baron, J.S., Poff, N.L., Angermeier, P.L., Dahm, C.N., Gleick, P.H., Nelson, G.H., Jackson, R.B., Johnston, C.A., Richter, B.D., Steinman, A.D. 2002. Meeting Ecological and Societal Needs for Freshwater. Ecological Applications 12, 1247-1260.

Bernhardt, E. S., Bunn, S. E., Hart, D. D., Malmqvist, B., Muotka, T., Naiman, R. J., Pringle, C., Ruess, M. , Van Wilgen, B. 2006. The challenge of ecologically sustainable water management. Water Policy 8:475-479.

Bernhardt, E.S., Palmer, M.A., Allan, J.D., Alexander, G., Barnas, K., Brooks, S., Carr, J., Clayton, S., Dahm, C., Follstad-Shah, J., Galat, D., Gloss, S., Goodwin, P., Hart, D., Hassett, B., Jenkinson, R., Katz, S., Kondolf, G.M., Lake, P.S., Lave, R., Meyer, J.L., O'Donnell, T.K., Pagano, L., Powell, B., Sudduth, E. 2005. Ecology - Synthesizing US River Restoration Efforts. Science 308, 636-637.

Brierley, G. J., Fryirs. K. A. 2005. Geomorphology and River Management: Applications of the River Styles Framework. Blackwell Publishing, Oxford, UK, 398pp.

Davies, S.P., Jackson, S.K. 2006. The Biological Condition Gradient: A Descriptive Model for Interpreting Change in Aquatic Ecosystems. Ecological Applications 16, 1251-1266.

Dudgeon, D., Arthington, A. H., Gessner, M.O., Kawabata, Z., Knowler, D., Lévêque, C., Naiman, R.J., Prieur-Richard, A. H., Soto, D., Stiassny, M.L.J., Sullivan, C.A. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biological Reviews 81: 163-182.

Frissell, C.A., Bayles, D. 1996. Ecosystem Management and the Conservation of Aquatic Biodiversity and Ecological Integrity. Water Resource Bulletin 32, 229-240.

Ham, K.D., Pearsons, T.N. 2000. Can Reduced Salmonid Population Abundance be Detected in Time to Limit Management Impacts? Canadian Journal of Fisheries and Aquatic Sciences 57, 17-24.

Karr, J.R., Dudley, D.R. 1981. Ecological Perspective on Water-Quality Goals. Environmental Management 5, 55-68.

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

Karr, J.R., Chu, E.W. 2000. Sustaining Living Rivers. Hydrobiologia 422-423, 1-14.

NRC (National Research Council) 1992. *Restoration of Aquatic Ecosystems; Science, Technology, and Public Policy.* National Academy Press, Washington, D.C.

Palmer, M.A., Bernhardt, E.S., Allen, J.D. , Lake, P.S., Alexander, G., Brooks, S., Carr,

J., Clayton, S., Dahm, C.N., Follstad Shah, J., Galat, D.L., Gloss, S., Goodwin, P., Hart, D.D., Hassett, B., Jenkinson, R., Kondolf, K.M., Lave, R., Meyer, J.L., O'Donnell, T.K., Pagano, L.,. Sudduth, E. 2005. Standards for ecologically successful river restoration. Journal of Applied Ecology 42:208-217.

Roni, P., Hanson, K., Beechie, T. 2008. Global Review of the Physical and Biological Effectiveness of Stream Habitat Rehabilitation Techniques. North American Journal of Fisheries Management 28, 856-890.

Stanford, J.A., Ward, J.V., Liss, W.J., Frissell, C.A., Williams, R.N., Lichatowich, J.A., Coutant, C.C. 1996. A General Protocol for Restoration of Regulated Rivers. Regulated Rivers-Research & Management 12, 391-413.

USEPA (U.S. Environmental Protection Agency). 2002. 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.

Chapter 2. [1] Evaluation of suitability and comparability of stream assessment indices using macroinvertebrate data sets from the Northern Lakes and Forests Ecoregion

1. Introduction

Whether the goal is to protect a relatively pristine ecosystem, manage an actively used system, or restore a degraded one, the approach and success relies on our knowledge and assessment of the physical and biological condition of ecosystems. Aquatic biological monitoring has been recognized as the first step in protecting biological integrity (Karr and Chu, 1999). Assessing the ecological condition of a site may be approached through multiple methods, often with the estimated biological condition dependent on many factors, including the organisms selected for use in the interpretation (Carter and Resh, 2001), how data are interpreted (Cao et al., 2005), and methods used to collect the data (Hughes and Peck, 2008).

Numerous national, regional and local organizations have independently developed aquatic assessment programs producing many innovative technical approaches for data acquisition and interpretation (Davies and Jackson, 2006) but with little standardization; therefore, determining the comparability of data collected and resulting assessments is needed (Cao and Hawkins, 2011). The ability to utilize multiple sources of data could benefit programs by allowing for validation of assessments if they are shown to be comparable (Herbst and Sillfdorff, 2006; Gerth and Herlihy, 2006; Rehn et al., 2007).

Biotic indices have been developed to aid in the interpretation of biological assessment data. The product of a biotic index is a single site- and time-specific numeric

[1] The material contained in this chapter was previously published in the journal *Ecological Indicators*.

score that can be interpreted within a regional gradient of condition (Karr and Chu, 1999). Assessment of the utility and applicability of these indices over spatial (Ode et al., 2008) and temporal scales is also necessary (Mazor et al., 2009). Determining comparability of this numeric score and inferences derived from these endpoints has become necessary to improve regulatory credibility, reduce redundancy, increase efficiency, improve long-term monitoring programs, expand assessments to a broader scale and generally increase sample size, which would improve assessment (Cao and Hawkins, 2011). In the Upper Midwest of the United States there are numerous indices available; however, determining the appropriate index and when to apply it is problematic. One biological data set can be interpreted in different ways and subsequently indicate different courses of action based on which index is applied.

Agencies within the Upper Midwest currently use disparate sampling methodologies and biological indices to assess stream systems. Often, management agencies use indices that are not directly comparable, having varying scales and different classification schemes. One of our goals for this study was to determine if indices developed for use at different spatial scales in the Upper Midwest (Figure 2.1) would produce concordant index scores within and across sites. We used a nested approach to evaluate sites based on scores from indices developed with increasing geographic scope. By nested approach we mean that the data set from the Big Manistee River watershed is within the state of Michigan, which is within the Northern Lakes and Forests ecoregion within the Upper Midwest. Scoring of sites is in comparison to reference condition or theoretical natural state utilized in the original development of the index. The natural variation across a larger region may limit discrimination of site specific differences in a

regionally derived index. A locally derived index may be necessary for discrimination of smaller changes in biotic integrity (Ode et al., 2008). A nested approach to data interpretation may lead to better understanding of variation in ecological condition and the geographic scope appropriate for interpretation.

Evaluation of stream condition is also dependent on the temporal stability of a system (Milner et al., 2006). Temporal variation in community assemblage occurs both seasonally and annually. Seasonal variability has been shown by others to be dependent on the system evaluated (Linke et al., 1999; Morais et al., 2004; Maloney and Feminella 2006; Callanan et al., 2008; Kappes et al., 2010). Annual variation has been less well studied (Jackson and Fureder, 2006) however; it has been shown that understanding annual variation is necessary to improve bioassessment when disturbance is subtle (Huttunen et al., 2012).

We evaluated the utility of currently available macroinvertebrate indices of biotic integrity to assess macroinvertebrate community data from the Big Manistee River watershed data set from the northwest Lower Peninsula of Michigan, USA. The five indices evaluated include the Hilsenhoff (HBI) (family and genus level) biotic indices (Hilsenhoff, 1987, 1988), the benthic community index for the Northern Lakes and Forests (NLFBCI) (Butcher et al., 2003), the Great Lakes Environmental Assessment Survey procedure 51(GLEAS) for macroinvertebrates (Creel et al., 1998) and a Biological Condition Gradient model (BCG) for the Upper Midwest (Gerritsen and Stamp, 2012). The HBI was developed to evaluate organic stream pollution based on genus or family level tolerance values (G-HBI, F-HBI, respectively) for Wisconsin macroinvertebrates. Community-based indices are used to assess biological integrity

using a combination of metrics such as native composition and relative sensitivity to environmental conditions. For example, the NLFBCI is a genus level assessment useful for delineating impaired sites from non-impaired sites in the Northern Lakes and Forests Ecoregion. The GLEAS was developed for use in Michigan with separate family level scoring for each ecoregion in the state resulting in a narrative classification of site scores as excellent, acceptable, or poor. The BCG, originally described by Davies and Jackson (2006), was calibrated for use in the Upper Midwest (Gerritsen and Stamp, 2012) and is based on the relationship between stressors in the environment and corresponding ecological response of the aquatic community indicated with a numeric value from one to six. In this study, macroinvertebrate community data collected through the Little River Band of Ottawa Indians (LRBOI) baseline monitoring and assessment program as well as State of Michigan Department of Environmental Quality (MI-DEQ) macroinvertebrate community data from the trend monitoring program were compiled and analyzed with available indices.

The objectives of this study were to 1) determine if data from multiple agencies could be effectively combined and integrated into a larger watershed dataset and 2) assess concordance of regional indices.

2. Methods

2.1 Study Area

The Big Manistee River watershed (Figure 2.1) is in the northern Lower Peninsula of Michigan, has an area of approximately 490,000 ha, spans 11 counties and includes the 1836 Reservation of the Little River Band of Ottawa Indians (LRBOI). The

watershed is primarily forested (56%), with scrub/shrub and grassland covering 16% and wetlands comprising an additional 13%. There is some agricultural use in the form of grazing and row crops (9%) with developed land covering 6% of the watershed (NLCD, 2006). There are 3191 km of stream within the Big Manistee River watershed (NLCD, 2006). The lower portion of the Big Manistee River is federally recognized as a wild and scenic river with upper portions of the mainstem and sections of tributaries designated by the State of Michigan as Natural Rivers and Blue Ribbon Trout Streams.

2.2 Data Acquisition

The LRBOI Natural Resources Department sampled benthic macroinvertebrates annually, beginning in 2002, using a multihabitat rapid bioassessment protocol (Barbour et al., 1999) to provide data for biological assessment of the watershed. Sampling occurred seasonally in the spring and fall of each year (2002-2011) at four long-term, fixed monitoring sites with reach lengths 40 x stream width. Habitat types (e.g., riffles and pools) were sampled in approximate proportion to their representation of surface area. Macroinvertebrates were preserved and identified in a laboratory. Additionally, three simultaneous replicate samples were collected from nine independent stream reaches in 2009. Three reaches were located on Sickle Creek, Bear Creek, and Pine Creek respectively (n=9), and were separated by a distance of 40 x stream width. Macroinvertebrate data was also acquired from State of Michigan assessments. In 2009 the State of Michigan MI-DEQ conducted an assessment of 23 sites in the Big Manistee River Watershed as part of the state monitoring program, which is on a 5 year watershed rotation (Lipsey, 2010). Macroinvertebrate assessments conducted through this effort followed the Great Lakes and Environmental Assessment Section (GLEAS) Procedure 51

(Creel et al., 1998) protocols. This protocol is used by the State of Michigan for biological assessments throughout the State and is very similar to the LRBOI protocol with proportional habitat being sampled for macroinvertebrate community composition. The GLEAS samples were subsampled to 100 organisms and processed to family in the field while the LRBOI samples were subsampled to 300 organisms and processed to family in a laboratory from 2002-2007 and to genus from 2008-2011.

2.3 Index Applicability

Regional macroinvertebrate indices derived from three spatial scales were evaluated for interpretation of bioassessment data. Macroinvertebrate indices were considered if they were developed for use in the Upper Midwest, the Northern Lakes and Forests Ecoregion (Omernik, 1987) or the State of Michigan (Figure 2.1). Sampling protocols for each index including collection method, timeframe and thermal regime were included in the comparison. The applicability matrix highlights differences in the requirements for each index that met the above criteria (Table 2.1). A total of five index scores were calculated: the NLFBCI, the G-HBI and F-HBI, the GLEAS and a BCG model for the Upper Midwest (Table 2.2). All indices were generally similar in approach, though each had slight variations in sampling protocols and data requirements. All indices were developed for use in cold water systems with multiple habitats in the stream reach sampled during field collections. However, there were some discrepancies in taxonomic resolution requirements (family or genus) and spatial scales of index development. Scale of index development ranged from the entire Upper Midwest to a specific ecoregion within the state of Michigan.

2.4 Data Resolution and Source

The feasibility of integrating datasets was evaluated through comparison of taxonomic resolution and precision of duplicate assessment. To evaluate the effect of taxonomic resolution on index sensitivity, index scores from each of four long-term monitoring sites were calculated from data with genus and family level resolution. Comparisons of truncated data (i.e., family level) to original genus level data were completed for eight paired samples (Spring and Fall for each year, 2008-2011) for each site utilizing a Wilcoxon Signed Rank Test in SigmaPlot version 12.2 (Systat Software Inc., 2012).

Consistency of macroinvertebrate sampling between agencies was verified using data from three duplicate sites sampled by both MI-DEQ and LRBOI in 2009. Both agencies sampled the same reaches independently using their respective protocols. Duplicate sites were of average quality and similar in size to other sites in the watershed assessment. LRBOI data was converted from genus level to family level resolution to match the lowest taxonomic unit available for MI-DEQ data. Index scores generated from these samples were evaluated with a Mantel test (Mantel, 1967) to determine if the two sampling efforts produced similar data (Mazor et al., 2010). Multivariate analysis (Mantel test) was conducted in PC-ORD version 6.0 (McCune and Grace, 2002; McCune and Mefford, 2006) for MI-DEQ and LRBOI macroinvertebrate community data. Mantel's R was used to determine correlation between community compositions of samples. Sorensen distance was used as a dissimilarity measure for the paired LRBOI and MI-DEQ matrices.

2.5 Index Precision

Replicate samples were collected to determine the effect of within site variability on index scores. To assess index score repeatability, three simultaneous replicate samples were collected by LRBOI in 2009 at each of nine independent site locations. Coefficient of variation (CV) and standard deviation (SD) were calculated for each site. The comparability of index scores was evaluated using numeric scales, thresholds, and classification systems among indices. To determine whether the different indices resulted in the correlated assessment of the macroinvertebrate communities, Spearman rank order correlation, were completed in SigmaPlot version 12.2 (Systat Software Inc., 2012), among index scores from 30 independent site assessments conducted throughout the watershed in 2009. Data was aggregated from the 23 MI-DEQ assessments, the four long-term monitoring LRBOI sites and an additional three sites from the replicate sampling (one site from each stream) for a total of 30 sites. For comparisons of NLFBCI, GLEAS and HBI indices, scores were calculated at the family level as that was taxonomic resolution available for MI-DEQ data. To compare the BCG scores, genus level data had to be used, which was available from the four long-term monitoring sites collected seasonally over four years (n=8 for each site). All comparisons with the BCG were made with indices calculated from genus level resolution. The intent of the analysis was to look at scores generated from samples and the relationship of those scores.

2.6 Site Assessments

Index scores from 30 sites throughout the Big Manistee River watershed were assessed to determine if sites scored similarly across indices. Data was aggregated from the 23 MI-DEQ assessments, the four long-term monitoring LRBOI sites and an additional three sites from the replicate sampling (one site from each stream) for a total of

30 sites. Indices were calculated based on family level resolution as that was the lowest taxonomic level available for MI-DEQ data. Proportion of sites in various numeric and categorical rankings of indices were calculated. Site assessments based on the nested indices (GLEAS being locally calibrated and NLFBCI being regionally calibrated) were compared by assessing proportionate divergence of scores away from a specific threshold in the index. A Wilcoxon Signed Rank test was completed in SigmaPlot version 12.2 (Systat Software Inc., 2012) to test the pairs of index scores (GLEAS and NLFBCI) generated from each site and the difference in the proportionate divergence from the threshold.

2.7 Temporal Trends

Seasonal index scores from 2002-2011 from the four long-term monitoring locations were analyzed using a Wilcoxon Signed Rank test in SigmaPlot version 12.2 (Systat Software Inc., 2012) to determine if there were seasonal affects discernible by index score. Spring and fall samples were paired for the analysis and if a season was missing data, that pair was omitted in the statistical analysis. Index scores were also plotted against time to examine seasonal trends by year at each site and evaluate variability (CVs). Each of the index scores were calculated for four long-term monitoring stations in the watershed over ten years to track index output over time.

3. Results

3.1 Data Resolution and Source

The index scores calculated from family and genus level taxonomic data were not significantly different for the HBI (family and genus) at any of the sites (P>0.37 Wilcoxon Signed Rank test) (Figure 2.2a). The index scores from the GLEAS and the

NLFBCI (Figure 2.2b, 2.2c) genus and family level pairs were significantly different at all sites tested (P<0.01Wilcoxon Signed Rank test). The output scores from the GLEAS and NLFBCI were greater (suggesting better condition) when calculated from genus level data.

Family level community composition data from the three sites with both MI-DEQ and LRBOI data were analyzed with a Mantel test in PC-ORD (Table 2.3). The Mantel test compared matrices based on species composition and indicated paired matrices were significantly correlated (P<0.01). Sampling completed by MI-DEQ and LRBOI at three sites indicated index scores were similar. Data collected by these two agencies resulted in identical scores for both the NLFBCI and the GLEAS at one site (PLD). The other two sites only varied by one point for the GLEAS scores and two points (one classification level) for the NLFBCI. The F-HBI scores fell within the same scoring level for PLD (Good) and PLR (Very Good) while BCR was scored as an Excellent site based on MI-DEQ sampling and a Very Good site based on LRBOI sampling.

3.2 Index Precision

Coefficient of variation and standard deviation based on three simultaneous replicate samples were variable in precision depending on the index used (Figure 2.3). The F-HBI and G-HBI ranged from 0 to 9% CV and 0.01 to 0.40 SD and 0 to 8% CV and 0.01 to 0.30 SD respectively. Sites scored with the NLFBCI ranged from 0 to 8% CV and 0 to 2.31 SD. When scored with the BCG, samples ranged from 0 to 25% CV and 0 to 0.58 SD. The GLEAS scores had the most variability and ranged from 0 to 265% CV and 0 to 2.65 SD for specific site replicates. The average coefficient of variation (CV) across sites was under 10% for all metrics except for the GLEAS, which was 68%.

Results were similar for the average standard deviation (SD) with the F-HBI, G-HBI and the BCG all under 0.25 SD. Both the NLFCBI and the GLEAS had higher average standard deviations above 1.0 SD.

Correlations among indices were varied (Figure 2.4). Spearman rank order correlation analysis showed output scores from the G-HBI were not significantly correlated with the scores from the BCG ($P=0.60$, $\rho=0.204$) however, the F-HBI was correlated with both the NLFBCI ($P<0.01$, $\rho=0.553$) and the GLEAS ($P<0.01$, $\rho=0.629$). The NLFBCI was significantly correlated with the GLEAS ($P<0.010$, $\rho=0.698$) and had the highest correlation coefficient. The BCG was weakly correlated with both the NLFBCI ($P<0.01$, $\rho=0.376$) and the GLEAS ($P<0.01$, $\rho=0.350$). The NLFBCI, GLEAS and F-HBI/G-HBI all increased in relation to each other as did the BCG and the NLFBCI and the GLEAS. The indices with no significant correlation were the G-HBI and the BCG.

3.3 Site Assessments

When using indices to assess sites throughout the watershed, the F-HBI, which has seven classifications, generally indicated 26 of the 30 sites ranked above the good threshold while the remaining four sites were in the fair category. Overall, 86% of sites were scored as BCG Tier 3 sites, which correspond to the narrative of the BCG model that states that there is loss of some rare native taxa and some shifts in relative abundance (Davies and Jackson 2006). The GLEAS index scored 50% the sites as “excellent” and 50% as “acceptable”, the top two of the three tiers of classification in the GLEAS and while the NLFBCI also showed a similar trend in ranking based on category (43% in the top and 53% in the fair category); however, discrepancies in scoring appear when actual

score values are analyzed (Figure 2.5). When evaluated based on a proportional measure away from the good or acceptable threshold, the two indices scored the sites differently ($P<0.01$, Wilcoxon Signed Rank test). The GLEAS often scored sites higher, and gave a more favorable view of the watershed as a whole, than the NLFBCI which often produced scores very close to the threshold between fair and good.

3.4 Temporal Trends

Four sites were assessed with the BCG (genus level resolution) seasonally from 2008-2011. Only one site (Sickle) had different scores for fall and spring and the divergence increased through time (Figure 2.6c). All other sites were similar to the oldhouse site (Figure 2.6g) and did not show differences in spring and fall BCG scores though statistical analysis was not completed due low sample size. The HBI, NLFBCI and GLEAS site assessments from 2002 to 2011 (Figure 2.6a,b,d,e,f,h) indicated that none of the output scores showed significant seasonality ($P>0.01$ Wilcoxon Signed Rank test). However, when sites were assessed on an annual basis by season they occasionally indicate different categories of classification (Figure 2.6). For example, in 2003 the GLEAS produced different categories for the analysis of the sickle site (Figure 2.6a). Figure 2.6 indicates that the seasonal sampling, though not significantly different in scores, can lead to different categories of classification with the HBI, GLEAS and the NLFBCI. Sites show temporal variation (CVs over year and season) in output scores with NLFBCI CVs ranging from 14% (Sickle) to 9% (Oldhouse); GLEAS score CVs from 80% (Sickle) to 61% (Oldhouse); BCG score CVs from 33% (Sickle) to 0% (Oldhouse) and HBI score CVs from 29% (Sickle) to 11% (Oldhouse).

4. Discussion

Important decisions about the management and the use of natural resources are often influenced by the estimated condition of a site (USEPA, 2011), which in recent times tends to be based on metric calculations or an index such as an IBI (Karr and Chu, 1999). Results of this study reveal that estimation of site quality can be influenced by the choice of the index and taxonomic resolution of data. We have shown that assessments of environmental condition are generally concordant among different indices; however, vary in magnitude (fair, good, excellent). Thus, which index is used has management implications, and awareness of biases and strengths of each index improves assessment and interpretation of site scores and results within a regional context.

4.1 Data Resolution and Source

The nature (e.g., resolution) and the source of the data (i.e., by whom and how collected) that is used to develop site scores can influence our characterization of different systems and our ability to merge data sets for broader spatial and temporal coverage of system assessment. For example, while there is argument for fine resolution in taxonomy for discriminating subtle ecological signals (Waite et al., 2004, Feio et al., 2006, Hawkins, 2006) there is also indication that biotic index scores may not always be sensitive to taxonomic resolution and for some applications more coarse taxonomic resolution (e.g., family) may be acceptable. The F-HBI has been described as less accurate than the G-HBI (Hilsenhoff, 1988) but there was no significant difference based on family or genus level scoring for the range of values exhibited at the long-term monitoring sites in this study. For both the GLEAS and the NLFBCI the score derived from genus level data was higher than family level data scores. This can be partially

explained by the use of individual metrics based on richness values and the inherent increase in richness values as taxonomic resolution increases. Both Bailey et al. (2001) and Chessman et al. (2007) found small differences in sensitivity between family and genus but determined no appreciable information was gained by the added effort of lower taxonomic resolution for bioassessment. In the Big Manistee watershed, where sites are of generally good quality, greater taxonomic resolution is necessary if the goal is to distill small differences in the relatively high quality sites.

Agencies have historically built bioassessment programs to suit specific monitoring and regulatory needs. These long-term monitoring programs provide consistency that is necessary for tracking trends over time (Herbst and Silldorff, 2006). If methods, data, and results were comparable there would be benefit to collaboration and sharing of data for greater regional determination of environmental conditions. In a survey of methods used by state agencies Carter and Resh (2001) found a large range of field and laboratory methods that could limit effective integration of data sets. Though field and laboratory methods between MI-DEQ and LRBOI varied slightly, comparisons between community composition and index scores derived from this data show similar results. Species composition of samples collected by the two agencies was not significantly different at the three replicate sites; however, index score classification derived from this data did vary at one of the three duplicate sites (e.g., LRBOI data scored the site as “excellent” while the MI-DEQ data scored it as “acceptable” using the GLEAS index. Differences in classification levels could be an issue if these indices were utilized for listing sites as impaired. If management recommendations were based on

bioassessment, effects of variation in classification could be rectified by conducting multiple assessments for a specific site.

4.2 Index Precision

Stream habitat and the associated macroinvertebrate assemblages are spatially variable (Palmer et al., 1997, Lake et al., 2000), yet bioassessment may be based on a single sample or multiple samples from a small area to represent the integrity of a stream reach. In a summary of state agencies that use macroinvertebrates for biomonitoring it was found that 56.1% of programs surveyed (48 States and District of Columbia) conducted replicate sampling for site characterization (Carter and Resh, 2001). Our study found that index scores based on concurrent replicate samples from a site differed in variability depending on index used. The GLEAS index had a much higher average variability (65% CV) and replicates ranged over 5 points for a single site (265% CV) with three replicates, whereas the average CV was below 10% for the NLFBCI, the BCG and the HBI. Mazor et al., (2009) found that average CVs for replicates ranged from 22-27% for IBI scores. Herbst and Sildorff (2006) used CVs of 15-20% as their data quality objective for aggregate multimetric IBI scores at reference sites. Nichols et al. (2006) concluded that a single macroinvertebrate collection would be acceptable if the habitat was not variable and was in good condition, but if there was a higher level of habitat heterogeneity then multiple collections were necessary. Depending on the index used there is evidence that replicate samples are necessary for a more accurate assessment of condition. Specifically, with higher variability in scores generated by the GLEAS samples we would advocate using multiple samples for assessments that lead to management decisions if using the GLEAS.

Indices were concordant except when comparing the BCG with the G-HBI. The range of condition determined for our sites was limited in scale for both the G-HBI and the BCG. Also, the HBI was developed to indicate issues from organic pollution and this tolerance-based index may not be comparable to scoring using community composition and comparison to reference conditions. Spearman correlation coefficients for significant relationships among indices (NLFBCI, GLEAS, F-HBI, G-HBI and BCG) were low (r=0.35-0.698) compared with other previous research. Herbst and Sildorff (2006) found moderate correlations (Spearman's r=0.70-0.86) among indices from sites in the eastern Sierra Nevada of California. Hawkins et al. (2010), also found moderate correlations (Pearson's r= 0.63-0.92) among index scores at sites within the Columbia River basin. When evaluations were completed spanning seven countries throughout Europe, Birk and Hering (2006) found more variable correlations (Pearson's r=0.20-.077) among indices. Five streams with 11 sites tested in Australia showed moderate correlations (r=0.66-0.89) between bioindicators (Nichols et al., 2010). Because results from the indices were concordant, sites scored with the NLFBCI, GLEAS, or the BCG will generally reflect similar patterns of biotic condition if tracked over time.

4.3 Site Assessments

Based on the region for which they were calibrated, all indices examined in this study were appropriate for use with Big Manistee River watershed dataset. Meador et al. (2008) found in a study of the western US that regional IBIs can work at multiple spatial scales and corroborate those developed at more local geographic scale. Over three geographically separate regions in Oregon and California, models have been developed that contain metrics that function well across ecoregions (Waite et al., 2010). However,

locally calibrated indices have also been found to outperform regional indices for site specific assessments (Mykrä et al., 2008; Ode et al., 2008). Overall, the GLEAS (locally calibrated index) and the NLFBCI (regional index) assessment scores provide a favorable view of the watershed where approximately half of the sites were in the top level classification for both indices. However, when assessed using the proportional divergence of the site score away from the highest threshold of condition, the NLFBCI index generally scored sites lower than the GLEAS. This may be an artifact of the taxonomic resolution of the dataset and an indication that genus level resolution is needed for valid assessment using the NLFBCI. Considering the nested nature of the indices, the local calibration of the GLEAS may indicate that, of the Michigan NLF ecoregion, sites in the Big Manistee River watershed sites rank well comparatively. The NLFBCI may give a better overall evaluation of how sites rank in relation to the rest of the ecoregion. Expanding further, the BCG scored six of the seven watershed sites as Tier three and the remaining site as Tier four. This model may give better insight as to the condition of sites relative to a larger regional picture including the state, ecoregion and Upper Midwest. These results exemplify that care must be taken in choosing indices as well as interpreting results from the scores.

4.4 Temporal Trends

Over 10 years, paired seasonal index scores were not significantly different; however, on an annual basis there were differences in seasonal scores that could lead to variation in classification of stream sites. We detected no significant trend where one season produced a consistently higher index score. This is in contrast to other studies that found consistent differences in seasonal index scores (Linke et al., 1999; Callanan et al.,

2008; Kappes et al., 2010). Others have found multimetric scores to be insensitive to season and showed no differences in scores based on season (Morais et al., 2004; Maloney and Feminella 2006). Index scores for long-term monitoring sites in the Big Manistee River watershed showed seasonal differences in classification on an annual basis sufficient to alter interpretation of system condition. Ensuring consistency in sampling season is important for accurate assessment and comparisons to reference sites especially if scores are to be used for designating impairment or management action.

Jackson and Fureder (2006), in a review of bioassessment papers (1987-2004), found only 46 papers with long-term (>5 yr) data sets. They stressed the need for long-term consistent research to accurately describe the variation, type, magnitude and direction of response signals. Long-term assessments are necessary to begin to understand the natural fluctuations and track anthropogenic effects. There have been conflicting reports in the literature as to the stability of assemblage composition over time. Temporal changes have been observed in Mediterranean systems (Feio et al., 2010) and in pristine Alaskan streams (Milner et al., 2006). In reference sites Nichols et al. (2010) found persistent communities and no significant change in bioindicators over 15 years. Mazor et al. (2009) determined through a twenty-year assessment of four sites that a snapshot approach to bioassessment could lead to incorrect conclusions if natural fluctuations are not taken into account. Huttunen et al. (2012) also found that even with low annual variation there were discrepancies in index scores describing ecological status of sites and that use of one year of data would be problematic for making informed management decisions.

By tracking index scores over multiple years, variation over time is revealed. Conclusions based on a one year assessment would likely be very different than conclusions based on data from 2008-2011 (BCG data). This pattern is even more pronounced when specific watershed sites were evaluated over 10 years utilizing the HBI, NLFBCI and the GLEAS. Observing similar patterns in multiple indices through time can distinguish long-term natural variability as compared to anthropogenic effects. Long-term variability has not been well studied in stream systems, though its suggested importance is well-documented (Jackson and Fureder, 2006). Documenting long-term variability will improve assessment of biological quality specifically where disturbance is subtle (Huttunen et al., 2012).

The results found here highlight the benefits and difficulty of utilizing multiple indices developed at different scales with geographically small data sets. Aggregating data from multiple agencies, assessing comparability issues and ensuring index scores are comparable is necessary for an expanded scope of site characterization. The benefit of being able to assess local sites at multiple scales and broadening the scope of assessments leads to a better understanding of ecological condition. The goal of bioassessment is to evaluate the ecological condition of a site, reach, watershed or region. Using multiple lines of evidence in the form of multiple indices will help assess the condition of a site and put it into a larger regional perspective. If indices and thresholds are to be used in management decision making process, replicates or multiple samples over time should be used due to variance in index scores. Long-term assessments are necessary to evaluate site condition and assess natural fluctuation. Each index used in this study was originally

developed for different geographic scales and we found that use of the NLFBCI provides an effective assessment of the Big Manistee River watershed.

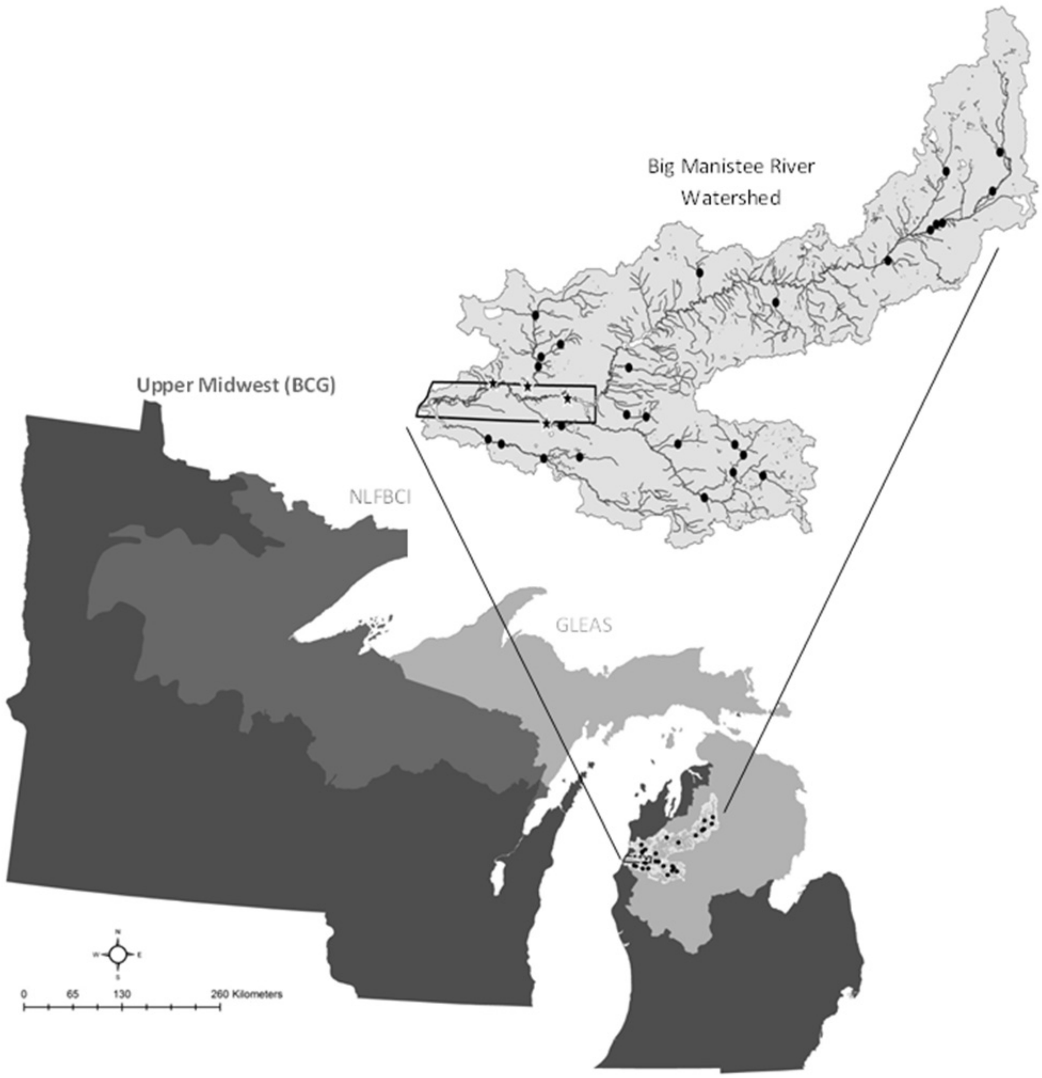

Figure 2.1. Location of watershed sampling sites within the Big Manistee River watershed located in Michigan, USA. Area depicted includes Upper Midwest (dark grey) where the Biological Condition Gradient (BCG) model was calibrated, Northern Lakes and Forests ecoregion (grey) where a biotic condition index (NLFBCI) was calibrated, the Michigan portion of the ecoregion (light grey) where the Great Lakes Environmental Assessments Section (GLEAS) index was calibrated and the Big Manistee River watershed where data was collected. The dark rectangle in the watershed is the Little River Band of Ottawa Indians 1836 exterior reservation boundaries where long term data was collected. All dots in the watershed are sampling locations with long term stations identified with a star.

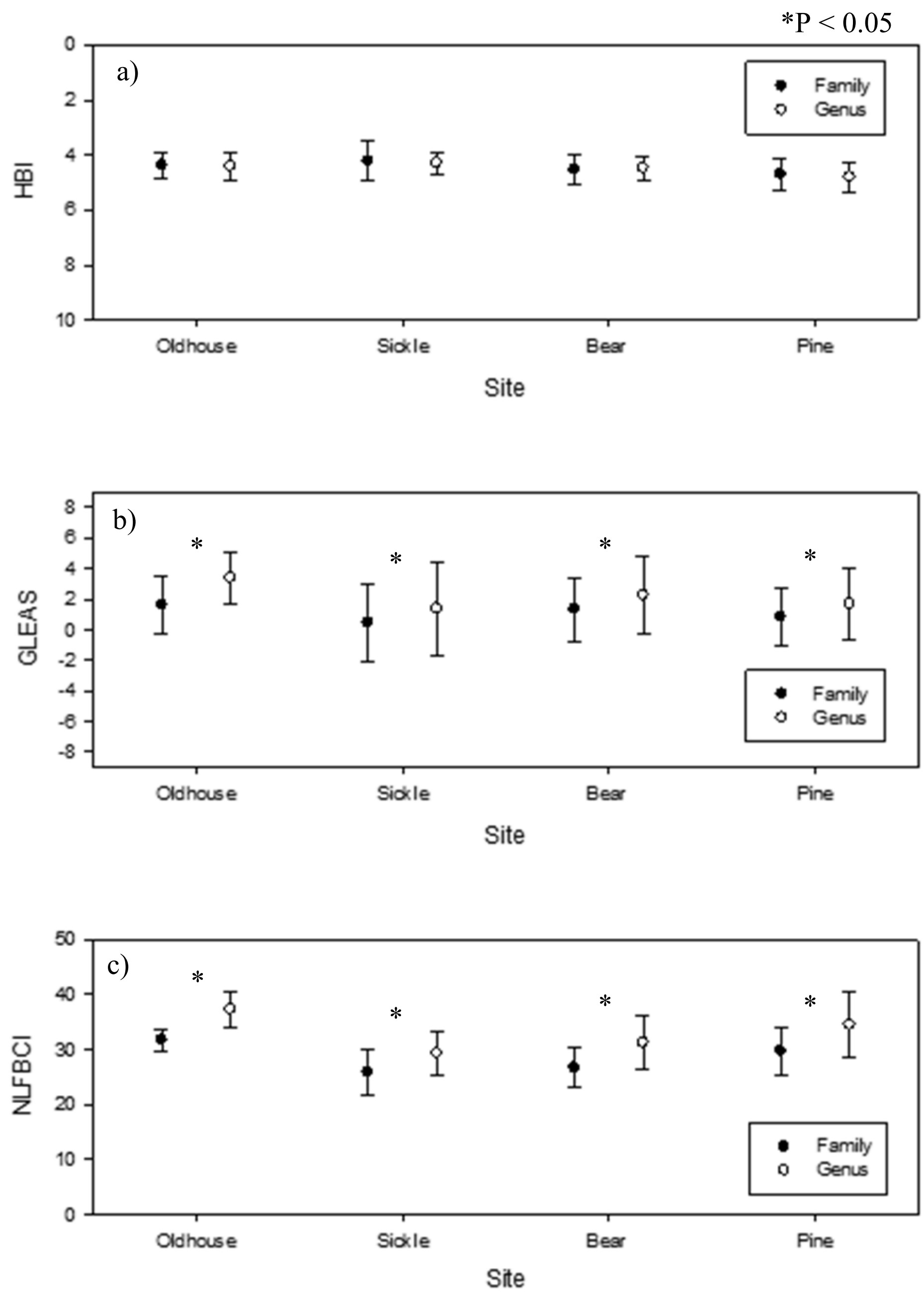

Figure 2.2. Comparison of family and genus level scores for three indices: a) Hilsenhoff biotic index, b) Great Lakes Environmental Assessment Section index and c) Northern Lakes and Forests benthic community index. Paired scores (genus and family outputs) at each site were tested using 8 samples at each site collected from 2008-2011 (Wilcoxon Signed Rank test).

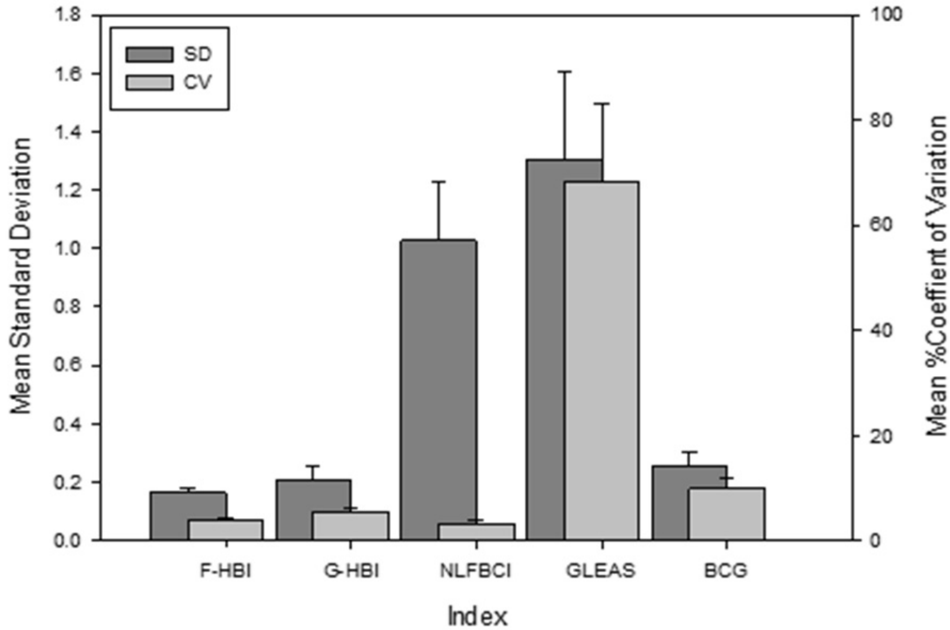

Figure 2.3. Mean (+ 1 S.E.) standard deviation and coefficient of variation among replicate samples (3) collected at nine locations in 2009. Black bars represent the mean standard deviation (SD) of the replicates for each index across the nine sites. The grey bars represent the mean percent coefficient of variation (CV).

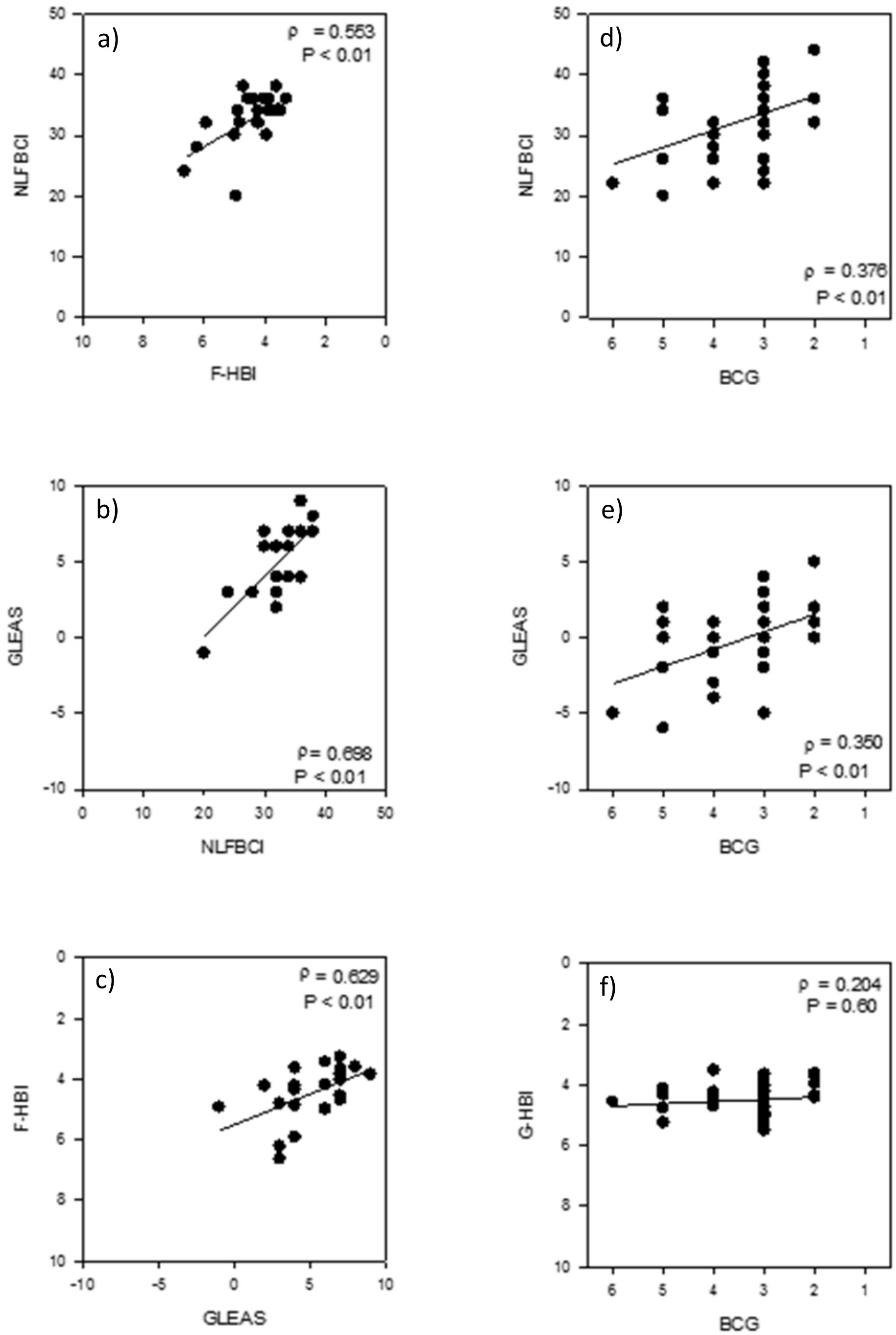

Figure 2.4. Spearman's rank order correlation (ρ) between index scores. Figure labels a,b, and c utilized family level data from the 30 sites taken throughout the watershed in 2009. To analyze the BCG genus level data were required and therefor data (n=32) used were from seasonal sampling that occurred at four sites from 2008-2011 (Figure d,e and f).

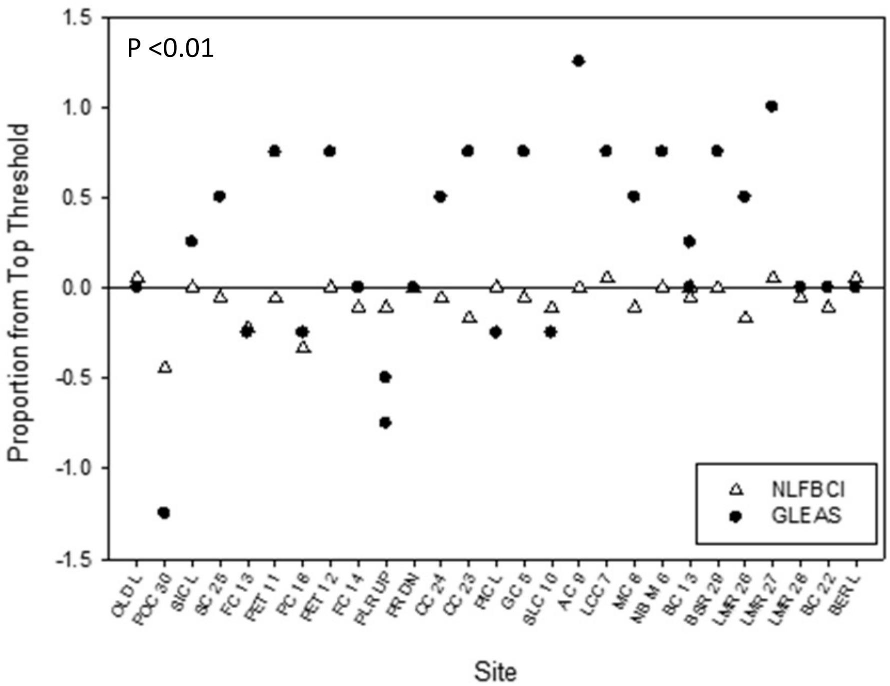

Figure 2.5. Representation of the proportion away from the good or acceptable threshold each site scored with the Northern Lakes and Forest benthic community index (NLFBCI) and the Great Lakes environmental assessment section index (GLEAS). This allows for a comparison of the relative scores for the two community indices and how assessment scores rate the condition of a given site. When paired site scores were analyzed with the Wilcoxon Signed Rank test there was a significant difference ($P<0.01$). The horizontal line indicates the good or acceptable threshold for the indices.

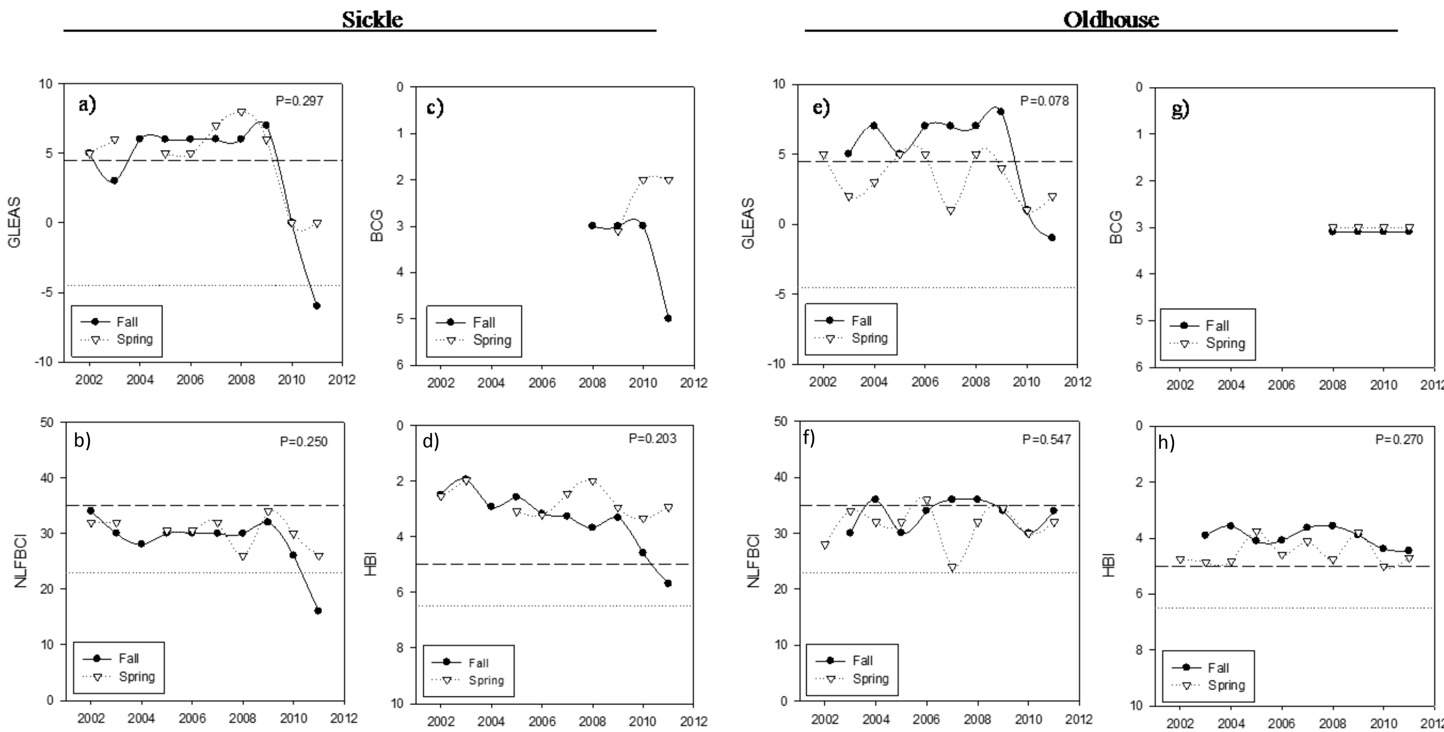

Figure 2.6. Spring and fall index scores for two long-term monitoring sites in the Big Manistee River watershed from 2002-2011. The Great Lakes environmental and assessment section index (GLEAS) scores (a,e), the Northern Lakes and Forests ecoregion benthic community index (NLFBCI) (b,f) and the Hilsenhoff biotic index (HBI) (d,h) spring and fall output pairs were analyzed using a Wilcoxon Signed Rank test which did not indicated differences in the seasonal index scores. The biological condition gradient (BCG) data was only available for four years and did not provide enough data to run the statistical analysis (c,g). Only the most variable site (sickle) and the least variable site (oldhouse) are shown for reference as they represent the extremes of the data. Long dashed horizontal lines in each panel represent the top threshold for each index while the fine dotted line represents the bottom threshold for each index.

Table 2.1. Index applicability matrix describing core attributes for each of five indices used for analysis: Hilsenhoff biotic index (F-HBI, G-HBI), Northern Lakes and Forests benthic community index (NLFBCI) , Great Lakes Environmental Assessment Section index (GLEAS) and the biological condition gradient (BCG). The BCG has different models for cool and cold water streams based on mean July temperatures. We used the cold water model.

	F-HBI	**G-HBI**	**NLFBCI**	**GLEAS**	**BCG**
Development Region	WI	WI	Ecoregion	MI	MN,WI,MI
Taxonomic Resolution	Family	Genus	Genus	Family	Genus
Sampling Protocol*	Multihabitat[a]	Multihabitat[a]	Multihabitat[b]	GLEAS 51[c]	RBP[d]
Temperature Regime	Regional	Regional	Regional	Ecoregion	(<17.5°C)

*References for sampling protocols: a) Hilsenhoff 1987, b) Chirhart 1998, c) Creel et al 1998, d) Gerritsen and Stamp 2012

Table 2.2. Numeric index scores and associated classification levels for the Hilsenhoff biotic index (HBI-F, HBI-G), the Northern Lakes and Forests benthic community index (NLFBCI), the Great Lakes and environmental assessment section index (GLEAS) and the numeric levels for the biological condition gradient (BCG) model for the Upper Midwest. Color gradations indicate groupings based on similarities in classification levels.

HBI (F)[a]		HBI (G)[b]		NLFBCI[c]		GLEAS[d]		BCG[e]
0-3.75	Excellent	0-3.50	Excellent					
3.76-4.25	Very Good	3.51-4.50	Very Good	36-50	Good	5 - 9	Excellent	1
4.26-5.0	Good	4.51-5.50	Good					2
5.01-5.75	Fair	5.51-6.50	Fair	24-34	Fair	-4 - 4	Acceptable	3
5.76-6.50	Fairly Poor	6.51-7.50	Fairly Poor					4
6.51-7.25	Poor	7.51-8.50	Poor	10-22	Poor	-9 - -5	Poor	5
7.26-10.0	Very Poor	8.51-10.00	Very Poor					6

*References for scoring: a) Hilsenhoff 1987, b) Hilsenhoff 1988, c) Butcher et al. 2003, d) Creel et al. 1998, e) Jackson and Davies 2006

Table 2.3. Sampling of same sites (PLD, PLU and BCR) independently completed by agency staff (LRBOI, MI-DEQ) and evaluated for index scores and Mantel's r based on family level community composition data. Index scores include the family level Hilsenhoff biotic index (F-HBI), the northern lakes and forests ecoregion benthic community index (NLFBCI) and the Great Lakes and Environmental Assessment Section (GLEAS) index.

	PLD		PLU		BCR	
	LRBOI	MI-DEQ	LRBOI	MI-DEQ	LRBOI	MI-DEQ
F-HBI	4.32	4.36	3.99	4.23	4.01	3.63
NLFBCI	36	36	32	34	36	34
GLEAS	4	4	1	2	5	4
Mantel's r	0.151		0.353		0.109	

References

Bailey, R., Norris, R.H., Reynoldson, T.B., 2001. Taxonomic resolution of benthic macroinvertebrate communities in bioassessments. J. N. Am. Benthol. Soc. 20, 280-286.

Barbour, M.T., Gerritsen, J., Snyder, B.D., Stribling, J.B., 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish. 2nd ed., U.S. Environmental Protection Agency, Office of Water, EPA Report 841-B-99-002, Washington, DC.

Birk, S., Hering, D., 2006. Direct comparison of assessment methods using benthic macroinvertebrates: a contribution to the EU Water Framework Directive intercalibration exercise. Hydrobiologia 566, 401–415.

Butcher, J.T., Stewart, P.M., Simon T.P., 2003. A benthic community index for streams in the Northern Lakes and Forests Ecoregion. Ecol. Indic. 3, 181-193.

Cao, Y., Hawkins, C.P., Storey, A.W., 2005. A method for measuring the comparability of different sampling methods used in biological surveys: implications for data integration and synthesis. Fresh. Biol. 50,1105-1115.

Cao, Y., Hawkins, C.P., 2011. The comparability of bioassessments: a review of conceptual and methodological issues. J. N. Am. Benthol. Soc 30, 680-701.

Callanan, M., Baars, J.R., Kelly-Quinn, M., 2008. Critical influence of seasonal sampling on the ecological quality assessment of small headwater streams. Hydrobiologia 610, 245–255.

Carter, J.L., Resh, V.H., 2001. After site selection and before data analysis: sampling, sorting, and laboratory procedures used in stream benthic macroinvertebrate monitoring programs by USA state agencies. J. N. Am. Benthol. Soc 20, 658–682.

Creel, W.S., Hanshue, S., Kosek, S,. Oemke, M., Walterhouse, M., 1998. GLEAS Procedure 51 Metric Scoring and Interpretation. Chapter 25 in Schneider, J.C. (ed) 2000. Manual of fisheries survey methods II: with periodic updates. Michigan Department of Natural Resources, Fisheries Special Report 25, Ann Arbor MI.

Chessman, B., Williams, S., Besley, C., 2007. Bioassessment of streams with macroinvertebrates: effect of sampled habitat and taxonomic resolution. J. N. Am. Benthol. Soc. 26, 546-565.

Chirhart, J., 1998. Invertebrate sampling procedures for Northern Lakes and Forests

streams. In: Simon, T.P., Stewart, P.M. (Eds.). Standard Operating Procedures for Development of Watershed Indicators in REMAP: Northern Lakes and Forests Streams. U.S. Environmental Protection Agency, Region 5 Water Division, Watershed and Non-Point Source Branch, Chicago, IL.

Davies, S.P., Jackson, S.K., 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. Ecol. Appl. 16, 1251–1266.

Feio, M.J., Reynoldson, T.B., Craca, M.A.S., 2006. The influence of taxonomic level on the performance of a predictive model for water quality assessment. Can. J. Fish. Aquat. Sci. 63, 367–376.

Feio, M.J., Coimbra C.N., Craca, M.A.S, Nichols S.J., Norris R.H., 2010. The influence of extreme climatic events and human disturbance on macroinvertebrate community patterns of a Mediterranean stream over 15 years. J. N. Am. Benthol. Soc., 29 1397-1409.

Gerth, W.J., Herlihy, A T., 2006. Effect of sampling different habitat types in regional macroinvertebrate bioassessment surveys. J. N. Am. Benthol. Soc. 25, 501–512.

Gerritesn, J., Stamp, J., 2012. Calibration of the biological condition gradient (BCG) in cold and cool waters of the Upper Midwest: BCG-based indexes (BCG-I) for fish and benthic macroinvertebrate assemblages. Tetra Tech, Inc. Owings Mills, MD.

Hawkis, C.P., Cao, Y, Roper, R., 2010. Method of predicting reference conditions affects the performance and interpretation of ecological indices. Fresh. Biol. 55, 1066–1085.

Hawkins, C.P., 2006. Maintaining and restoring the ecological integrity of freshwater ecosystems: Refining biological assessments. Ecol. Appl. 16,1249-1311.

Herbst, D.B., Silldorff, E.L., 2006. Comparison of the performance of different bioassessment methods: similar evaluations of biotic integrity from separate programs and procedures. J. N. Am. Benthol. Soc. 25**,** 513-530.

Hilsenhoff, W.L. 1987. An improved biotic index of organic stream pollution. Great Lakes Entomol. 20, 31-39.

Hilsenhoff, W.L. 1988. A modification of the biotic index of organic stream pollution to remedy problems and permit its use throughout the year. Great. Lakes. Entomol. 31, 1-12.

Hughes, R.M., Peck, D.V., 2008. Acquiring data for large aquatic resource surveys: the art of compromise among science, logistics and reality. J. N. Am. Benthol. Soc. 27, 837-859.

Huttunen, K.L., Mykrä, H., Muotka, T., 2012. Temporal variability in taxonomic completeness of stream macroinvertebrate assemblages. Fresh. Sci. 31**,** 423-441.

Jackson, J.K., Fureder, L., 2006. Long-term studies of freshwater macroinvertebrates: a review of the frequency, duration and ecological significance. Fresh. Biol. 51**,** 591-603.

Kappes, H., Sundermann, A., Haase, P., 2010. High spatial variability biases the space-for-time approach in environmental monitoring. Ecol. Indic. 10**,** 1202-1205.

Karr, J.R., Chu E.W., 1999. Restoring life in running waters: better biological monitoring. Island Press. Washington, DC.

Lake, P.S., Palmer, M.A., Biro, P., Cole, J., Covich, A.P., Dahm, C., Gibert, J., Goedkoop, W., Martens, K., Verhoeven J., 2000. Global change and the biodiversity of freshwater ecosystems: Impacts on linkages between above-sediment and sediment biota. BioScience 50,1099–1107.

Linke, S., Bailey, R.C., Schwindt, J., 1999. Temporal variability of stream bioassessment using benthic macroinvertebrates. Fresh. Biol. 42, 575–584.

Lipsey, T., 2010. MI-DEQ staff report. Biological and water chemistry surveys of selected stations in the Manistee, Little Manistee, and Big Sable river watersheds Grand Traverse, Kalkaska, Lake, Manistee, Mason, Osceola, and Wexford Counties, Michigan. MI/DNRE/WB-10/016.

Maloney, K. O., Feminella J.W., 2006. Evaluation of single- and multi-metric benthic macroinvertebrate indicators of catchment disturbance over time at the Fort Benning Military Installation, Georgia, USA. Ecol. Indic. 63, 469-84.

Mantel, N.A., 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27, 209–220.

Mazor, R.D., Schiff K., Ritter K., Rehn, A., Ode P., 2010. Bioassessment tools in novel habitats: an evaluation of indices and sampling methods in low-gradient streams in California. Environ. Monit. Assess. 167, 91-104.

Mazor, R. D., Purcell A. H., Resh V. H., 2009. Long-term variability in bioassessments: a twenty-year study from two northern California streams. Environ. Manage. 43, 1269–1286.

McCune, B., Grace, J.B., 2002. Analysis of ecological communities. Mjm Software Design, Gleneden Beach, OR, US.

McCune, B., Mefford, M.J., 2006. PC-ORD: multivariate analysis of ecological data. MJM Software. Version 5.14. Gleneden Beach, Oregon.

Meador, M.R., Whittier T.R., Goldstein, R.M., 2008. Evaluation of an index of biotic integrity approach used to assess biological condition in western US streams and rivers at varying spatial scales. Trans. Am. Fish. Soc. 137, 13–22.

Milner, A.M., Conn S.C., Brown L.E., 2006. Persistence and stability of macroinvertebrate communities in streams of Denali National Park, Alaska: implications for biological monitoring. Fresh. Biol. 51, 373-387.

Morais, M., Pinto, P., Guilherme, P., Rosado, J., Antunes, I., 2004. Assessment of temporary streams: the robustness of metric and multimetric indices under different hydrological conditions. Hydrobiologia. 516, 229-49.

Mykrä, H., Aroviita, J., Kotanen, J., Hämäläinen H., Muotka, T., 2008. Predicting the stream macroinvertebrate fauna across regional scales: influence of geographical extent on model performance. J. N. Am. Benthol. Soc. 27, 705–716.

Nichols, S.J., Robinson, W.A., Norris R.H., 2006. Sample variability influences on the precision of predictive bioassessment. Hydrobiologia. 572, 215-233.

Nichols, S.J., Robinson, W.A., Norris R.H., 2010. Using the reference condition maintains the integrity of a bioassessment program in a changing climate. J. N. Am. Benthol. Soc. 29, 1459-1471.

NLCD, 2006. National Oceanic and Atmospheric Administration, Coastal Services Center. 1995-present. The Coastal Change Analysis Program (C-CAP) Regional Land Cover. Charleston, SC: NOAA Coastal Services Center. Accessed at www.csc.noaa.gov/digitalcoast/data/ccapregional.

Omernik, J.M., 1987. Ecoregions of the conterminous United States. map (scale 1:7,500,000). Ann. Assoc. Am. Geogr. 77, 118-125.

Ode, P.R., Hawkins C.P., Mazor, R.D., 2008. Comparability of biological assessments derived from predictive models and multimetric indices of increasing geographic scope. J. N. Am. Benthol. Soc. 27, 967–985.

Palmer, M.A., Covich A.P., Findlay B.J., Gilbert J., Hyde, K.D., Johnson R K., Kairesalo T., Lake S., Lovell, C.R., Naiman, R. J., Ricci C., Sabater, F., 1997. Biodiversity and ecosystem processes in freshwater sediments. Ambio 26, 571–577.

Rehn, A.C., Ode, P.R., Hawkins, C.P., 2007. Comparisons of targeted-riffle and reach-

wide benthic macroinvertebrate samples: implications for data sharing in stream-condition assessments. J. N. Am. Benthol. Soc. 26, 332–348.

Systat Software Inc., 2012. Systat Software SigmaPlot Users Guide 12.2. San Jose, CA

USEPA (US Environmental Protection Agency). 2011. A Primer on using biological assessments to support water quality management. EPA-810-R-11-01. Office of Science and Technology, Office of Water, Washington, DC.

Waite, I.R., Herlihy, A.T., Larsen, D.P., Urquhart, N.S., Klemm, D.J., 2004. The effects of macroinvertebrate taxonomic resolution in large landscape bioassessments: an example from the Mid-Atlantic Highlands, USA. Fresh. Biol. 49, 474-489.

Waite, I.R., Brown, L.R., Kennen, J.G., May, J.T., Cuffney, T.F., Orlando, J.L., Jones, K.A., 2010. Comparison of watershed disturbance predictive models for stream benthic macroinvertebrates for three distinct ecoregions in western US. Ecol.Indic. 10, 1125-1136.

Chapter 3.[2] Awareness of bias in fish indices of biotic integrity improves interpretation of bioassessment

Introduction

Biotic integrity has been defined as the ability to support an integrated, adaptive community of organisms with species composition, diversity, and functional organization comparable to that of a natural assemblage of the region (Frey 1977; Karr and Dudley1981; Karr et al. 1986). As such, a system with intact biotic integrity supports a complex of native biodiversity with natural processes and services (Karr and Chu 1999). For the past thirty years, assessment of the ecological condition of aquatic systems has routinely been accomplished using various modified and regionally calibrated indices of biotic integrity (IBIs; Karr 1981; US EPA 2002). Though criticisms of the IBI have been noted (Suter 1993), citations of Karr (1981) have increased since its original publication through 2007 with over 700 papers being documented (Ruaro and Gubiani 2013). IBIs have been found to be effective monitoring tools that can provide qualitative information to communicate assessments to the public and policy makers as well as provide quantitative data for hypothesis testing (Fore et al. 1994).

Fish assemblages can be ideal integrated indicators of ecological integrity given that they are relatively easy to collect and they can have unique species- and population-specific responses to environmental conditions that are reflected in their relative abundance and composition (Karr et al. 1986; Barbour et al. 1999). Indices based on fish community composition provide an approach for classifying stream reaches based on a continuum of biological condition and these measures of stream condition integrate

[2] The material contained in this chapter has been submitted to *the North American Journal of Fisheries Management*.

multiple influences making them particularly useful for assessing overall stream integrity (Allen 2004). Multiple fish based IBIs (Lyons et al. 1996; Mundahl and Simon 1999; Lyons 2012) and biological condition gradient (BCG) models (Gerritsen and Stamp 2012) have been developed and validated to assess ecological integrity in the Laurentian Great Lakes Region. With multiple options for biomonitoring, critical examination of applicability and interpretation is necessary to produce accurate and biologically meaningful assessments of stream condition.

A primary goal of bioassessment is to characterize the site and determine where it falls along a continuum of impairment; however, selection of the appropriate index to assess the site is influenced by the physical conditions of the site within the regional range. Biotic and abiotic factors interact in complex ways at multiple spatial scales to influence organism abundance and distribution in fluvial systems (Hynes 1975; Vannote et al. 1980; Poff and Allan 1995; Faush et al. 2002). Stream hydrology and thermal regime are important factors and they are influenced by large scale drivers such as climate and surficial geology (Roth et al. 1996; Allan et al. 1997; Wang et al. 1997). While large scale influences such as geology provide a template for the range of potential habitat conditions, local factors such as groundwater input, are important site specific drivers of local habitat quality (*sensu* Tonn et al. 1990). Climatic variation, predation and competition (Moyle and Cech 1996) and the availability and quality of thermal habitat (Magnuson et al. 1979; Connor et al. 2003; Sloat et al. 2005) also modify the resulting fish community and therefore the resulting IBI score. The combination of large scale landscape and small scale local conditions produce a hierarchical filter that influences or

limits assemblage composition (Frissell et al. 1986; Tonn et al. 1990; Gregory et al. 1991; Jackson et al. 2001).

In Upper Midwest streams, flow pattern and thermal regime are strong predictors of local fish assemblages (Lyons 1996; Zorn et al. 2002; Wang et al. 2003; Wehrley et al. 2003; Zorn et al. 2008). Thermal regimes of streams in this region have been described to fall along a gradient from coldwater to warmwater with some uncertainty about the bounds of coolwater classifications between the two extremes (Wehrly et al. 2003; Lyons et al. 2009). Coldwater stream fish communities are generally dominated by a low number of species restricted by thermal thresholds and are dominated by the families Salmonidae and Cottidae (Lyons 1996; Lyons et al. 1996; Wehrly et al. 2007). Warmwater systems are generally more diverse and fish assemblages may be dominated by individuals in the families Centrarchidae, Catostomidae, Ictaluridae, and Cyprinidae (Lyons 1996).

Streams with different thermal and flow regimes require different assessment tools (Karr and Chu 1999) and interannual variation in temperatures of streams with mean temperatures near the upper or lower thresholds for thermal classifications can lead to difficulty or error in determination of their thermal classification, therefore an ideal classification would be founded on a long-term record (Wehrley et al. 2003). Lyons (2012) indicates that modeling of temperature regime at larger landscape scales may also be an appropriate predictor of a streams true potential as a cool or coldwater system. Further, there are unique considerations in these systems because indices that give higher scores for diversity in warmwater systems do not work well in coldwater systems beacause a lack of diversity is associated with high quality coldwater systems and

degradation is generally marked by an increase in diversity (Karr 1999, Lyons et al. 2009). A site must be classified accurately and an appropriate index calculated to serve as an effective indicator and be of use to managers, decision makers and community members. Application of inappropriate indices can underestimate biotic integrity (Wang et al. 2003; Baker et al. 2005) and lead to coolwater streams being mis-classified as degraded coldwater streams.

We assessed the performance of five indices developed for the Upper Midwest using data from sites in the Big Manistee River watershed, in the Lower Peninsula of Michigan. The original Wisconsin coldwater IBI (Lyons et al. 1996) uses five metrics and was developed for use in coldwater Wisconsin streams (maximum daily mean <22°C). Some concern has been noted that due to a reduced number of metrics small differences in biotic integrity may be difficult to detect (Lyons et al. 1996). Thereafter, Mundahl and Simon (1999) used data from coldwater (<22°C) streams in Michigan, Minnesota and Wisconsin to develop and test an upper Midwest IBI that uses 12 metrics. This was initially compared to the Wisconsin coldwater IBI with some discrepancies in classification being noted. Lyons (2012) developed an index based on fish assemblages in coolwater Wisconsin streams that are cool-cold transitional (20.7-22.5°C max daily mean). Lastly, two biological condition gradient (BCG) models were developed for use in coolwater (17.5-19.0°C July mean) and coldwater (<17.5 °C) stream systems of the upper Midwest (Gerritsen and Stamp 2012). The BCG models differ from traditional IBI's in that they aim to provide a consistent assessment of ecological integrity that can be applied across regions and are based not on reference sites but on a theoretical, pristine condition (Davies and Jackson 2006). All of the above indices assess ecosystem integrity

either based on reference conditions or a defined condition and place current sites on a continuum of ecological condition. We evaluated index performance including annual and seasonal trends, correlation and concordance of assessments based on index scores, and explored potential biases that might result from each index.

Methods

Four streams within the Big Manistee River watershed in the northern Lower Peninsula of Michigan were assessed for this study (Figure 1). The watershed has an area of approximately 490,000 ha and spans 11 counties and the 1836 exterior reservation boundary of the Little River Band of Ottawa Indians. This watershed is typical of many Michigan systems having streams that are sand and gravel dominated, low gradient, and hydrologically-stable with temperatures influenced by groundwater inputs (Seelbach et al. 1997; Wehrly et al. 2006). Hendrickson and Doonan (1972) estimated that 90% of the annual discharge from the Big Manistee River watershed is from groundwater sources that are common in the outwash plain the watershed drains. The watershed is primarily forested (56%), with scrub/shrub and grassland covering 16% and wetlands comprising an additional 13% (NLCD, 2006). There is minor agricultural use in the form of grazing and row crops (9%) with developed land covering 6% of the watershed (NLCD, 2006). Sampling locations were located in two large sub-watersheds (Bear Creek and Pine Creek) and in two mainstem tributaries (Figure 3.1). The lower portion of the Big Manistee River is federally recognized as a wild and scenic river and upper portions of the mainstem and sections of tributaries are designated by the State of Michigan as Natural Rivers and Blue Ribbon Trout streams.

Data collection.- Fish assemblage data have been collected annually by the LRBOI biological assessment program since 2002 at four long-term monitoring sites each located in a separate sub-watershed (Figure 3.1). Sampling station lengths were set as 40x stream wetted width if greater than a minimum length of 120 m and less than a maximum length of 400 m. Backpack electrofishing was conducted using a Smith-Root LR-14 unit at summer base flow between June and September. Community composition was assessed with all fish identified, measured, and counted. Following the same protocol, additional sampling was conducted in three streams (five sites per stream) in May and August 2005 to characterize seasonal differences in community structure. For broader spatial coverage, a total of 26 additional sites throughout the watershed were also sampled in 2007 during summer base flow. These sites were located in the Bear Creek (n=8), Pine Creek (n=12), Sickle Creek (n=4), and Oldhouse Creek (n=2) (Figure 3.1). Temperature data were collected with Onset® HOBO ® water temperature pro loggers set for hourly increments from May through September, deployed periodically from 2002 through 2012 at the long-term monitoring stations. Mean daily July temperatures (°C) were calculated for each year sampled.

Indices.- Regional fish indices were considered applicable for, 1) use in the Upper Midwest and the Northern Lakes and Forest Ecoregion (Omernik 1987), 2) streams with appropriate temperature regimes, and 3) sampling protocols that matched the collection method and timeframe of the data. Due to the transitional temperatures found in the Big Manistee watershed, indices for both coolwater and coldwater stream classifications were

considered. Regional indices calculated from community composition included IBIs for coldwater steams in Wisconsin (Lyons et al. 1996) and Upper Midwest (Mundahl and Simon 1999), a cool-cold transitional IBI (Lyons 2012) and both coolwater and coldwater BCG models for the Northern Lakes and Forests Ecoregion (Gerritsen and Stamp 2012). Throughout this paper these will be referred to as Lyons 96, M&S, Lyons Cool, BCG Cool and BCG Cold, respectively. All indices were calculated as directed by the original publications. The ability for direct comparisons of scores (i.e., assessments of condition) across indices was evaluated by comparing numeric scale, thresholds and classification systems among indices. The indices vary in range of possible numeric scores, thresholds for classification groupings, and how they were designed to be categorically interpreted (Table 3.1). Though the three IBIs (Lyons 96, M&S, and Lyons Cool) had similar categories, the numeric thresholds that apply to these categories are not uniform. The BCG models rely on numeric categories that refer to a certain condition associated with a particular score (tier). For example, a tier 4 stream is described as having the condition of moderate changes in structure of the biotic community with minimal changes in ecosystem function (Davies and Jackson 2006).

Index performance analysis- To determine long-term temporal variability, mean coefficients of variation (CVs) were calculated for scores derived from summer surveys conducted from 2002-2010. For three streams, sampled in two seasons (May and August 2005), seasonal effects on index scores were analyzed utilizing Wilcoxon Signed Rank test in SigmaPlot version 12.2 (Systat Software Inc., 2012) based on data from five independent segments from each stream. Index scores from long-term monitoring sites

were regressed against date to evaluate temporal trends within each site, and slopes were compared to zero (Mazor et al. 2009) to determine if significant trends could be identified. A Bonferroni correction was used ($\alpha = 0.01$) to account for multiple comparisons across indices. Spearman rank order correlations between calculated scores resulting from each index using all sites sampled in 2007 (n= 30) were conducted using SigmaPlot version 12.2, (Systat Software Inc., 2012).

Nonmetric multidimensional scaling (NMDS) was conducted in PC-ORD version 6.0 (McCune and Grace 2002; McCune and Mefford 2006) to evaluate community composition patterns for each stream and determine if compositional data clustered similarly to the calculated index score. Fish community data including species lists and numeric composition was used to create matrices for each site collection. Streams were used as potential clusters and examined for groupings. A total of 30 sites sampled in 2007 were analyzed from four streams (Oldhouse Creek n=3, Sickle Creek n=5, Pine Creek n=13, Bear Creek n=9). Analysis of variance (Kruskal-Wallis) with a post hoc multiple comparisons procedure (Dunn's test) was used to determine if the mean index score differed among streams (n ranged from three to 13)and showed similar patterns to the NMDS.

Results

Index applicability

We detected long-term interannual and short-term variation in daily mean water temperatures in July from long-term monitoring stations (2002-2012) in four tributary

streams of the Big Manistee River (Figure 3.2). Mean (°C ±1 SD) daily July temperature was 17.8 ±1.70, 14.03 ±1.30, 17.84 ±1.85, 18.73 ±1.90 for Oldhouse Creek, Sickle Creek, Pine Creek and Bear Creek, respectively. Only one site (Sickle Creek – Fig 2b) was consistently below all four maximum thresholds for designation as a coldwater stream as classified by Lyons et al. 1996; Wehrly et al. 2003; Lyons et al. 2009; Gerritsen and Stamp 2012; and Lyons 2012. Daily mean temperatures in the other three sites were periodically above the coldwater thresholds depending on the year. Because daily mean temperatures were variable and close to thresholds for coolwater and coldwater designations (depending on reference used) we analyzed both coolwater and coldwater indices developed for the region.

Index performance

Long term monitoring. - Species occurrence (2002-2010) at the long-term monitoring sites indicated a total of 12 families were represented (Table 3.2). This did include species that may have only been recorded in one sampling event. Mean index score indicated that Sickle Creek scored the highest with the coldwater indices and Bear and Pine Creeks scored poorly with coldwater indices and all sites scored well with the coolwater indices. Bear Creek had the highest mean species richness (15), followed by Pine Creek (10) with both Sickle and Oldhouse Creeks having an average annual richness of seven species. Richness was not indicative of metric scoring. Though Oldhouse Creek generally had a low richness value it also scored poorly in Lyons 96 as well as in the M&S. Presence of Brook Trout (*Salvelinus fontinalis*) or percent salmonids as Brook Trout is common to the coldwater IBIs as an indicator of high quality. None of the long

term monitoring sites had more than 1% salmonids as Brook Trout. Index score CV over the eight years ranged from 0 – 44 CV depending on index and site analyzed.

Temporal trends.- Seasonal (spring vs. fall) index scores were not significantly different (P>0.25, Wilcoxon Signed Rank test) based on data from five sites in each of three streams in 2005 (Figure 3.3). Mean scores within each stream all fell within the same index classification (e.g. fair) across season; however, individual scores varied by up to one classification level (e.g. fair v. poor).

Long-term (2002-2010) site scores varied in assessment over time and by index (Figure 3.4, 3.5). Lyons Cool produces the highest scores for all sites while Lyons 96 and M&S vary in ranking by site and over time (Figure 3.4). Overall long-term trend analysis of scores for Oldhouse Creek showed a significant decline through time based on M&S (r^2 = 0.867, P= 0.007). The BCG models indicated relatively stable results for Pine and Bear Creeks while more pronounced temporal variation was observed in Oldhouse Creek and Sickle Creek BCG scores (Figure 3.5). Long-term monitoring sites were classified differently depending on year and index used. For example in a single year, both Oldhouse Creek and Bear Creek sites could be classified as poor, fair, or excellent depending on the index used.

Stream site grouping.- NMDS exploration of the 30 study sites throughout the watershed surveyed in 2007 grouped sites based on their fish community composition (Figure 3.6). Sickle Creek sites grouped apart from Oldhouse Creek sites and both were distinct from Bear Creek and Pine Creek sites which overlapped. General patterns of relative similarity of the sites were supported by differences in mean index scores between the

streams (Kruskal-Wallis ANOVA) assessed with the three coldwater indices (Figure 3.7). Lyons 96 (H= 15.217 df = 3, P=0.002), M&S (H=15.265 df=3 P=0.002) and the BCG Cold model (H=11.299 df=3, P=0.010) differentially scored (Dunn post-hoc test P <0.05) Sickle and Oldhouse Creeks from each other as well as from Pine and Bear Creeks. Pine and Bear Creeks were not significantly different in mean score. There was no significant discrimination among streams based on the BCG Cool (H=5.285 df=3, P=0.152) or the Lyons Cool (H=5.903 df= 3 P=0.190).

Overall assessment of watershed stream sites in 2007 (n=30) indicated that the different indices produce disparate site classifications. The number of sites classified in top levels of condition varied by index such that 100% were classified as excellent using Lyons Cool index and 70% were classified to be within tiers 1,2 or 3 based on the BCG Cool index, whereas only 40% of the sites were classified to be within tiers 1,2 or 3 using the BCG Cold index. Based on the M&S index, 24% of sites were classified as good or better while assessments with Lyons 96 classified 30% of sites as good or better.

Correlation of indices.- Spearman's rank order correlations from data collected in 2007 at 30 sites throughout the watershed indicated that scores from Lyons 96 were positively correlated with those from M&S (Figure 3.8a; P= 0.000033, ρ= 0.715) and the BCG Cold (Figure 3.8b; P= 0.00595, ρ= 0.517). M&S scores were also positively correlated with those of BCG Cold (Figure 3.8c; P= 0.012, ρ= 0.477). Neither the Lyons Cool nor the BCG Cool scores were significantly correlated with other index scores or each other. Both of these indices tended to rate the sites similarly and did not often vary in site scores.

Discussion

The use of IBIs, though common in the Upper Midwest, is complicated and can lead to conflicting results depending on the index used for assessment. Score disparity can be confounded by thermal classification of a site. We found coolwater indices scored sites as consistently high quality when compared to scoring from coldwater indices, where sites were differentially scored. Inappropriate index application has been found to alter scores of biotic integrity (Wang et al. 2003; Baker et al. 2005) and may bias interpretation. Though all indices we applied were developed for the Upper Midwest region, they resulted in different site assessments for our study in the Big Manistee River watershed. Coolwater indices scored all sites to be of better condition when compared to coldwater indices, and although output from the coldwater indices were correlated, they varied in magnitude and classification of sites. Understanding primary index drivers and regional reference condition or the predicted natural condition that is the basis of each index will assist in selection of the most appropriate assessment tools.

Index applicability

Determining thermal classification of stream systems is imperative for accurate assessment and application of indices (Lyons 1992, Lyons et al. 1996). Our findings indicated that thermal regimes of streams we monitored were at the threshold between coolwater and coldwater classifications based on actual, not modeled temperatures. All streams assessed in this study are listed as Michigan trout streams (Section 48701(o), as amended being Sections 324.48701(o) of the Michigan compiled laws) and as such are

considered coldwater streams. Further, statewide modeled water temperature classified our study streams as coldwater (Wehrly et al. 2009). Although our actual temperature measurements in these sites occasionally exceeded coldwater thresholds, when coupling actual temperatures, modeled data and performance of the coolwater indices there is support for using coldwater classifications and associated indices for these streams. When the Lyons Cool and the BCG Cool were applied to datasets for the Big Manistee River watershed the results did not match those from the coldwater indices. Overall, Lyons Cool consistently scored sites as excellent with little variation and although the BCG Cool model displayed some sensitivity to temporal site trends, neither were correlated with other indices. However, all coldwater indices were correlated with each other. Additionally, the NMDS separated sites by stream and was supported by average coldwater index site scores analyzed with ANOVA, in contrast with coolwater indices (Figures 3.6 and 3.7).

Index performance

The three coldwater indices differ in parameters that have the greatest influence on the score for a given site. Indices generally produce higher scores (or a lower tier for the BCG) when there are few total species, and when native coldwater species are present and dominate the community. Though all three include Brook Trout as a focal species that leads to higher scores, Lyons 96 only has a total of five metrics resulting in relatively greater weight of each metric including the percent of salmonids that are Brook Trout. The low number of metrics has been identified as a limitation of the Lyons 96 index (Lyons et al., 1996; Angermeier and Karr 1986; Miller et al., 1988). M&S uses 12

metrics, thereby not placing as much weight on proportion of Brook Trout though it remains an important scoring metric. Similarly, in the BCG, Brook Trout are one of six coldwater species that improve site scores. This sensitivity in Lyons 96, based on a reduced number of metrics, was exemplified in Oldhouse Creek long-term monitoring site where Brook Trout were not encountered. The overall site scoring was very low with Lyons 96 and generally higher with M&S though variable (Table 3.2, Figure 3.4). Interestingly, the lack of Brook Trout and the presence of non-native salmonids, Brown Trout (*Salmo trutta)* and Rainbow Trout (*Oncorhynchus mykiss*), also resulted in lower scores from the BCG Cold index until additional coldwater taxa (e.g., Cottidae) started to be represented in the samples (2007-2009).

There was not a seasonal difference indicated by any of the indices for streams sampled. For example, similar index scores resulted from data collected in May and August/September. Therefore, while consistent sampling is important for monitoring programs, our results do not support that these indices are sensitive to seasonality or alternatively, there may be little variation due to seasonality. Based on this result, concern about combining data sets from sampling done at different times during the summer may be unwarranted; however, we would caution this result may only apply to this region where coldwater streams likely have more stable environmental conditions and biotic communities. The original IBI was developed to be robust to seasonality (Karr et al. 1986) and this has been supported by others, (Bozzetti and Schulz 2004) though Roset et al. (2007) recommended more rigorous analysis should be performed on seasonal comparisons of indices.

Analysis of site scores (2002-2010) indicated natural annual variability at long-term monitoring sites. In previous analyses of IBI variability over time, more highly degraded sites were found to exhibit more variability than less-impacted ones (Steedman 1988; DeShon 1994; Niemela and Feist 2000, Paller 2002). However, Hughes et al. (1998) cautioned how these results can be misleading if too few sites are included in the analysis. We observed mixed results from our long-term monitoring where three of the four sites were relatively degraded (fair to poor scores) and had higher CVs than the site that ranked in the "good" classification for biotic integrity when using Lyons 96 and M&S. As noted in Mundahl and Simon (1999: M&S), the assessment of site condition by this index (M&S) and Lyons 96 diverge as impairment of the sites increased. They described how M&S includes seven metrics sensitive to low levels of biotic integrity and may be better able to differentiate sites with higher levels of degradation. Though both indices scored the long-term monitoring site on Pine Creek as poor and Sickle Creek as good, they diverged in assessment of Oldhouse Creek and Bear Creek long-term monitoring sites. The Oldhouse Creek monitoring indicated a significant trend of degradation over time using M&S. The BCG Cold indicated that three of the sites were relatively stable with only Sickle Creek ranging over three tiers of the index. These examples suggest that with continued long-term monitoring, natural variability can be detected and therefore trends of degradation or improvement can be tracked with indices. Though Trebitz et al. (2003) found IBIs insensitive to subtle changes in a simulated system and warned against using only IBIs as an early detection system for degradation, others have stressed the need for long term datasets, including IBIs, as a tool for

increasing our understanding of variability and change in a system (Fore et al. 1994; Jackson and Fureder 2006; Kennen et al. 2012).

Separation of stream sites between coldwater index scores (Lyons 96, M&S, and BCG Cold) and community composition (NMDS) was similar. The coldwater indices discriminate among differing fish community compositions and can be used to rank and compare the stream systems. Sickle Creek had the highest average scores for all indices in 2007 due to all species being coldwater, including Burbot (*Lota lota*) and Brook Trout. Oldhouse Creek sites also had low diversity (6 species in 2007) however; they generally scored poorly due to presence of tolerant species and absence of Brook Trout. Pine Creek and Bear Creek sites did not separate well, due to overlap in species, and general compositional similarities with richness of 10 and 16 species, respectively. The coldwater IBIs allow for differential assessment of stream fish assemblages and are sensitive to the types of species alterations that we encountered in the Northern Lakes and Forests Ecoregion.

Interpretation of Index scores

Interpretation of IBI scores should incorporate an understanding of the effect of certain drivers of the index. Coldwater indices tended to similarly assess trends and agree in their relative scoring of sites from high to low. However, if a specific site/year is analyzed with all three indices it could be ranked fair, poor, or good depending on the index chosen (e.g., Oldhouse site; Figure 3.4, 3.5). Though the three coldwater indices are concordant and generally show similar temporal trends, magnitude of score was

variable and could lead to different conclusions regarding the general integrity of a given site.

Spatial interpretations (landscape and local views) of fish communities can also lead to seemingly contradictory results if the relative temporal and spatial coverage are not adequately represented within the data (Wiley et al. 1997). Wang et al. (2003) found that IBI scores in the Northern Lakes and Forests Ecoregion were not strongly related to land-use. They attributed this to mis-application of IBI and lack of a coolwater IBI, since many of their sites lacked noted environmental degradation. With empirical validation of temperature regimes at the sites we monitored in the same Ecoregion, long-term monitoring stations, and knowledge of historic and current land-use, we believe that the sites monitored in the Big Manistee River watershed are coldwater streams with signs of degradation. VanDusen et al. (2005) found that forestry management practices (i.e., selection logging) in the northern portion of the same Ecoregion left historic landuse signals that were present in the biotic communities of streams. In the same study Flashpoler et al. (2002), found indications of degradation in multiple biotic communities (birds, fish and macroinvertebrates) could persist for up to thirty years even where local habitat (30m riparian buffer) was maintained according to best management practices during selection logging. With much of the Big Manistee River watershed dominated by forest, this is a possible source of degradation in a seemingly unimpacted Ecoregion. As noted by Wiley et al. (1997) both spatially and temporally extensive designs are needed to accurately assess biotic communities.

Assessment of fish communities has effectively conveyed information on the status and ecological quality of aquatic ecosystems (Roset et al. 2007). The scientific

community should ensure interpretation of index scores is clear and well understood. This is exemplified by data from Michigan where Pine Creek and Bear Creek are considered "Tier 1 trout streams" (Tonello 2011) and yet score in the "poor" category for biotic integrity. The occurrence of non-native fish lowers the score resulting from the BCG, Lyons 96, and M&S and therefore each index includes explicit statements regarding how they were developed to assess the overall integrity of a fish community. In the Upper Midwest coldwater systems are commonly managed for non-native fish species, often through stocking, which can lower the overall index score and quality (Mundahl and Simon 1999). Biotic integrity is different than the status of a fishery so communicating a clear interpretation of score is imperative to reduce confusion and insure appropriate understanding of the assessment.

The indices are all developed in relation to regional reference condition or a predicted natural condition that provides a reference point for score interpretation and exemplifies the need for greater understanding of those conditions. For example, Lyons 96 is based on using Wisconsin coldwater streams as the reference condition, whereas M&S is based on streams from WI, MN, and MI. The BCG Cold model was also calibrated for streams in WI, MN, and MI; however, it was developed in relation to a presumed natural and undisturbed condition (Tier 1). The M&S scores represent how assessments score in relation to streams throughout WI, MN, and MI whereas, the Lyons 96 score represents how the site assessment ranks in relation to WI coldwater streams. The BCG Cold score represents how a sites ranks in relation to a natural, undisturbed condition for the entire Upper Midwest. BCG Cold generally scored sites as mediocre for the region where there was a larger gradient of condition expressed in scores from

M&S and Lyons. This difference in scoring may be partially explained by the reference condition that is used as the basis of each index.

Conclusion

Assessing and communicating the status of lotic systems is key to their management, restoration, and protection. While indices such as the IBI have been applied for over thirty years (Karr 1981) they have been refined and improved through time. Assessment of indices used in the Upper Midwest indicated that proper classification of thermal regime is imperative for appropriate and accurate scoring. For the coldwater indices evaluated, site scores revealed long-term trends and natural variability. Though indices were correlated, magnitude of site score was variable which can be explained by the reference condition used and the metrics that were drivers of each index. By defining the biotic integrity of highest value (reference condition), interpretation of each index becomes clearer. For example, if a stream community dominated by Brook Trout is the valued condition, then using an index that is heavily influenced by that species would be appropriate. If comparison to a pristine natural condition is desired then using an index based on that reference would be more appropriate. The indices we used classified sites into groups that are all based on the value of particular biotic communities. Recognizing and incorporating the bias of the index used into communication of the scores will assist in interpretation. The BCG Cold ranks sites compared to a natural undisturbed condition and indicated that most of our sites showed changes in community structure with tolerant taxa occurring. Lyons 96 scores were heavily weighted and influenced by Brook Trout, and though this may be

used as an indicator of biotic integrity, for streams that are slightly degraded and with minimal Brook Trout, there may be limitations. For areas like the Big Manistee River watershed where there is some degradation, the metric-rich M&S tracked trends over time, elucidated annual variation, and discriminated among sites. All three of the coldwater indices could be used for assessment as long at the drivers of the index and the reference conditions were articulated, understood, and included critical features of the goals of the assessment.

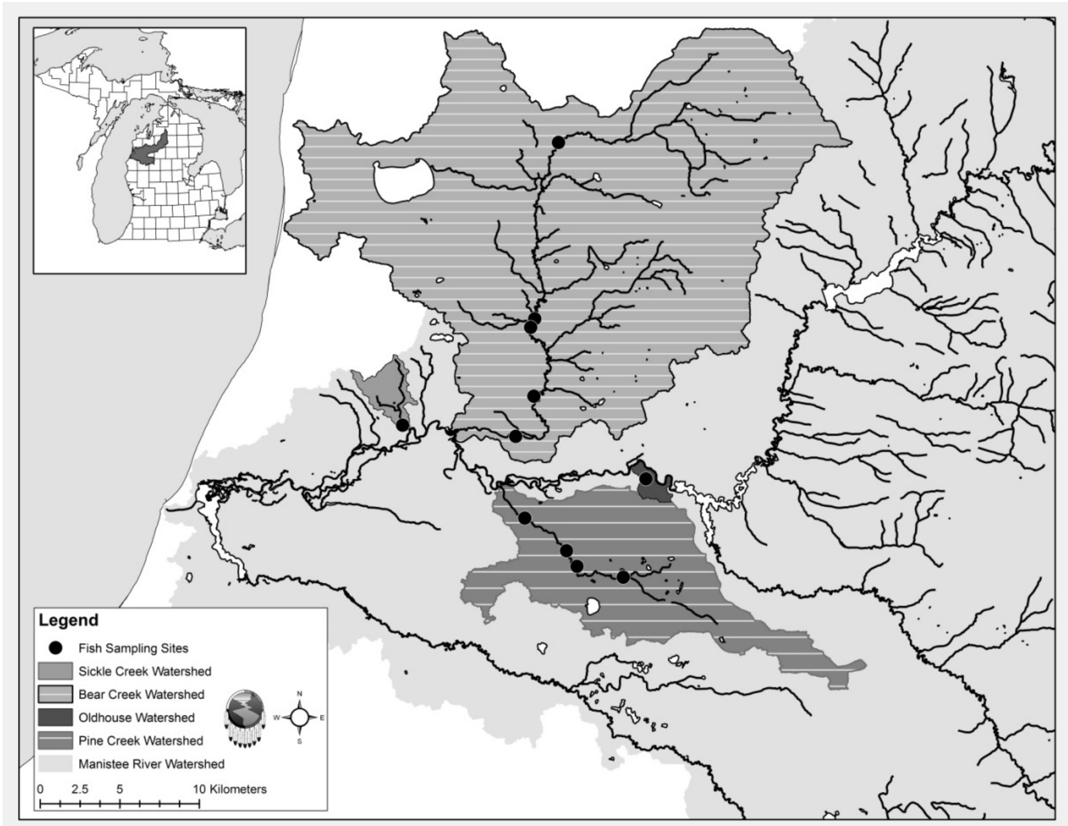

Figure 3.1. Location of fish sampling sites within the Big Manistee River watershed, Michigan, USA. Areas shaded include the sub-watersheds of Bear Creek, Pine Creek, Sickle Creek and Oldhouse Creek. All dots represent sampling reaches within the streams.

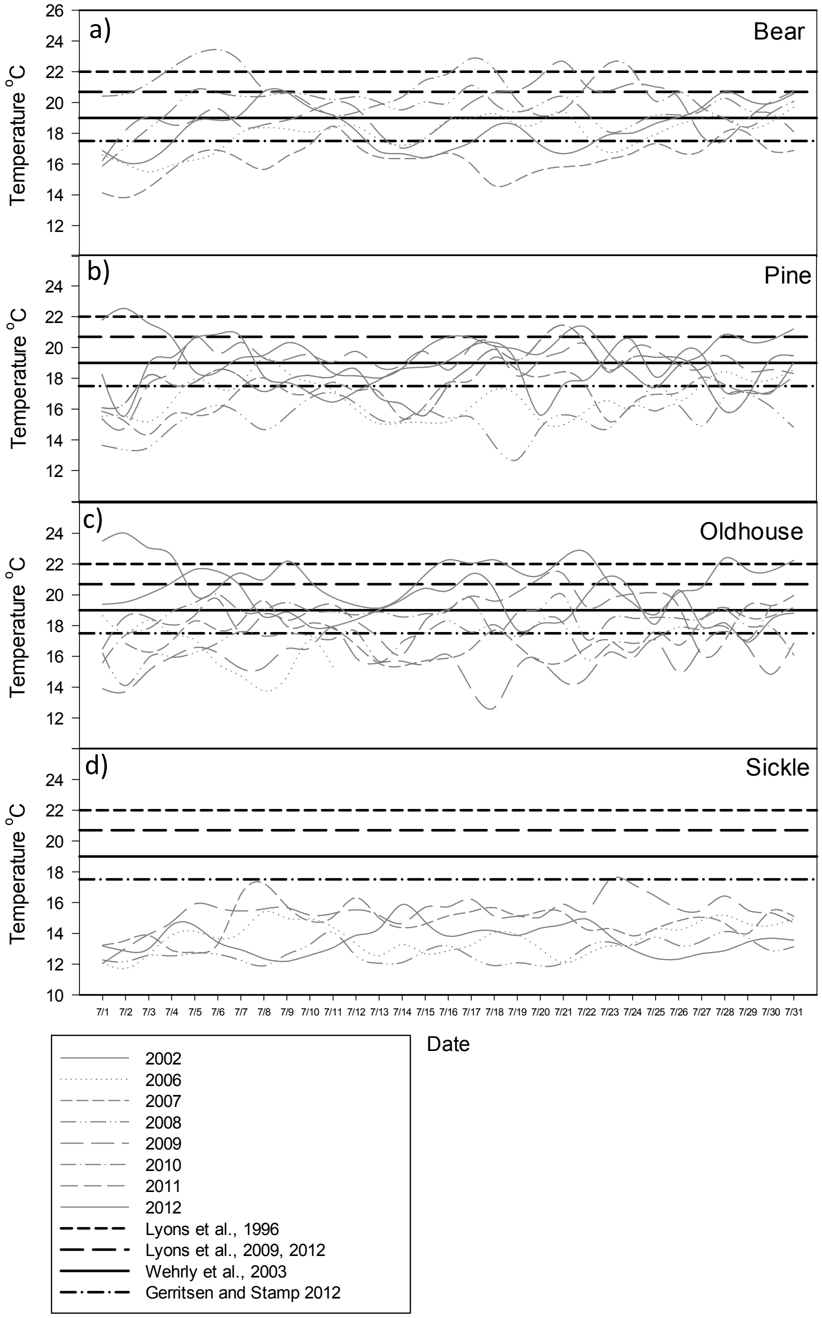

Figure 3.2. Daily mean temperatures (C°) in July at four long-term monitoring stations in the Big Manistee River watershed. Data were collected periodically from 2002 – 2012 depending on site location (ie: Sickle Creek only has 5 years of data). Four different maximum thresholds for cold water classification systems are depicted 1) Lyons et al 2009,2012; max daily mean <20.7, 2) Lyons et al 1996; max daily mean <22, 3) Wehrley et al 2003; July weekly mean <19, 4) Gerritesen and Stamp 2012; average July <17.5.

Table 3.1. Numeric thresholds and classification systems for indices assessed.

Lyons 96[a]		M&S [b]		Lyons Cool[c]		BCG[d]
100-90	Excellent	105-120	Excellent	70-100	Excellent	1 2
80-60	Good	70-100	Good	50-60	Good	
50-30	Fair	35-65	Fair	30-40	Fair	3 4
20-10	Poor	10-30	Poor	0-20	Poor	5 6
0	Very Poor	0-5	Very Poor			

*References for scoring: a) Lyons et al., 1996, b) Mundahl and Simon 1999, c) Lyons 2012, d) Gerritsen and Stamp 2012

Table 3.2. Total species richness by family for each of the long-term monitoring sites from 2002-2010 as represented by more than two years of occurrence with total in parentheticals. Mean index score (± CV) is given for each site.

	Site			
	Species richness per family			
Family	Bear	Pine	Oldhouse	Sickle
Cyprinidae	5 (9)	3(4)	3	2(3)
Centrarchidae	2 (5)	1(5)	0(2)	1(2)
Salmonidae	4	5	1(3)	5
Umbridae	1	1	0(1)	1
Percidae	2 (3)	1	1	1(2)
Cottidae	1	1	1	1
Petromyzontidae	0	0 (1)	0(1)	1
Catostomidae	1(2)	2	1	0
Gasterosteidae	0	0	1	1
Lotidae	1	0	0	1
Esocidae	2(3)	0	0	0
Ictaluridae	0	0	0(1)	0
Mean Annual Richness	15	10	7	7
Index	Mean Score (± CV)			
Lyons 96	48 (22)	27 (26)	23 (44)	72 (15)
M&S	32 (16)	28 (21)	44 (33)	74 (8)
Lyons - Cool	98 (5)	98 (4)	78 (10)	83 (12)
BCG Cool	3 (15)	4 (0)	3 (15)	3 (17)
BCG Cold	4 (0)	4 (0)	3 (16)	3 (25)

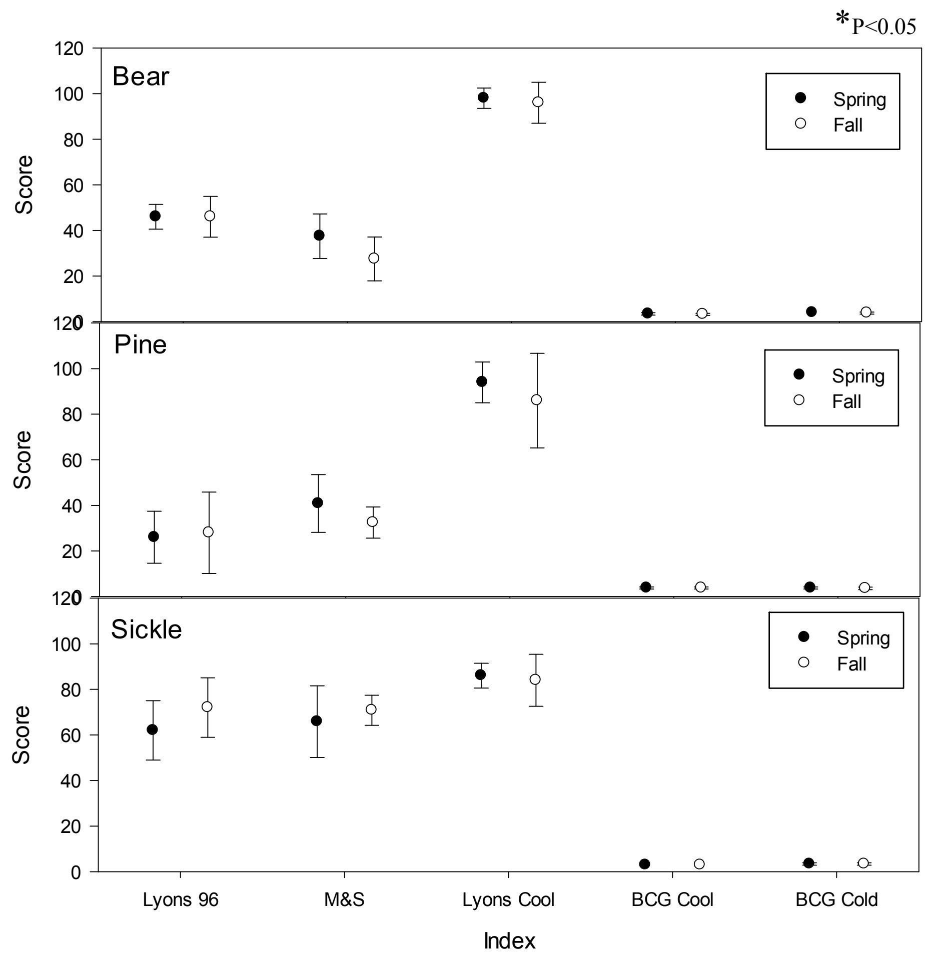

Figure 3.3. Mean index scores (±SD) did not indicate differences between seasons for three stream segments assessed in 2005 (Wicoxon Signed Rank test, p >0.05). Five reaches were sampled in May and in August for each stream.

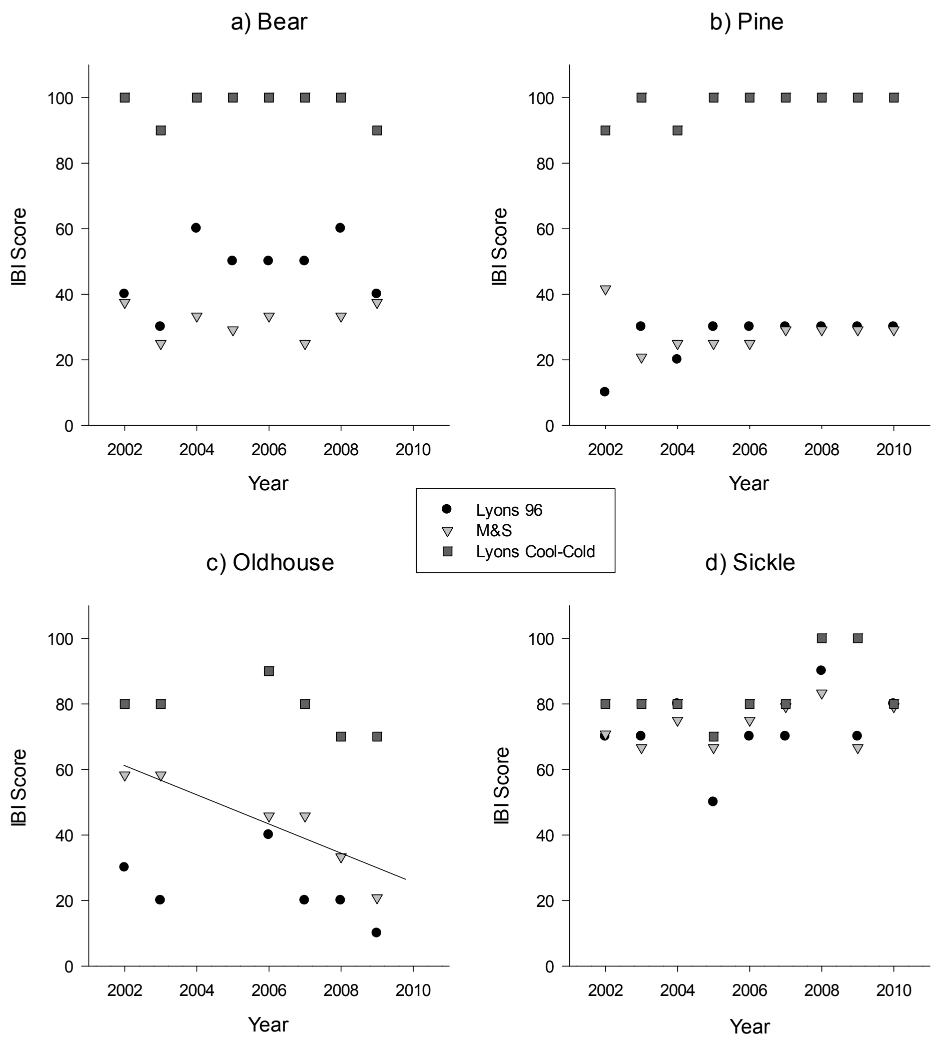

Figure 3.4. Trend analysis from 2002 to 2010 at four long-term monitoring stations in the Big Manistee River Watershed. Only one significant regression was detected (C: Oldhouse $r^2 = 0.867$, $P = 0.007$).

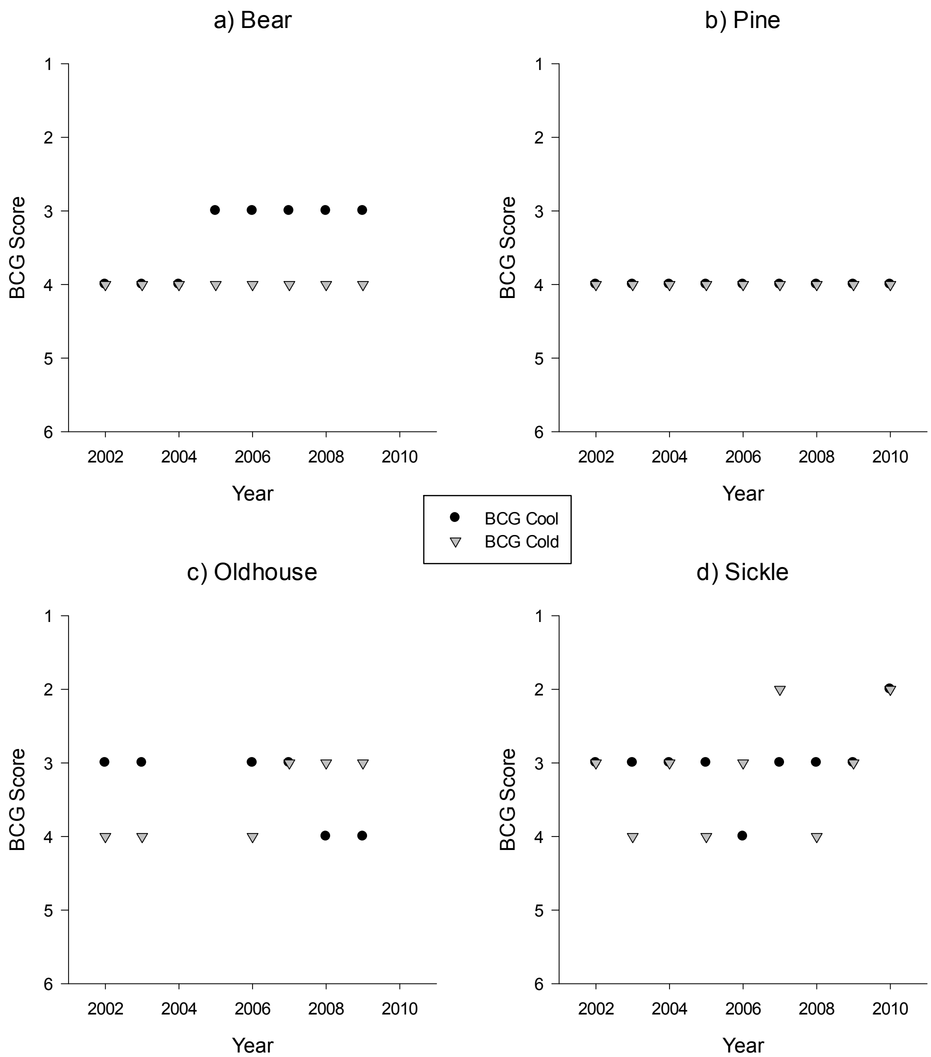

Figure 3.5. Trend analysis from 2002 to 2010 at four long-term monitoring stations in the Big Manistee River Watershed using the BCG cool and cold models indicated no significant relationships.

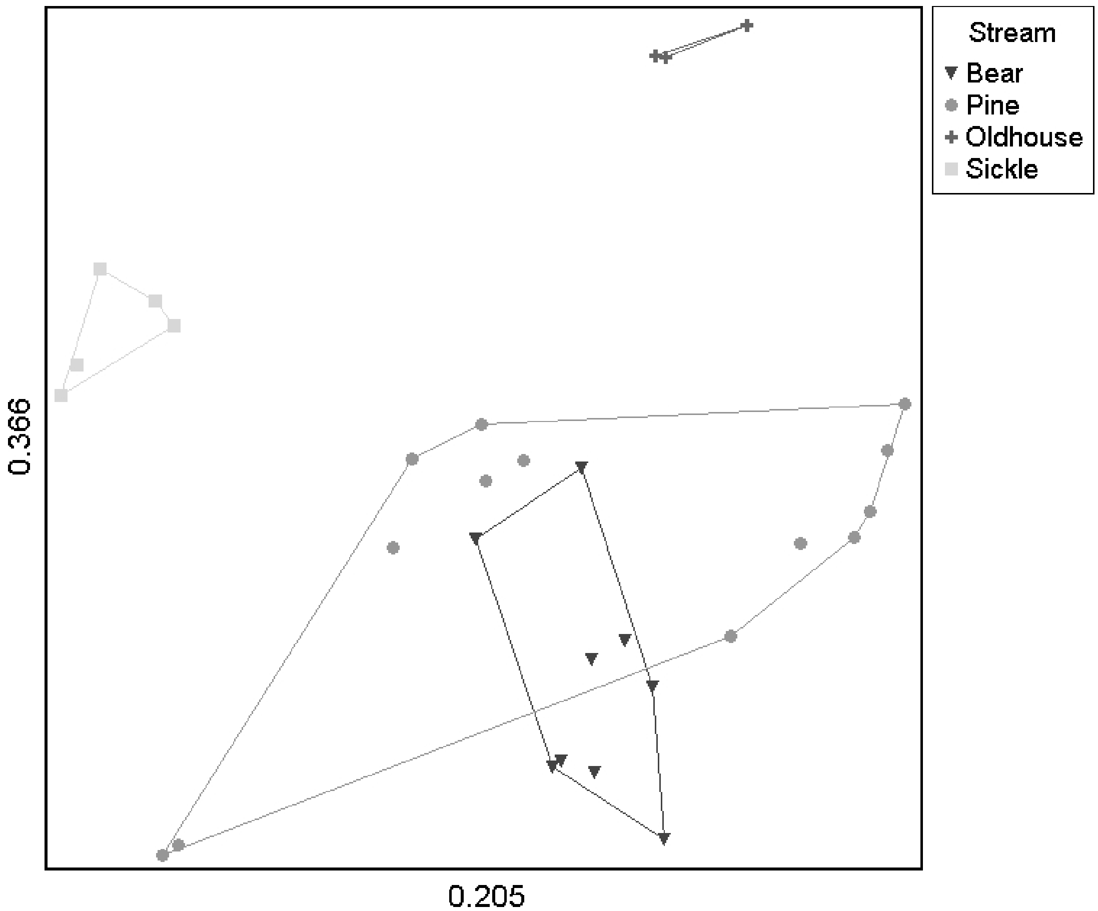

Figure 3.6. Nonmetric multidimensional scaling output of community composition data from 30 sites (2007). Groupings are based on convex hulls produced in PC-ORD.

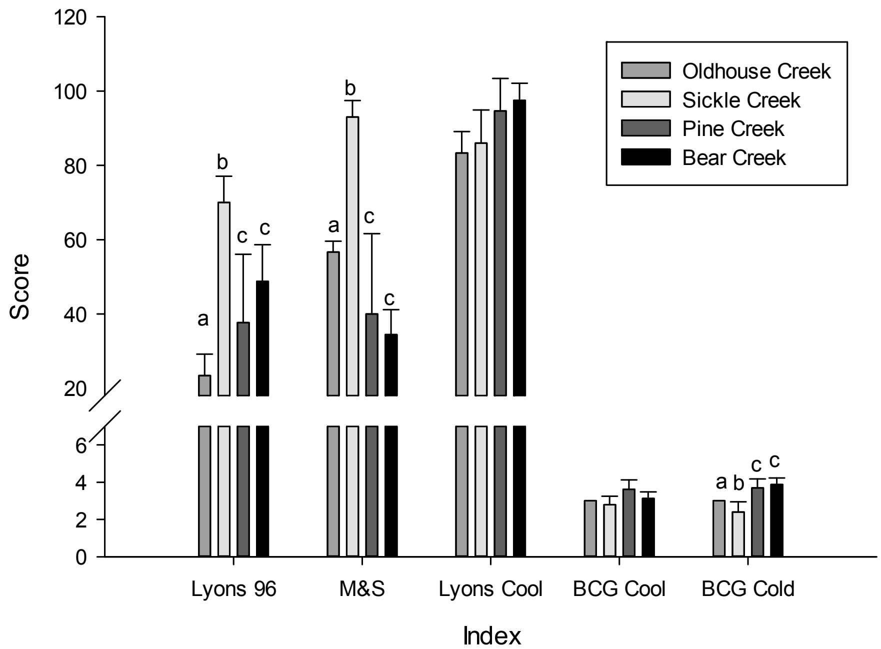

Figure 3.7. Index score for each stream in 2007, where a separate ANOVA was completed for each index. Lyons 96, M/S and BCG Cold all had similar patterns where indices showed significant differences (marked with unique letters) in mean site scores where Sickle Creek was different from Oldhouse Creek and from Pine Creek and Bear Creek. Oldhouse Creek is different than Pine Creek and Bear Creek although Pine and Bear are not significantly different. Lyons Cool and the BCG Cool index scores did not indicate these same patterns and no significant difference was found. Note that the relative condition of sites based on the tiers of the BCG are reversed from the other indices such that a lower tier is better condition.

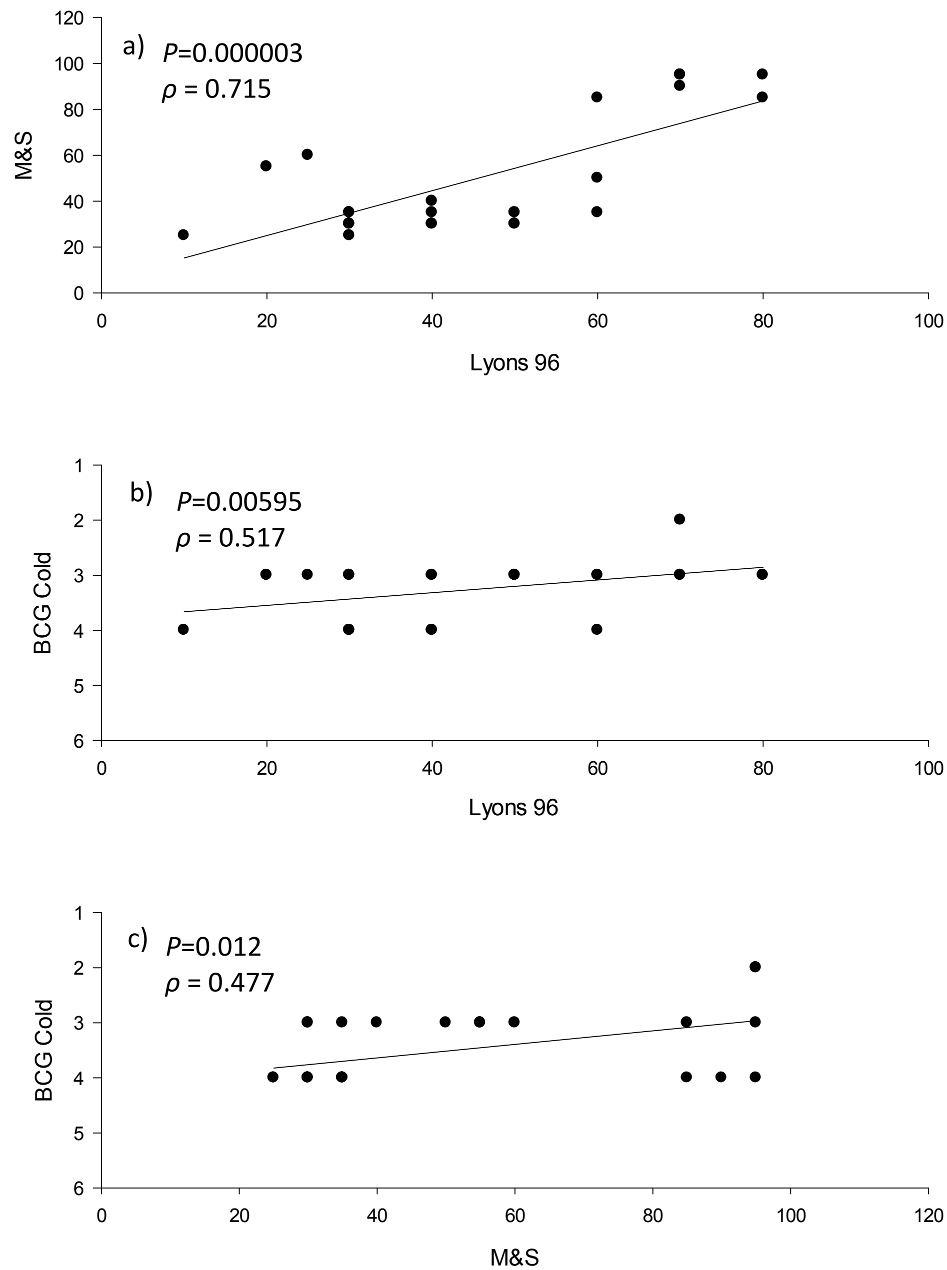

Figure 3.8. Spearman's rank order correlation among coldwater index scores for 30 sites throughout the Big Manistee River watershed in 2007.

References

Allen, D. J. 2004. Landscape and riverscapes: The influence of land use on river ecosystems. Annual Reviews of Ecology, Evolution and Systematics. 35:257-284.

Allan, D. J., D. L. Erickson and J. Fay. 1997. The influence of catchment land use on stream integrity across multiple spatial scales. Freshwater Biology. 37:149–161.

Angermeier, P. L., and J. R. Karr. 1986. Applying and index of biotic integrity based on stream-fish communities: considerations in sampling and interpretation. North American Journal of Fisheries Management. 6:418-429.

Baker, E. A., K. E. Wehrly, P. W. Seelbach, L. Wang, M. J. Wiley, and T. Simon. 2005. A multimetric assessment of stream condition in the Northern Lakes and Forests Ecoregion using spatially explicit statistical modeling and regional normalization. Transactions of the American Fisheries Society. 134:697-710.

Barbour, M. T., J. Gerritsen, B. D. Snyder, and J. B. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish. 2nd ed., U.S. Environmental Protection Agency, Office of Water, EPA Report 841-B-99-002, Washington, DC.

Bozzetti, M. and U.H. Schulz. 2004. An index of biotic integrity based on fish assemblages for subtropical streams in southern Brazil. Hydrobiologia 529:133-144.

Connor, W. P., H. L. Burge, J. R. Yearsley, and T. C. Bjornn. 2003. Influence of flow and temperature on survival of wild subyearling fall chinook salmon in the Snake River. North American Journal of Fisheries Management 23:362-375.

Davies, S. P., and S. K. Jackson. 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. Ecological Applications. 16:1251–1266.

DeShon, J. E. 1994. Development and application of the invertebrate community index (ICI). Pages 217-243 *in* W. S. Davis and T. P. Simon, editors. Biological assessment and criteria: tools for water resource planning and decision making. Lewis Publishers, Boca Raton, Florida.

Fausch, K. D., C. E. Torgersen, C. V. Baxter, and H. W. Li. 2002. Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes. BioScience 52:483–498.

Flaspohler D. J., C.J.F Huckins, B.R. Bub, and P.J. VanDusen. 2002. Temporal patterns in aquatic and avian communities following selective logging in the upper Great Lakes region. Forest Science 48: 339-349.

Fore, L.S., Karr, J.R., and L.L. Conquest. 1994. Statistical properties of an index of biological integrity used to evaluate water resources. Canadian Journal of Fisheries and Aquatic Sciences 51:1077–1087.

Frey, D. 1977. Biological integrity of water: an historical approach. Pages 127-140 *in* R.K Ballentine, and L.J Guarraia,. editors. The integrity of water. Proceedings of a symposium. March 10–12, 1975. U.S. Environmental Protection Agency, Washington, DC.

Frissell, C.A., W. J. Liss, C. E. Warren, and M. D. Hurley. 1986. A hierarchical framework for stream habitat classification: Viewing streams in a watershed context. Environmental Management 10:199-214.

Gerritsen, J., and J. Stamp. 2012. Calibration of the biological condition gradient (BCG) in cold and cool waters of the Upper Midwest: BCG-based indexes (BCG-I) for fish and benthic macroinvertebrate assemblages. Tetra Tech, Inc. Owings Mills, MD.

Gregory, S. V., F. J. Swanson, W. A. McKee, and K.W. Cummins. 1991. An ecosystem perspective of riparian zones. BioScience 41:540-551.

Hendrickson, G. E., and C. J. Doonan. 1972. Hydrology and recreation on the cold-water rivers of Michigan's southern peninsula: Geological Survey Water Information Series Report 3.

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.

Hynes, H. B. N., 1975. The stream and its valley. Verhandlungen des Internationalen Verein Limnologie 19:1-15.

Jackson, D. A., P. R. Peres-Neto, and J. D. Olden. 2001. What controls who is where in freshwater fish communities - the roles of biotic, abiotic, and spatial factors. Canadian Journal of fisheries and Aquatic Sciences 58: 157–170.

Jackson, J. K., and L. Fureder. 2006. Long-term studies of freshwater macroinvertebrates: a review of the frequency, duration and ecological significance. Freshwater Biology 51:591-603.

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

Karr, J. R. 1999. Defining and measuring river health. Freshwater Biology. 41:221-234.

Karr, J. R., and D. R. Dudley. 1981. Ecological perspective on water quality goals. Environmental Management 5:55–68.

Karr, J. R., and E. W. Chu. 1999. Restoring life in running waters: better biological monitoring. Island Press. Washington, DC.

Karr, J. R., K. D. Fausch, P. L. Angermeier, P. R. Yant, and I. J. Schlosser. 1986. Assessing biological integrity of running waters. A method and its rationale. Illinois Natural History Survey Special publication #5. Champaign, Il.

Kennen, J. G., D.J. Sullivan, J.T. May, A.H. Bell, K. M. Beaulieu, and D.E. Rice. 2012. Temporal changes in aquatic-invertebrate and fish assemblages in streams of the north-central and northeastern US. Ecological Indicators 18:312-329.

Lyons, J. 1992. Using the index of biotic integrity (IBI) to measure environmental quality on warmwater strean if Wisconsin. U.S. Forest Service General Technical Report NC-149.

Lyons, J. 1996 . Pattern in the species composition of fish assemblages among Wisconsin streams. Environmental Biology of Fishes 45:329-341.

Lyons, J. 2012. Development and validation of two fish-based indices of biotic integrity for assessing perennial coolwater streams in Wisconsin, USA. Ecological Indicators 23:402-412.

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

Lyons, J., T. G. Zorn, J. Stewart, P. W. Seelbach, K.E. Wehrly, and L. Wang. 2009. Defining and characterizing coolwater streams and their fish assemblages in Michigan and Wisconsin, USA. North American Journal of Fisheries Management 29:1130-1151.

Magnuson, J. J., L. B. Crowder, and P. A. Medvick. 1979. Temperature as an ecological resource. American Zoologist 19:331-343.

Mazor, R. D., A. H. Purcell., and V. H Resh. 2009. Long-term variability in bioassessments: a twenty-year study from two northern California streams. Environmental Management. 43:1269–1286.

McCune, B., and J. B Grace. 2002. Analysis of ecological communities. MjM Software Design, Gleneden Beach, Oregon USA.

McCune, B., and M. J Mefford. 2006. PC-ORD: multivariate analysis of ecological data. MJM Software. Version 5.14. Gleneden Beach, Oregon USA.

Miller, D. L., R. M. Hughes, J. R. Karr, P. M. Leonard, P. B. Moyle, L. H. Schrader, B. A. Thompson, R. A. Daniels, K. D. Fausch, G. A. Fitzhugh, J. R. Gammon, D. B. Halliwell, P. L. Angermeier, and D. J. Orth. 1988. Regional applications of an index of biotic integrity for use in water resource management. Fisheries 13:2-20.

Moyle, P. B., and J. J. Cech. 1996. Fishes: An introduction to ichthyology. Prentice Hall, New Jersey.

Mundahl, N. D. and T. P. Simon, 1999. Development and application of an index of biotic integrity for coldwater streams of the upper midwestern United States.Paged 383-416 *in* T. P. Simon, editor. Assessing the Sustainability and Biological Integrity of Water Resources using Fish Communities. CRC Press, Boca Raton, Florida.

Niemela, S. and M. Feist, 2000. Index of biotic Iintegrity (IBI) guidance for coolwater rivers and streams of the St. Croix River basin in Minnesota. Minnesota Pollution Control Agency, Biological Monitoring Program, St. Paul, MN.

NLCD, 2006. National Oceanic and Atmospheric Administration, Coastal Services Center. 1995-present. The Coastal Change Analysis Program (C-CAP) Regional Land Cover. Charleston, SC: NOAA Coastal Services Center. Accessed at www.csc.noaa.gov/digitalcoast/data/ccapregional.

Omernik, J.M., 1987. Ecoregions of the conterminous United States. map (scale 1:7,500,000). Annals of the Association of the American Geographers. 77:118-125.

Paller, M. H. 2002. Temporal variability in fish assemblages from disturbed and undisturbed streams. Journal of Aquatic Ecosystems Stress and Recovery 9:149–158.

Poff, N. L., and J. D. Allan. 1995. Functional organization of stream fish assemblages in relation to hydrologic variability. Ecology 76:606-627.

Roset, N., G. Grenouillet, D. Goffaux, D. Pont, and P. Kestemont. 2007 A review of existing fish assemblage indicators and methodologies. Fisheries Management and Ecology 14:393-405.

Roth, N. E., J. D. Allan and D. L. Erickson, 1996. Landscape influences on stream biotic

integrity assessed at multiple spatial scales. Landscape Ecology. 11:141–156.

Ruaro, R. and E.A. Gubiani. 2013. A scientometric assessment of 30 years of the Index of Biotic Integrity in aquatic ecosystems: Applications and main flaws. Ecological Indicators 29:105-110.

Seelbach, P. W., M. J. Wiley, J. C. Kotanchik, and M. E. Baker. 1997. A landscape-based system of ecological classifications for river valley segments in lower Michigan: VSEC Version 1.0. Michigan Department of Natural Resources, Fisheries Research Report 2036, Ann Arbor.

Sloat, M. R., B. B. Shepard, R. G. White, and S. Carson. 2005. Influence of stream temperature on the spatial distribution of westslope cutthroat trout growth potential within the Madison River Basin, Montana. North American Journal of Fisheries Management 25:225-237.

Steedman, R. 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.

Suter, G. W. 1993. A Critique of ecosystem health concepts and indexes. Environmental toxicology and chemistry 12:1533-1539.

Tonello, M. A. 2011. Michigan Department of Natural Resources Status of the Fishery Resource, Pine Creek MDNR 2011-22.

Tonn, W. M., J. J. Magnuson, M. Rask, and J. Toivonen. 1990. Intercontinental comparison of small-lake fish assemblages: The balance between local and regional processes. American Naturalist 136:345-375.

Trebitz, A. S., B. H. Hill, and F. H. McCormick. 2003. Sensitivity of indices of biotic integrity to simulated fish assemblage changes. Environmental Management 32:499-515.

Unites States Environmental Protection Agency, 2002. Summary of biological assessment programs and biocriteria development for states, tribes, territories, and interstate commissions: Streams and wadeable rivers. Technical Report EPA 822-R-02-048, U.S.Environmental Protection Agency, Office of Environmental Information and Office of Water, Washington, D.C.

VanDusen, P. H., C. J Huckins, and D.J. Flashpoler. 2005. Associations among selection logging history, brook trout, macroinvertebrates, and habitat in northern Michigan headwater streams. Transactions of the American Fisheries Society 134:762-774.

Vannote, R. L., G. W. Minshall, K. W. Cummings, J. R. Sedell, and C. E. Cushing. 1980.

The river continuum concept. Canadian Journal of Fisheries and Aquatic Sciences 37:130-137.

Wang, L., J. Lyons, P. Kanehl and R. Gatti, 1997. Influences of watershed land use on habitat quality and biotic integrity in Wisconsin streams. Fisheries 22:6–12.

Wang, L., J. Lyons, P. Rasmussen, P. Kanehl, P. Seelbach, T. Simon, M. Wiley, E. Baker, S. Niemela, and P. M. Stewart. 2003. Watershed, reach, and riparian influences on stream fish assemblages in the Northern Lakes and Forest Ecoregion, U.S.A. Canadian Journal of Fisheries and Aquatic Sciences 60:491-505.

Wehrly, K. E., L. Wang, and M. Mitro. 2007. Field-based estimates of thermal tolerance limits for trout: incorporating exposure time and temperature fluctuation. Transactions of the American Fisheries Society 136:365–374.

Wehrly, K. E., M. J. Wiley, and P. W. Seelbach. 2003. Classifying regional variation in thermal regime based on stream fish community patterns. Transactions of the American Fisheries Society 132:18-38.

Wehrly, K. E., M. J. Wiley, and P. W. Seelbach. 2006. Influence of landscape features on summer water temperatures in lower Michigan streams. Pages 113–128 *in* R. M. Hughes, L. Wang, and P.W. Seelbach, editors. Influences of landscapes on stream habitats and biological assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

Wehrly, K. E., T. O. Brenden, L. Wang. 2009. A comparison of statistical approaches for predicting stream temperatures across heterogeneous landscapes. Journal of the American Water Resources Association 45:986-997.

Wiley, M.J., S.L. Kohler, and P.W. Seelbach. 1997. Reconciling landscape and local views of aquatic communities: lessons from Michigan trout streams. Freshwater Biology 37:133-148.

Zorn, T. G., P. W. Seelbach, E. S. Rutherford, T. C. Wills, S. T. Cheng, and M. J. Wiley. 2008. A regional-scale habitat suitability model to assess the effects of flow reduction on fish assemblages in Michigan streams. Michigan Department of Natural Resources, Fisheries Research Report 2089, Ann Arbor.

Zorn, T. G., P. W. Seelbach, and M. J. Wiley. 2002. Distributions of stream fishes and their relationship to stream size and hydrology in Michigan's Lower Peninsula. Transactions of the American Fisheries Society 131:70-85.

Chapter 4. [3] Culvert replacements; improvement of stream biotic integrity?

Introduction

Human activities continue to disturb the natural structure and function of rivers despite (and also because of) the value of rivers for drinking water, agriculture, recreation, and subsistence (MA 2005; Dudgeon et al. 2006). Indeed, one of the greatest environmental challenges for the twenty-first century will be to sustain biologically diverse and functional river ecosystems (Bernhardt et al. 2006). Recently, substantial attempts to improve the quality and integrity of freshwater ecosystems at both national and international levels has occurred (NRC 1992; Frissell & Bayles 1996; Baron et al. 2002). River rehabilitation has become a common management activity to improve resources of economic, cultural and/or spiritual importance (Roni et al. 2008) and accordingly it is increasingly important we are clear about the motivation and justification of these projects. This in turn allows for proper monitoring of restoration projects to determine if they have achieved stated goals and that they are done at the appropriate scale to achieve the goal. This is especially true if the goal is biological.

Society has been made well-aware that all too often, restoration efforts have proceeded without substantive documentation of successes and failures of individual activities (sensu Palmer et al. 2005). Even when success is noted, data may be inadequate to identify specific results or management endpoints. Of 37,099 restoration projects in the National River Restoration Science Synthesis (NRRSS) database only 10% indicated any form of assessment or monitoring (Bernhardt et al. 2005) with 47% of project assessments based on qualitative site observations (Bernhardt et al. 2007). As

[3] The material contained in this chapter has been submitted to the Journal *Restoration Ecology*

part of the NRRSS Alexander and Allan (2007) examined results from Midwest (Michigan, Wisconsin and Ohio) restoration projects and found monitoring occurred in 79% of the projects; however, they rarely documented biological improvements.

Human land use, development and urbanization can alter water flow and degrade stream habitat and biotic conditions through, for example, draining agricultural fields and channelizing streams (Petersen 1992; Ogren &King 2008), increasing sedimentation from agricultural (Berkman & Rabeni 1987) and forestry management activities (Meehan 1991; VanDusen et al. 2005) and loss of important riparian buffer habitat (Naiman et al. 2005). A major component of human land use in aquatic systems is the construction, maintenance, and use of roads that occur as part of human infrastructure and the road/stream interface is one of the main pathways for sediment to reach waterways (Croke et al. 2005). Stream crossings, often culverts, can alter in-stream sediment accumulations and geomorphology of a stream (Wellman et al. 2000). The effects of sedimentation on macroinvertebrates and fish have been well documented (Wood & Armitage 1997; Jones et al. 2012); however, in a review of stream rehabilitation by Roni et al. (2008), only five of 345 examined papers documented positive impacts of road improvements, including culvert replacement, and only two discussed the effects on water quality and biota. A primary disconnect between restoration activities and their goals is highlighted by the pattern that improvement in biotic community is generally the main objective of stream restoration though it is not often monitored.

Improperly designed, deployed, or maintained culverts can pose negative impacts on the biotic community. For instance, a decrease in abundance of macroinvertebrates

directly below culverts was found by Khan and Colbo (2008). Blakely et al. (2006) found adult *Tricoptera* (caddisfly) upstream movement was reduced by road culverts. Studies on the implications of culverts for passage of fish, primarily *Salmonids*, have indicated reduced adult upstream movement and fragmentation and alteration of juvenile fish habitat (Burford et al. 2009; Davis & Davis 2011), as well as, reduced movement of non-sport fish species through culverts (Warren & Pardew 1998; MacPherson et al. 2012; Briggs & Galarowicz 2013).

Culvert replacement and road crossing improvement projects are relatively common, with multiple procedures for evaluating and prioritizing removal (Kemp & O'Hanely 2010). Januchowski-Hartley et al. (2013) estimated that in the North American Great Lakes Basin 268,818 road stream crossings could cause degradation to streams and only 36% of road crossing structures identified in the study were determined to be completely passable for fish. Within this same region, but focused at a smaller scale, the Big Manistee River watershed in the northwestern Lower Peninsula of Michigan was estimated to have 243 inventoried road-stream crossings and 80% were considered moderately or severely degraded (www.northernmichiganstreams.org). In this watershed we have been involved in an ongoing concerted effort to repair and improve road crossings, motivated largely to promote improved connectivity and biotic integrity.

In this study, we evaluated the effects of replacing three improperly designed and failing road stream crossings with fully spanning structures on the biotic integrity of stream reaches in the Big Manistee River watershed. We assessed biological data from these systems and applied indices of biotic integrity (IBIs) based on fish and

macroinvertebrate communities to evaluate 1) the short – term (3- 6 years post treatment) effect of culvert replacement and 2) the overarching goal that culvert replacement improves stream biotic integrity.

Methods

Watershed description

The Big Manistee River watershed (4,900km^2) spans 11 counties (Fig. 4.1) and is typical of many Michigan watersheds which consist of sand and gravel dominated streams with low gradient, hydrologically-stable environments with temperatures moderated by groundwater inputs (Seelbach et al. 1997). The lower portion of the Big Manistee River is federally recognized as a wild and scenic river while upper portions of the mainstem and sections of tributaries are designated by the State of Michigan as Natural Rivers and Blue Ribbon Trout Streams. Three culvert replacement projects were conducted at stream crossings in the lower portions of the Big Manistee River watershed in the Sickle Creek (S1, one site), and Pine Creek (PC1 and PC2, 2 sites, Fig. 4.1) subwatersheds. The Sickle Creek subwatershed (8.95km^2) is approximately 41% forested and 11% wetland with 25% agriculture (NLCD 2006). The Pine Creek subwatershed (124.05km^2) is 75% forested with 11% wetlands and <1% classified as agriculture (NLCD 2006). For comparison to the culvert replacement sites we used control sites from adjacent subwatersheds that possessed similar biotic communities (Ogren Chapters 2 and 3). Pine Creek control sites (C-PC1 and C-PC2) were located in the Bear Creek subwatershed and the S1 control site (C-S1) was located in a small forested subwatershed that is also a direct tributary to the mainstem of the Big Manistee River (Fig.4.1). All of

these systems were coldwater trout streams that have similar community compositions including fish species from the families *Salmonidae*, *Cottidae* and *Cyprinidae*, and macroinvertebrate communities including intolerant coldwater taxa typical of the southern range of the Northern Lakes and Forests Ecoregion (Ogren Chapters 2 and 3).

<u>Culvert Replacements</u>

Culvert replacement of all three treatment sites was completed in late summer of 2005. Prior to restoration, the S1 road crossing was a seasonal fish passage barrier with two undersized and perched round culverts that restricted water flow and created a large plunge pool. A CON/SPAN® arch bridge structure was installed to allow unimpeded stream flow. The road was graded and outflow diversions were created that allowed for sedimentation and road runoff to flow away from the stream and into a forested area adjacent to the stream. The PC1 road crossing was a double culvert crossing that constricted stream flow. The culverts were not perched even at summer baseflow; however, they did restrict flow and during large storm events the stream would overtop the road. The undersized culverts were replaced with a flat bottom, natural substrate culvert that was designed to accommodate the stream during bankfull flows. The road was paved to reduce erosion and sloped to direct road runoff into an adjacent wetland. The PC2 site was a poorly aligned double culvert on a heavily used gravel road. To flow through the culverts the stream made a 90 degree bend next to the road causing severe erosion and road safety issues. A bottomless arch was installed with wing walls and the stream was allowed to migrate to a more natural path under the newly paved road.

<u>Data collection</u>

Data were collected for both macroinvertebrate and fish communities before and after culvert replacement. Macroinvertebrate assessments included before and after restoration at S1 (2004-2007), PC1 (2004-2007) and PC2 (2004-2007) in both spring (May) and fall (September) excepting PC2, which was not sampled in spring 2007. Macroinvertebrate sampling included two spring samples and one fall sample prior to culvert replacement. Data were collected annually for fisheries assessments at S1 (2002-2010), PC1 (2004-2010) and PC2 (2004-2008). Data collected in 2005 was prior to construction for fish community assessment.

Macroinvertebrate samples were collected using a multihabitat rapid bioassessment protocol (Barbour et al. 1999) with habitat types (e.g., riffles and pools) sampled using a D-frame kick net in approximate proportion to their representation of surface area. For each culvert replacement stream, sampling occurred seasonally in spring and fall of each year at six fixed sites (three reaches upstream and downstream of each culvert) with reach lengths determined as 40 x stream width. Each sampling reach was separated by a length of 40 x stream width. Macroinvertebrates were preserved and transported to a laboratory where they were subsampled to 200 individual organisms, and identified to genus or lowest possible taxonomic unit. For each sampling period, control reaches and culvert replacement sites were sampled within one week of each other.

Fish community assessment occurred annually in mid-summer at the same locations as macroinvertebrate sampling though only two upstream reaches were assessed for fish community. Single pass backpack electrofishing was conducted using a Smith-

Root LR-14 unit with all fish identified and enumerated. The control reach was sampled within one month of the date the impacted sites were sampled.

Taxa composition data from both macroinvertebrate and fish communities were assessed using IBIs for each sampling event. The Northern Lakes and Forests benthic community index (NLFBCI) was used to estimate biotic integrity based on macroinvertebrate communities (Butcher et al. 2003). For an estimate of biotic integrity based on fisheries data, an IBI for coldwater streams of the Upper Midwestern United States (Fish-IBI) was used (Mundahl & Simon 1999). These indices were found to be good representatives of community integrity and provide reliable scores for the Big Manistee River watershed (Ogren & Huckins 2014; Chapter 2; Chapter 3).

Experimental Design

Analyses were designed to test the effects of culvert replacement on the biotic integrity of the streams as estimated by biological data and resulting index scores. We used a Before-After-Control-Impact Paired Series (BACIPS) design (Stewart-Oaten et al. 1986; Osenberg et al. 2006) to assess the differences (BACI differentials) of the index scores before and after the culvert replacements occurred at each site. Index scores for each stream segment were calculated independently and then averaged across sub-samples taken both upstream (n=2 for fish and n=3 for macroinvertebrates) and downstream (n=3 for fish and macroinvertebrates) of the culvert. The upstream reach and downstream reach were considered independent sites and treated separately. Score differentials (control – impact scores) were calculated for the periods before and after culvert replacement and two way analysis of variance (ANOVA) was completed with a

pairwise multiple comparison procedure (Holm-Sidak method), executed in SigmaPlot version 12.2 (Systat Software Inc., 2012). The two factors were location (site) and period (before or after) with the dependent variable being the BACI differentials.

Results

Macroinvertebrate community composition remained relatively similar with a few additional taxa detected post culvert replacement at a few sites. Upstream and downstream reaches at S1 had additional *Ephemeroptera, Plecoptera and Tricopetera* (EPT) genera and downstream PC1 had a higher percentage of *Tricoptera* post culvert replacement. The ANOVA indicated a significant difference in NLFBCI differentials among locations ($F_{5,35}$=4.487 p=0.005) and a significant interaction between location and period ($F_{5,35}$=5.868, p=0.001) (Table 1). Mean NLFBCI scores improved after culvert replacement at S1 and shifted from the “poor” to the “fair” classification in the IBI, and in contrast, no similar temporal trends were observed for the control stream (Fig. 4.2a & 4.2b). Index scores displayed a difference in the BACI differentials after culvert replacement at S1 in both upstream (t=2.994, p=0.006) and downstream (t=3.368, p=0.003) locations. Although similar improvements were detected in the upstream reach of the PC1 site (a shift from “poor” to “fair” scores), the difference in BACI differentials between the two periods was not significant (t=0.842, p=0.408). The NLFBCI score for the upstream reach did vary approximately 10 points (20%) before culvert replacement and though scores steadily increased post construction they did not surpass the highest score from the before period (Fig. 4.2c). The downstream reach site scores improved and surpassed the control site scores (Fig. 4.2d); however, there was not a significant effect of

period (t=2.035, p=0.053) and mean site scores were in the "fair" category for the index before and after culvert replacement. The upstream reach at PC2 indicated no effect of period (t=1.491, p=0.149) while the downstream reach was significantly different with lower scores in the after period (t=2.544, p=0.018). Both upstream and downstream site mean scores post culvert replacement were generally in the "poor" range of the index (Fig. 4.2e & 4.2f).

Fish community composition displayed slight shifts post culvert replacement where *Lota lota* (Burbot), *Salvelinus fontinalis* (Brook Trout), and *Cottus bairdii* (Mottled Sculpin) increased upstream and downstream at S1. This increase likely drove the coldwater and percent Salmonids as Brook Trout metrics that improved index scoring. The ANOVA indicated a statistical difference in FISH-IBI differentials among location ($F_{5,29}$=62.380, p<0.001); however, there was no significant interaction between location and period ($F_{5,29}$=0.246, p=0.937) indicating that there was not an effect on FISH-IBI score post culvert replacement at any of the sites (Table 2). The mean S1 site scores were in the "good" range of the Fish – IBI before construction and remained there throughout the study period (Fig. 4.3a & 4.3b). Fish community at PC1 indicated a "poor" community throughout the study period for both upstream and downstream reaches (Fig. 4.3c & 4.3d). The community was composed of few Brook Trout with additional tolerant species such as *Semotilus atromaculatus* (Creek Chub) and *Umbra limi* (Central Mudminnow). PC2 was also composed of more tolerant species similar to PC1. PC2 varied in Fish-IBI classification, with scores fluctuating over time between "poor" and "fair" site classifications (Fig. 4.3e & 4.3f).

Discussion

Culvert improvement projects are often developed based on the potential, or the perception that they will restore ecological integrity. The three sites assessed in this study provided an opportunity to evaluate the effectiveness of road-stream crossings to improve biotic integrity as estimated by index scores. While culverts have been reported to fragment stream systems, exacerbate localized erosion and sedimentation, and alter flow regimes, we detected little improvement in overall biotic integrity after replacement of poorly constructed culverts in the subwatersheds we monitored. As described by Naiman (2013), there is an expectation that stream systems can be restored and this expectation provides an opportunity to monitor and assess restoration attempts and work towards integrated, collective actions that inform restoration science. Parameters to assess the success of restoration projects can be objective (Palmer et al. 2005) or subjective (Jӑhnig et al. 2011). Using a robust BACIPS design, we were able to objectively determine if the improvement of degraded road-stream crossings affected biotic integrity in the upstream and downstream reaches adjacent to the crossing. Working through this process of culvert replacement and assessment we have improved understanding of the watersheds and suggest multiple scaled approaches may be needed to evaluate and prioritize restoration projects proposed for watershed management.

Culvert improvements are important to overall connectivity and a more natural flow regime but without ties to larger over-all management actions, ecological changes in biotic integrity scores may not be realized. In Sickle Creek, NLFBCI scores improved while Fish-IBI scores did not. In this small subwatershed the reach scale improvement

was matched by assessment at the reach scale to see reach scale effects in the NLFBCI scores. In the larger Pine Creek subwatershed these reach scale actions, while necessary, may have acted as localized spot-treatments whose effects may have been masked by larger watershed scale processes. This concurs with the findings of Bohn and Kershner et al. (2002), which indicated that larger watershed issues need to be ameliorated before a change from restoration is realized on a reach level. This analysis supports conceptual suggestions that watershed scale assessments, implementation and monitoring should be the framework for stream restoration actions (Lake et al. 2007; Bernhardt & Palmer 2011).

Sites responded differently to the road stream crossing modifications. The smallest site assessed (S1) showed the most change in IBI scores while the largest site showed no change or a degradation in scores (PC2). The catchments where activities occurred had small amounts of development (urban/agriculture); however, active forestry practices were occurring in the region at the time of this study. Interestingly, the site with noted improvement occurred in a catchment with 41% forested and 24% agriculture land use/land cover, while the larger sites with little improvement in scores were in a 75% forested watershed with less than 1% agriculture. While both regional and local scale factors influence stream community composition (Frissel et al. 1986), there are conflicting reports of the strongest driver for stream biotic integrity in the literature. In watersheds dominated by forests and agriculture, watershed and reach scale drivers are both important for biotic integrity (Allan et al. 1997). In the Northern Lakes and Forests Ecoregion where our study sites were located, Wang et al. (2003) found that local

conditions best explained the patterns in fish occurrence though, watershed variables also explained a substantial amount of variation. They conclude that local scale habitat improvement would be more effective in less degraded watersheds and less effective in degraded watersheds. Though the Northern Lakes and Forests Ecoregion was considered by Wang et al. (2003) to be undisturbed, VanDusen et al. (2005) suggested that the condition of macroinvertebrate and fish communities in headwater streams in the same ecoregion retained signals of historic selection logging that may have occurred decades prior within the local catchments (i.e., years since logging explained community composition). In forested catchments, like Sickle and Pine Creeks there are many variables (local and watershed based) that can affect macroinvertebrate and fish communities.

Prior to culvert replacement the three sites differed in biotic integrity based on fish and macroinvertebrate data. Of the three stream crossings, S1 was the most physically degraded site due to its perched, undersized culverts prior to modification. The NLFBCI scored the site as "fair" to "poor" while the S1 fish community scored "good" with the Fish-IBI. The macroinvertebrate communities improved through time and were getting closer to a "good" classification three years post culvert replacement. Sundermann et al. (2011) suggests that sites that show improvement in macroinvertebrate communities may be in close proximity to potential colonists and an intact system. Though this subwatershed was forested (41%) with some agriculture in the upper reaches (25%), the riparian in the reach where work was completed was owned by the U.S. Forest Service and was forested with some wetland areas near the mainstem of the Big Manistee

River. The fish community scored “good” and remained in that condition post culvert replacement. The lack of response in the Fish-IBI was expected as the site started out with a “good” fish community both upstream and downstream and would have been hard to improve upon as composition was mainly intolerant coldwater species such as Brook Trout, Burbot, and Mottled Sculpin.

The Pine Creek subwatershed contained two culvert replacement sites, was larger in size than the Sickle Creek subwatershed, and land cover was dominated by forest (75%). The PC1 site NLFBCI scores improved post restoration downstream from “poor” to “fair”, likely driven by an increase in the EPT genera, while the upstream scores did not change. A more natural flow regime, coupled with reduced sedimentation may have led to improved NLFBCI scores downstream of the culvert. The Fish-IBI did not improve and remained in a “fair” to “poor” classification throughout the study. The largest site assessed, by catchment, stream width and discharge was PC2. Fish-IBI and NLFBCI scores failed to indicate a response upstream and actually decreased at the downstream reach. This site is in the mid-reaches of the watershed and while there was erosion of the road prior to the culvert replacement, no improvement was noted in biotic integrity directly within the stream reach we assessed.   As suggested by Bernhardt and Palmer (2011), in local restoration projects there are likely other issues in the watershed that are affecting overall biotic integrity recovery. Pine Creek is a designated cold water trout stream; however, it appears that there may be some thermal degradation (i.e., increased temperatures) that could be affecting the fishery (Ogren Chapter 3; Tonello 2011). This system had generally low Fish-IBI scores and one may have expected

improved communities after restoration; however, Rahel (2002) indicated that when a species pool is comprised of only resistant species (degraded system), local restoration does not tend to result in species composition improvements. Within the Pine Creek subwatershed there are two known impoundments in the headwaters and these, coupled with active forestry practices may account for some of the thermal stress and more tolerant fish species that were found in the stream reaches along with cold water species.

Culvert replacements are common and provide an opportunity to assess stream modification processes. Many of these types of projects may have value regardless of their over promised predicted benefits to biotic integrity. Success and failure of restoration projects has been debated and there are a multitude of metrics that can be used to justify projects and document success. We would put forth that having a clear objective, communicating that objective and monitoring for that objective is imperative. Often, projects are touted as providing more than they can deliver. Connectivity is imperative for streams to function properly and improving connectivity and removing barriers to migration can change patterns of fish movement. Unfortunately, culvert replacement is often tied to improved biotic integrity, which may happen with varying degrees of success at various scales. There are opportunities to be realistic in expectations and to evaluate the true impacts of restoration attempts. Improving natural stream hydrology and roadway safety are highly legitimate justifications for improvement of road crossings and they may also have long-term benefits to biotic communities and sustainability as well.

Implications for Practice

- Monitoring with indicators of biotic integrity can objectively determine if restoration practices had an effect on the overall biotic community when using a robust BACIPS design.
- Restoration projects should have a clearly defined goal and that goal should be the target used as a measure of success.
- Proposing realistic and justifiable physical and biological goals is a critical component of public acceptance of future restoration projects.
- Incorporating watershed scale assessment prior to restoration implementation will aid in directing ecologically significant efforts and determining appropriate expectations for recovery.

Table 4.1. Nortern lakes and forests benthic community index BACIP AVOVA results.

Source of Variation	DF	SS	MS	F	P
Location	5	121.510	24.302	4.487	0.005
Period	1	24.448	24.448	4.514	0.044
Location x Period	5	158.887	31.777	5.868	0.001
Residual	24	129.976	5.416		
Total	35	434.822	12.423		

Table 4.2. Mundahl and Simon coldwater fish index of biotic integrity BACIP ANOVA results.

Source of Variation	DF	SS	MS	F	P
Location	5	19258.630	3851.726	62.380	<0.001
Period	1	18.958	18.958	0.307	0.586
Location x Period	5	75.856	15.171	0.246	0.937
Residual	18	1111.423	61.746		
Total	29	20455.542	705.364		

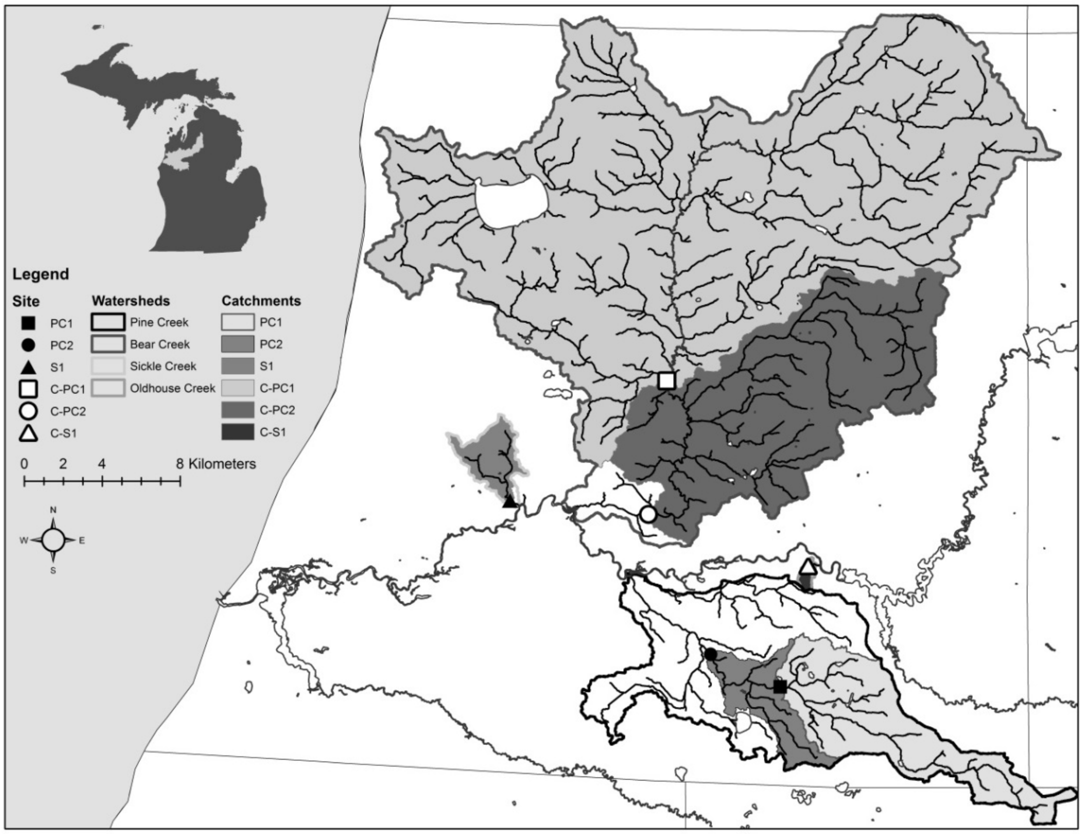

Figure 4.1. The Big Manistee River Watershed in the northern lower peninsula of Michigan, USA. The Bear Creek sub-watershed is shown with the catchments delineated to the two control sites located in that watershed (C-PC1 and C-PC2). The Pine Creek subwatershed is delineated with the catchments of two culvert replacement sites shown (PC1 and PC2). The Sickle Creek subwatershed is delineated with the S1 site and catchment shown and the control subwatershed for the S1 site is also shown (C-S1). All treatment locations are in solid symbols with the corresponding control reach designated by and open symbol of the same shape.

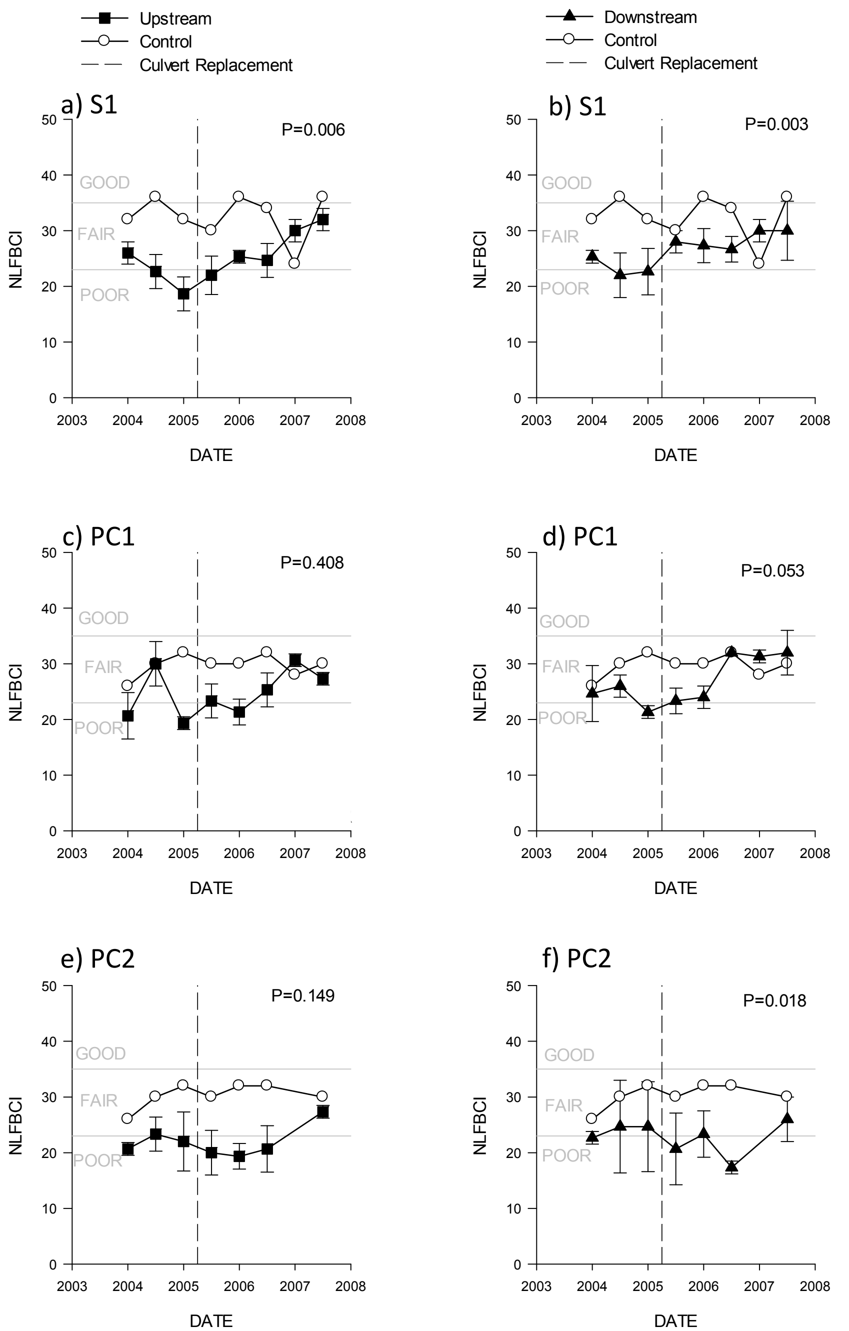

Figure 4.2. Results of the Northern Lakes and Forests benthic community index (NLFBCI) scores for the three culvert replacement streams. Graphs on the left are the upstream reach results and the right are the downstream reach results. Each panel has the treatment reach (solid symbol) as well as the control reach (open symbol) plotted over time with the culvert replacement year marked with a vertical dashed line. Individual points are means ± SD (control *n*=1, upstream *n*=3, downstream *n*=3). P-values represent the results from a BACIPS ANOVA.

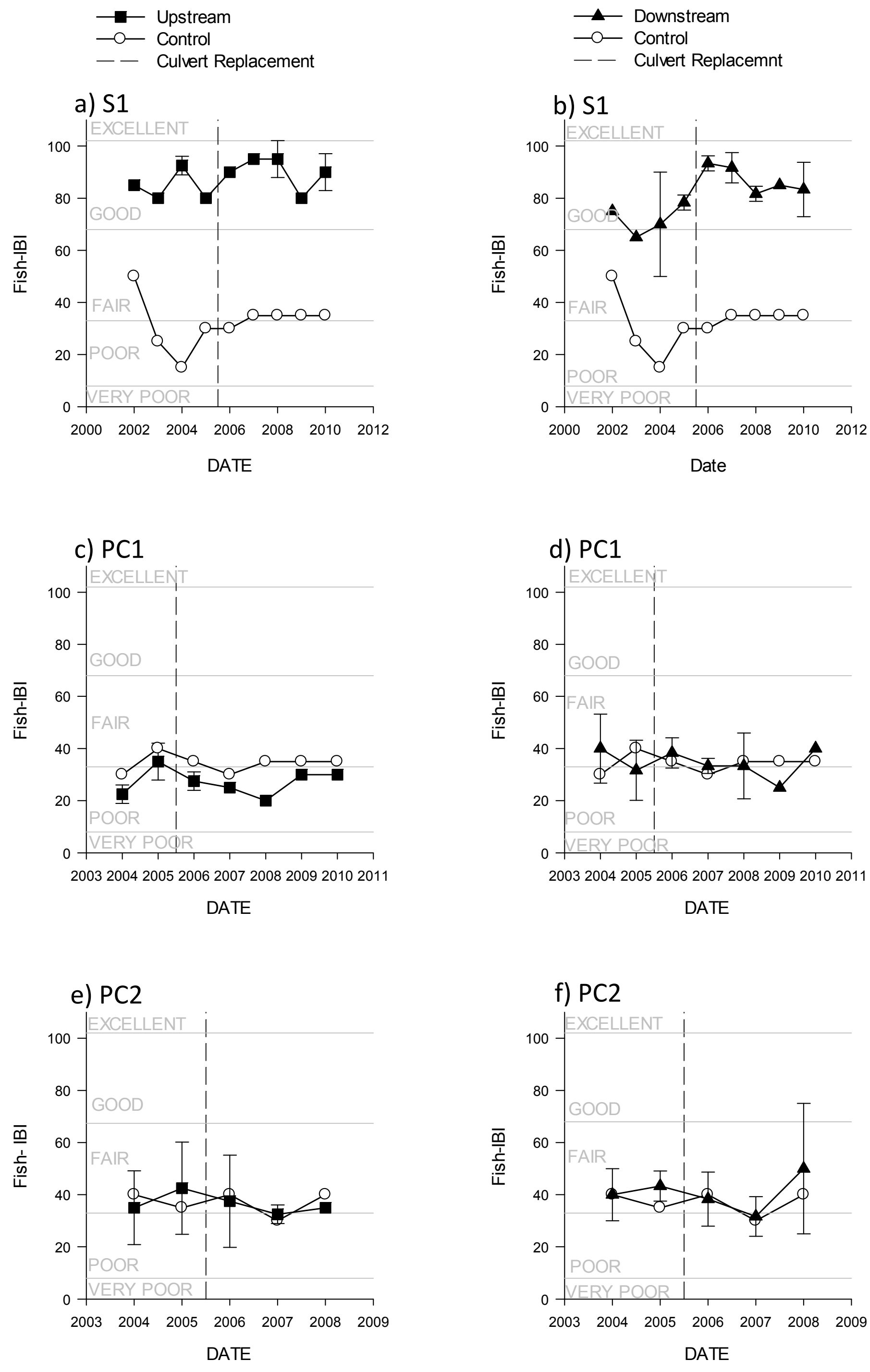

Figure 4.3. Results of the Fish-IBI scores for the three culvert replacement streams. Graphs on the left are the upstream reach results and the right are the downstream reach results. Each panel has the treatment reach (solid symbol) as well as the control reach (open symbol) plotted over time with the culvert replacement year marked with a vertical dashed line. Individual points are means ± SD (control n=1, upstream n=2, downstream n=3).

Literature Cited

Alexander, G., and J. D. Allan. 2007. Ecological Success in Stream Restoration: Case Studies from the Midwestern United States. Environmental Management **40**:245-255.

Allan, J. D., D. L. Erickson and J. Fay, 1997. The influence of catchment land use on stream integrity across multiple spatial scales. Freshwater Biology **37:** 149–161.

Barbour, M. T., J. Gerritsen, B. D. Snyder, and J. D. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish. EPA 841-B-99-002. United States Environmental Protection Agency, Washington, D.C.

Baron, J. S., N. L. Poff, P. L. Angermeier, C. N. Dahm, P. H. Gleick, G. H. Nelson, R. B. Jackson, C. A. Johnston, B. D. Richter, and A. D. Steinman. 2002. Meeting ecological and societal needs for freshwater. Ecological Applications **12**:1247-1260.

Berkman, H. E., and C. F. Rabeni. 1987. Effect of siltation on stream fish communities. Environmental Biology of Fishes **18**:285-294.

Bernhardt, E. S., and M. A. Palmer. 2011. River restoration: the fuzzy logic of repairing reaches to reverse catchment scale degradation. Ecological Applications 21:1926-1931.

Bernhardt, E. S., E. B. Sudduth, M. A. Palmer, J. D. Allan, J. L. Meyer, G. Alexander, J. Follastad-Shah, B. Hassett, R. Jenkinson, R. Lave, J. Rumps, and L. Pagano. 2007. Restoring rivers one reach at a time: results from a survey of US river restoration practitioners. Restoration Ecology **15**:482-493.

Bernhardt, E. S., M. A. Palmer, J. D. Allan, G. Alexander, K. Barnas, S. Brooks, J. Carr, S. Clayton, C. Dahm, J. Follstad-Shah, D. Galat, S. Gloss, P. Goodwin, D. Hart, B. Hassett, R. Jenkinson, S. Katz, G. M. Kondolf, P. S. Lake, R. Lave, J. L. Meyer, T. K. O'Donnell, L. Pagano, B. Powell, and E. Sudduth. 2005. Ecology - synthesizing US river restoration efforts. Science **308**:636-637.

Bernhardt, E. S., S. E. Bunn, D. D. Hart, B. Malmqvist, T. Muotka, R. J. Naiman, C. Pringle, M. Ruess, and B. Van Wilgen. 2006. The challenge of ecologically sustainable water management. Water Policy **8**:475-479.

Blakely, T. J., J. S. Harding, A.R. McIntosh, and M. J. Winterbourn. 2006. Barriers to the recovery of aquatic insect communities in urban streams. Freshwater Biology **51**:1634-1645.

Bohn, B., and J. L. Kershner. 2002. Establishing aquatic restoration priorities using a watershed approach. Journal of Environmental Management **64**: 355-36.

Briggs, A., and T. L. Galarowicz. 2013. Fish passage through culverts in Central

Michigan warmwater streams. North American Journal of Fisheries Management **33**:652-664.

Burford, D. D., T. E. McMahon, J. E. Cahoon, and M. Blank. 2009. Assessment of trout passage through culverts in a large Montana drainage during summer low flow. North American Journal of Fisheries Management **29**:739-752.

Butcher, J. T., P. M. Stewart, and T. P. Simon. 2003. A benthic community index for streams in the Northern Lakes and Forests Ecoregion. Ecological Indicators **3**:181-193.

Croke, J., S. Mockler, P. Fogarty, and I. Takken. 2005. Sediment concentration changes in runoff pathways from a forest road network and the resultant spatial pattern of catchment connectivity. Geomorphology **68**:257-268.

Davis, J.C. and G.A. Davis. 2011. The influence of stream-crossing structures on the distribution of rearing juvenile Pacific salmon. North American Benthological Society **30**:1117-1128.

Dudgeon, D., A. H. Arthington, M.O.Gessner, Z. Kawabata, D. Knowler, C. Lévêque, R.J. Naiman, A. H. Prieur-Richard, D. Soto, M.L.J. Stiassny, and C.A. Sullivan. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biological Reviews **81**: 163-182.

Frissell, C. A., and D. Bayles. 1996. Ecosystem management and the conservation of aquatic biodiversity and ecological integrity. Water Resources Bulletin **32**:229-240.

Frissell, C., W. Liss, C. Warren, and M. Hurley. 1986. A hierarchical framework for stream habitat classification: Viewing streams in a watershed context. Environmental Management **10**:199-214.

Jähnig, S. C., A. W. Lorenz, D. Hering, C. Antons, A. Sundermann, E. Jedicke, and P. Haase. 2011. River restoration success: a question of perception. Ecological Applications **21**:2007-2015.

Januchowski-Hartley, S. R., P.B. McIntyre, M. Diebel, P.J Doran, D. M. Infante, C. Joseph, and J.D. Allen. 2013. Restoring aquatic ecosystem connectivity requires expanding inventories of both dams and road crossings. Frontiers in Ecology and the Environment **11**:194-202.

Jones, J. I., J. F. Murphy, A. L. Collins, D. A. Sear, P. S. Naden, and P. D. Armitage. 2012. The impact of fine sediment on macroinvertebrates. River Research and Applications **28**:1055-1071.

Kemp, P. S., and J. R. O'Hanely. 2010. Procedures for evaluating and prioritizing the removal of fish passage barriers: a synthesis. Fisheries Management and Ecology **17**:297-322.

Khan, B., and M. H. Colbo. 2008. The impact of physical disturbance on stream communities: lessons from road culverts. Hydrobiologia **600**:229-235.

Lake, P. S., N. Bond, and P. Reich. 2007. Linking ecological theory with stream restoration. Freshwater Biology **52**:597-615.

MA (Millenium Ecosystem Assessment). 2005. Ecosystems and human well-being: Current State and Trends Vol 1. Island Press, Washington D.C.

MacPherson, L. M., M. G. Sullivan, A. L. Foote, and C. E. Stevens. 2012. Effects of culverts on stream fish assemblages in the Alberta Foothills. North American Journal of Fisheries Management **32**:480-490.

Meehan, W. R. 1991. Influences of forest and rangeland management on Salmonid fishes and their habitats. Special Publication 19. American Fisheries Society, Bethesda, Maryland.

Mundahl, N. D., and T. P. Simon, 1999. Development and application of an index of biotic integrity for coldwater streams of the upper midwestern United States. Pages 383–416 in T. P. Simon, editor. Assessing the sustainability and biological integrity of water resources using fish communities. CRC Press, Boca Raton, FL.

Naiman, R. 2013. Socio-ecological complexity and the restoration of river ecosystems. Inland Waters **3**:391-410.

Naiman, R. J., H. Décamps, and M. E. McClain. 2005. Riparia: ecology, conservation and management of streamside communities. Elsevier, San Diego, California, USA.

NLCD, 2006. National Oceanic and Atmospheric Administration, Coastal Services Center. 1995-present. The Coastal Change Analysis Program (C-CAP) Regional Land Cover. Charleston, SC: NOAA Coastal Services Center. Accessed at www.csc.noaa.gov/digitalcoast/data/ccapregional.

NRC. 1992. Restoration of Aquatic Ecosystems, National Academy Press, Washington, DC

Ogren, S.A., and C.J Huckins. 2014. Evaluation of suitability and comparability of stream assessment indices using macroinvertebrate data sets from the Northrern Lakes and Forestes Ecoregion. Ecological Indicators **40**:117-126

Ogren, S. A., and D. K. King. 2008. The effect of large woody debris on macroinvertebrate communities and epilithon detritus composition in a channelized headwater stream. Journal of Freshwater Ecology **23**:65-77.

Osenberg, C. W., B. M. Bolker, J. S. White, C. St. Mary, and J. S. Shima. 2006. Statistical issues and study design in ecological restorations: Lessons learned from

marine reserves. Pages 280–302 in D. A. Falk M. A. Palmer, and J. B. Zedler, editors. Foundations of restoration ecology. Island Press, Washington, D.C.

Palmer, M.A., E.S. Bernhardt, J.D. Allen, P.S. Lake, G. Alexander, S. Brooks, J. Carr, S. Clayton, C.N. Dahm, J. Follstad Shah, D.L. Galat, S. Gloss, P. Goodwin, D.D. Hart, B. Hassett, R. Jenkinson, K.M. Kondolf, R. Lave, J.L. Meyer, T.K. O'Donnell, L. Pagano, and E. Sudduth. 2005. Standards for ecologically successfull river restoration. Journal of Applied Ecology **42**:208-217.

Petersen, R. C. 1992. The RCE: A riparian, channel, and environmental inventory for small streams in the agricultural landscape. Freshwater Biology **27**:295-306.

Rahel, F. J. 2002. Homogenization of freshwater faunas. Annual Review of Ecology and Systematics **33**:291-315.

Roni, P., K. Hanson, and T. Beechie. 2008. Global review of the phyS1al and biological effectiveness of stream habitat rehabilitation techniques. North American Journal of Fisheries Management **28**:856-890.

Seelbach, P. W., M. J. Wiley, J. C. Kotanchik, and M. E. Baker. 1997. A landscape-based ecological classification system for river valley segments in lower Michigan (MI-VSEC version 1.0). Fisheries Research Report 2036. Michigan Department of Natural Resources, Ann Arbor.

Stewart-Oaten, A., W. M. Murdoch, K. R. Parker. 1986. Environmenatl impact assessment: “Pseudoreplication” in time? Ecology **67**:929-940.

Sundermann, A., S. Stoll, and P. Haase. 2011. River restoration success depends on the species pool of the immediate surroundings. Ecological Applications **21**:1962-1971.

Tonello, M. A. 2011. Status of the fishery resource, Pine Creek Michigan. Michigan Department of Natural Resources report 2011-22. Lansing.

VanDusen, P. J., C. J F. Huckins, and D. J. Flaspohler. 2005. Associations among selection logging history, brook trout, macroinvertebrates, and habitat in northern Michigan headwater streams. Transactions of the American Fisheries Society **134**:762-774.

Wang, L., J. Lyons, P. Rasmussen, P. W. Seelbach, T. Simon, M. J. Wiley, P. Kanehl, E. Baker, S. Niemela, and P. M. Stewart. 2003. Watershed, reach, and riparian influences on stream fish assemblages in the Northern Lakes and Forest Ecoregion, U.S.A. Canadian Journal of Fisheries and Aquatic Sciences **60**:491-505.

Warren, M. L., and M. G. Pardew. 1998. Road crossings as barriers to small-stream fish movement. Transactions of the American Fisheries Society **127**:637-644.

Wellman, J. C., D. L. Combs, and S. B. Cook. 2000. Long-term impacts of bridge and culvert construction or replacement on fish communities and sediment characteristics of streams. Journal of Freshwater Ecology **15**:317-328.

Wood, P. J., and P. D. Armitage. 1997. Biological effects of fine sediment in the lotic environment. Environmental Management **21**:203-217.