ENVIRONMENTALLY STRATIFIED SAMPLING DESIGN FOR THE DEVELOPMENT OF GREAT LAKES ENVIRONMENTAL INDICATORS

NICHOLAS P. DANZ[1,*], RONALD R. REGAL[2], GERALD J. NIEMI[1,3], VALERIE J. BRADY[1], TOM HOLLENHORST[1], LUCINDA B. JOHNSON[1], GEORGE E. HOST[1], JOANN M. HANOWSKI[1], CAROL A. JOHNSTON[4], TERRY BROWN[1], JOHN KINGSTON[1] and JOHN R. KELLY[5]

[1]Center for Water and the Environment, Natural Resources Research Institute, University of Minnesota Duluth, 5013 Miller Trunk Highway, Duluth, Minnesota, USA; [2]Department of Mathematics and Statistics, University of Minnesota Duluth, 10 University Drive, Duluth, Minnesota, USA; [3]Department of Biology, University of Minnesota Duluth, 10 University Drive, Duluth, Minnesota, USA; [4]Center for Biocomplexity Studies, South Dakota State University, Brookings, South Dakota, USA; [5]Mid-Continent Ecology Division, U.S. Environmental Protection Agency, 6201 Congdon Blvd., Duluth, Minnesota, USA

*(*author for correspondence, e-mail: ndanz@d.umn.edu)*

(Received 20 June 2003; accepted 24 February 2004)

Abstract. Understanding the relationship between human disturbance and ecological response is essential to the process of indicator development. For large-scale observational studies, sites should be selected across gradients of anthropogenic stress, but such gradients are often unknown for a population of sites prior to site selection. Stress data available from public sources can be used in a geographic information system (GIS), to partially characterize environmental conditions for large geographic areas without visiting the sites. We divided the U.S. Great Lakes coastal region into 762 units consisting of a shoreline reach and drainage-shed, and then summarized over 200 environmental variables in seven categories for the units using a GIS. Redundancy within the categories of environmental variables was reduced using principal components analysis. Environmental strata were generated from cluster analysis, using principal component scores as input. To protect against site-selection bias, sites were selected in random order from clusters. The site selection process allowed us to exclude sites that were inaccessible and was shown to successfully distribute sites across the range of environmental variation in our GIS data. This design has broad applicability, when the goal is to develop ecological indicators using observational data from large-scale surveys.

Keywords: anthropogenic stress, ecological indicators, GIS, Great Lakes, human disturbance gradient, sampling design

1. Introduction

The goal of biological monitoring and assessment is to measure and evaluate the consequences of human activities on biological systems. Ecological indicators have become important tools for the assessment and monitoring of natural resources, but management and monitoring programs have a history of using indicators that have lacked scientific rigor because of a failure to use a defined protocol for selecting the indicators (Dale and Beyeler, 2001). An additional limitation of many

Environmental Monitoring and Assessment **xxx:** 1–25, 2004.

current indicators is that they lack connection with specific anthropogenic stresses, making unclear the cause of ecosystem change and how to implement restorative management (Suter *et al*., 2002). Several recent methodological papers have proposed protocols and criteria for indicator development and selection (Hunsaker and Carpenter, 1990; Cairns *et al*., 1993; Barber, 1994; Jackson *et al*., 2000; Andreasen *et al*., 2001; Dale and Beyeler, 2001). A common thread among these papers is that indicators must be evaluated for properties including variability, error, discriminatory ability and responsiveness (to stress). Thus, to determine if indicators are robust, it is clear that at some point in the development process ecological data must be collected, analyzed, and interpreted. The process of deciding where to collect data is termed *sampling design* (Stevens and Urquhart, 2000). Because the sampling design imposes constraints upon the interpretation of the data, special care needs to be taken to ensure that the data meet the needs of the project (Overton and Stehman, 1995; Schreuder *et al*., 2001). Considerable effort has been devoted to appropriate sampling designs for monitoring programs that have the goal of reporting on ecological condition across a system of interest (Skalski, 1990; Urquhart *et al*., 1993; Larsen *et al*., 1994; Olsen *et al*., 1999, Stevens and Olsen, 1999; Herlihy *et al*., 2000). However, there is little information about sampling designs for detecting and understanding human-caused changes in biological systems (Karr and Chu, 1999), especially for observational studies with a wide geographic extent. The sampling design planned by Holland (1990), with results reported in Weisberg *et al*. (1993), is a notable exception.

Understanding the relationship between human activity and ecological response is essential to the process of indicator development; an indicator is not useful unless it varies predictably across a gradient of stress (Dale and Beyeler, 2001). Although potential indicators can be shown to be responsive to stress in laboratory or field experiments, for large observational studies the best way to demonstrate responsiveness is by evaluating the potential indicator at sites along a gradient from relatively pristine to highly disturbed (U.S. EPA, 1998). Statistical approaches such as curve fitting can then be used to describe relationships between stresses (x variables) and potential indicators (y variables). Studies that furnish a wider range of variation in the x variable are expected to give more precise estimates of the effect on y (Cochran, 1965). When a study is concerned with a single stress, the sampling design may be conceptually simple. Sites could be selected at either the extreme ends of the stress gradient or at several values along the stress gradient, depending upon the study objectives. In most circumstances, however, natural ecosystems are simultaneously influenced by many types of anthropogenic stress, making the sampling design more complex if the goal is to evaluate many potential indicators at several levels of stress and for many stresses.

Indicator development must also be concerned with understanding how patterns of response to anthropogenic stress are related to natural physical features and processes (Karr and Chu, 1999). Responses of interest must be isolated from noise introduced by natural spatial and temporal variability (Osenberg *et al*., 1994).

Indicators also should incorporate environmental conditions encountered during routine monitoring (Barber, 1994) and embody diversity in key environmental gradients across the ecological system of interest that are not anthropogenic stresses, such as soils, temperature, and hydrology (Dale and Beyeler, 2001). Hence, an additional consideration for indicator development is to distribute the sample across sources of environmental variation that may influence potential indicators but are not directly representative of stress.

How can sites be selected widely across many dimensions of stress and other environmental variation? Simple random sampling will tend to produce a sample in which the xs are spread throughout the range of x values in the population if the sample size is large, but this should not be left to chance if sample size is small or there is a need to ensure that a certain range of x values are covered (Royall, 1970). Systematic samples over large geographic regions also do not guarantee that important x variables are covered. This was recently demonstrated by Austin *et al.* (2001), who applied the sampling design of the U.S. EPA Environmental Monitoring and Assessment Program (EMAP) in the prairie pothole region and found that sample points tended to be clumped at one end of the range of landscape variables.

Alternatively, if environmental conditions are quantified for a study region, stratification can be used in the sampling design to ensure the sample is distributed across important gradients (Austin and Heyligers, 1991). Indeed, an impressive amount of data is available for many geographic regions and can inform us about a study area prior to sampling. Via the internet, one can quickly access publicly available data representing anthropogenic stresses and other types of natural environmental variation at various resolutions and spatial extents. For example, for the U.S. Great Lakes region, we obtained point locations for sewage treatment facilities, land use data at 30 m resolution (Vogelmann *et al.*, 2001), and estimates of agricultural runoff for United States Geological Survey hydrologic units (eight-digit HUC) (Seaber *et al.*, 1987). We propose such data can be used to partially characterize environmental conditions for sampling locations across large geographic areas without visiting sites, and can be used as stratification variables in a sampling design. Whether such data can also be used to evaluate responsiveness of potential indicators will depend upon the scale at which an indicator is influenced and whether the data are representative of the important stresses.

The objective of this paper is to describe a sampling design to develop indicators for the U.S. Great Lakes coastal region. In particular, we describe the way in which the coastal region was subdivided into observational units and the process we developed to ensure that the samples collected were distributed across a range of environmental conditions in the Great Lakes region. Results for stress/response relationships and indicator evaluation are not discussed here and will be reported elsewhere. Although our design is specific to the coastal region of the Great Lakes, the methodology has general applicability when the goal is to develop indicators using observational data from large-scale surveys.

2. Project Background

The Great Lakes Environmental Indicators (GLEI) project has the overall goal of developing indicators of ecological condition for the Great Lakes coastal region. Because of restrictions on funding and the size of the study area, our project was limited to the U.S. portion of the basin. Our study includes a wide variety of potential indicators representing individual, population, community, and landscape attributes to reflect the move toward using multiple measures to assess condition (U.S. EPA, 2002). The project was organized into five subcomponents that individually focus on ecosystem aspects related to current management concern in the coastal Great Lakes (Environment Canada and U.S. EPA, 2003): (i) birds and amphibians, (ii) diatoms and water quality, (iii) fish and macroinvertebrates, (iv) wetland vegetation, and (v) environmental contaminants. Numerous recent examples in the literature demonstrate indicator development using similar indicator categories (e.g., O'Connell *et al*., 1998; Simon *et al*., 2000; Cole, 2002; Fore and Grafe, 2002). Areas of focus within subcomponents were paired partly because of similarity of sampling protocols for taxonomic groups (e.g., both birds and amphibians are sampled using auditory surveys).

Indicators will be developed by approaching stress/response relationships from both stress and response perspectives. For example, we will (i) identify biological responses that indicate the presence or amount of a particular kind of stress, and (ii) identify which of the several stresses has the greatest influence on a particular biological response. Indicators will be developed for subcomponents individually (e.g., fish indicators of ecosystem condition) and by integrating indicators across subcomponents (O'Connor *et al*., 2000). Integrated measures are thought to better assess the ecological condition of an area (Karr and Chu, 1999; U.S. EPA, 2002). A challenge in the study design was to allow for maximum overlap in sampling locations, given different sample size requirements and sampling methodologies across the subcomponents. For example, the bird/amphibian subcomponent could visit many more sites than the other subcomponents because the sampling protocol takes much less time per site (Table I).The environmental contaminants subcomponent had a slightly different sampling design due to a much smaller sample size and different project goals compared to the other groups. The design for environmental contaminants is not addressed here and will be described elsewhere.

3. Study Area

The Great Lakes basin is an immense area that covers more than 30 million ha, holds 23,000 km^3 of water, and represents 18% of the world's surface freshwater (U.S. EPA and Government of Canada, 1995)(Figure 1). The basin is within one of the most industrialized regions of the world and contains about 10% of the U.S.

TABLE I

Targeted number of sites per cluster (stratum) for coastal ecosystem types for project subcomponents

		Subcomponents			
Coastal ecosystem	*n* Clusters	Birds and amphibians	Diatoms and water Quality	Fish and macro-invertebrates	Wetland vegetation
Nearshore uplands	60	3			
Nearshore wetlands	60	5			
Open	20		1	1	1
Protected	20		1	1	1
River-influenced	20		1	1	1
Embayments	20		1	1	
High-energy shoreline	20		1	1	
Total sites per subcomponent		480	100	100	60

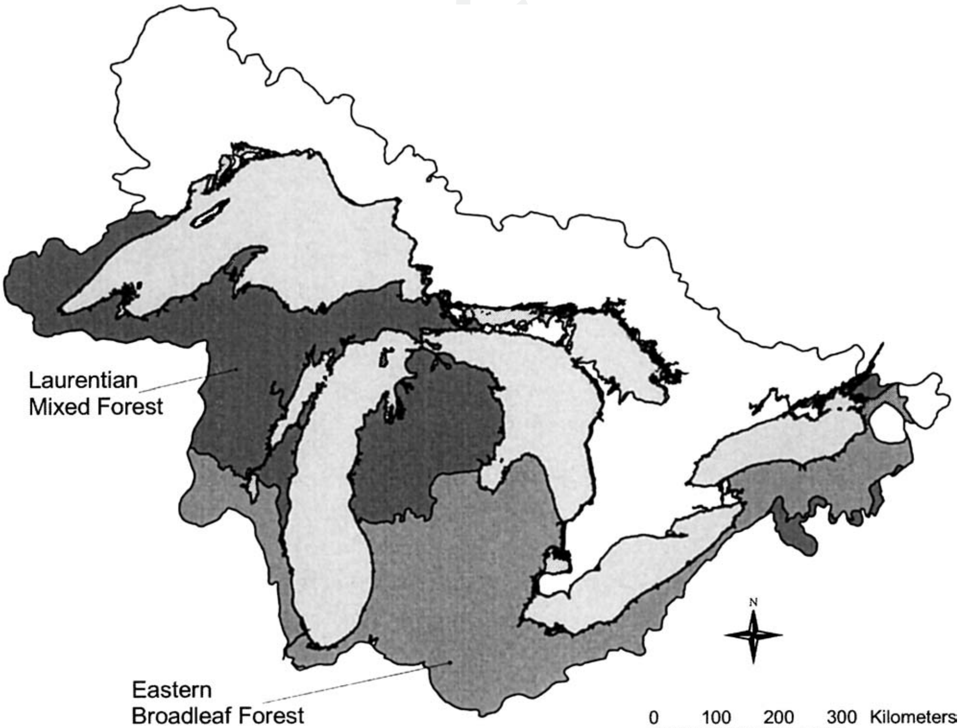

Figure 1. Watershed boundary of the Great Lakes basin, with the U.S. portion divided into two Ecological Provinces.

population. The region has been identified as an area of high ecological significance because of the presence of 131 elements (100 species and 31 communities) that are critically imperiled, threatened or rare on a global basis (The Nature Conservancy, 1994). The basin exhibits a wide range of environmental variation from relatively pristine wetlands and headwater streams to highly disturbed ecosystems near industrial areas. A substantial body of literature exists on the history and biota of the basin. Primary human pressures to coastal ecosystems in the basin include land use and landscape change (Brazner, 1997; Richards and Johnson, 1998; Detenbeck *et al*., 1999), climate change (Hartmann, 1990; Mortsch and Quinn, 1996; Magnuson *et al*., 1997; Kunkel *et al*., 1998, Mortsch, 1998), exotic species (Griffiths, 1993; Brazner *et al*., 1998; Brazner and Jensen, 1999), point and non-point source pollution (The Nature Conservancy, 1994), atmospheric deposition (Vitousek *et al*., 1997; Nichols *et al*., 1999), and various hydrological modifications (e.g., dredging, breakwaters, docks, harbors).

4. Units of the Great Lakes Coastal Region

4.1. COASTAL ECOSYSTEMS

Coastal regions of the Great Lakes basin subject to anthropogenic stress include land margins, nearshore waters, wetlands, estuaries, and bays (Minc and Albert, 1998; Keough *et al*., 1999, Detenbeck *et al*., 1999). Our units for indicator development are six types of ecosystems that occur in these regions. *Nearshore upland* is defined as the terrestrial region from the shoreline to 1 km inland. We defined *embayments* as shoreline indentations, where the width of the indentation mouth is less than the depth of the indentation, the total area is greater than 1 km^2, and there are fewer than two smaller embayments contained within. *High-energy shoreline* consists of lengths of shoreline not defined as embayment where emergent vegetation is not a dominant shoreline feature (e.g., sandy beach, cliffs, rock outcrops). Three types of coastal wetlands include *open-coast wetlands*, *drowned-river mouth and flooded-delta wetlands* (river-influenced), and *protected wetlands* as defined by Keough *et al*. (1999). The goal of sampling is to obtain representative measurements from the six types of coastal ecosystems, with project subcomponents having different sampling requirements for the ecosystem types (Table I).

4.2. SEGMENT-SHEDS

A primary step of study design is to identify the sampling frame—the list of all units that could potentially be selected for sampling (Figure 2)(Cochran, 1965). Conceptually, our sampling frame included all individual coastal ecosystem units (as defined above) in the U.S. Great Lakes basin. Because of the large size of the

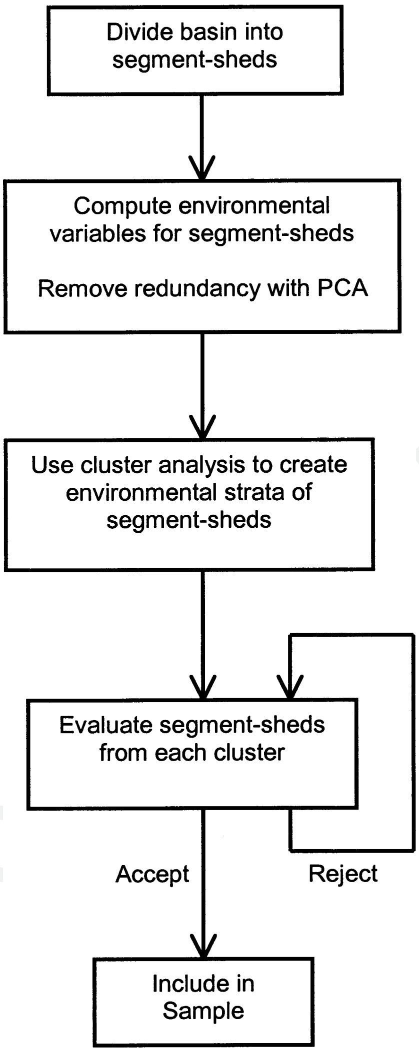

Figure 2. Sample design process.

basin it was impossible to delineate and compute environmental variables for the entire sampling frame prior to site selection. Instead, we defined a manageable number of coastal portions that contained our sampling units, and for the purpose of sampling design we computed environmental variables for the coastal portions rather than for ecosystem units individually (Figure 2). These coastal portions consisted of coastline segments with their associated drainage areas and accordingly are labeled "segment-sheds."

Segment-sheds were delineated in a two-step process using a geographic information system (GIS). First, segments were defined as lengths of shoreline beginning

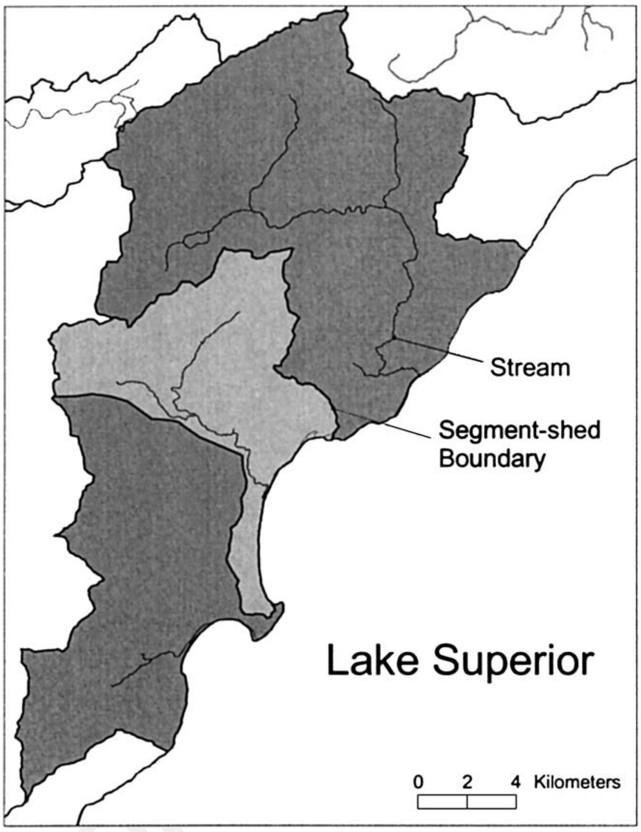

Figure 3. Example segment-sheds near Houghton, Michigan. Each segment-shed consists of the drainage area surrounding a second-order or higher stream.

and ending halfway between each second order or higher stream reaching the coastline using Reach File version 3.0, (RF3) (U.S. EPA, 1994). Second, the drainage area associated with each segment, including the stream and adjacent coastline, was delineated using the National Elevation Dataset (Gesch *et al*., 2002). This process resulted in 762 segment-sheds for the U.S. portion of the Great Lakes basin (Figure 3).We used a watershed-based approach to define coastal portions because coastal ecological condition is strongly influenced by upstream human activity (NRC, 2000). In addition, ecological assemblages are affected by geologic and anthropogenic factors operating at a watershed scale (Johnston *et al*., 1990; Hunsaker *et al*., 1992; Detenbeck *et al*., 1990, 1993; Richards *et al*., 1996; Johnson and Gage, 1997), and watersheds are being increasingly used as units for management (e.g., Total Maximum Daily Load [TMDL], Section 303[d]).

Using a GIS, we identified as accurately as possible the existence of individual ecosystem units within each segment-shed with United States Geological Survey

(USGS) digital orthophoto quadrangle images (DOQs) having 1-m resolution, 7.5-min USGS digital raster graphic images (DRGs), and existing wetland inventories (Herdendorf *et al*., 1981; Johnston, 1984; National Wetlands Inventory, 1990). All segment-sheds contained at least one ecosystem type, and some segment-sheds contained several. Nearshore uplands, high-energy shorelines, and large embayments sometimes crossed segment-shed boundaries. For the purpose of sampling design, we defined the portion of a coastal ecosystem within a segment-shed's boundaries as a discrete site. Coastal wetlands usually had well defined natural boundaries that occurred entirely within individual segment-sheds, and each individual wetland was considered a site. When a segment-shed contained sites of different ecosystem types, all of the sites were considered as candidates for sampling.

4.3. ENVIRONMENTAL VARIABLES

Using primarily public sources, we collected GIS data for one category of environmental variation not reflective of stress (i.e., soils), and for six primary categories of human disturbance that are of current management concern in the Great Lakes region (Environment Canada and U.S. EPA, 2003): agriculture (including agricultural chemicals), atmospheric deposition, land use and land cover, human population density and development, point and nonpoint source pollution, and shoreline modification. The latter six categories included a combination of natural land cover (e.g., forests, wetlands), along with types of human activities (e.g., amount of agricultural land), and specific stressors (e.g., agricultural nitrogen runoff). A total of 207 data layers were collected across the seven categories. The variables are principally land-based, which reflects our focus on developing coastal ecological indicators related to land-based human activities in the basin rather than stresses from the open water. These data were at various spatial resolutions, and it was necessary to rescale them to the resolution of segment-sheds. For example, land cover data existed as 30 m^2 pixels assigned to 1 of 20 classes; these data were summarized as the areal proportion of the segment-sheds comprised by each class. Table IIincludes several representative variables for each category, along with data sources and original resolution.

In summary, we computed 207 variables for 762 segment-sheds that were defined using drainage patterns. Because the primary sources of stress to coastal ecosystems are upstream human activities in coastal watersheds (Kennish, 2002), we are confident that using stresses computed for segment-sheds will result in our sampled sites, e.g., individual river-influenced wetlands, being spread over desired gradients of environmental stress. Future work will include computing stress variables corresponding to the individual coastal ecosystems that were actually sampled, which will further allow us to check how well stresses computed for segment-sheds correspond to stresses at individual sites within segment-sheds.

TABLE II
Representative GIS variables for the seven categories of environmental variation

Category	Resolution/Scale	Units	Variable	Agency	Program
Agriculture & Ag. chemical	8-digit HUC	Proportion	Area with animal facility nutrient application	USDA–NRCS	Performance Results Measurement System (PRMS)
	8-digit HUC	tons/ha/yr	Estimated soil loss	USDA–NRCS	Natural Resources Conservation Service (NRCS)
	8-digit HUC	$kg/km^2/yr$	P export from fertilizer into streams	USGS	NAWQA SPARROW
	County	Proportion	Area treated with agricultural herbicides	USGS	Census of Agriculture
Atmospheric deposition	Point	kg/ha/yr	Calcium deposition from atmosphere	Multi-agency	National Atmospheric Deposition Program (NADP)
	Point	kg/ha/yr	Chloride deposition from atmosphere	Multi-agency	National Atmospheric Deposition Program (NADP)
	Point	kg/ha/yr	Sulfate deposition from atmosphere	Multi-agency	National Atmospheric Deposition Program (NADP)
	Point	$kg/km^2/yr$	N export from atmosphere into streams	Multi-agency	National Atmospheric Deposition Program (NADP)
Land use and land cover	8-digit HUC	Proportion	Amount of grazing land	USDA–NRCS	National Resources Inventory (NRI)
	30 m × 30 m	Proportion	Evergreen Forest	USGS	National Land Cover Database (NLCD)
	30 m × 30 m	Proportion	Commercial/ Industrial/ Transportation	USGS	National Land Cover Database (NLCD)
	30 m × 30 m	Proportion	High Intensity Residential	USGS	National Land Cover Database (NLCD)

Point and non-point pollution	Point	#/shoreline km	Mine density in segment	USGS	Mineral resources spatial data
	8-digit HUC	$kg/km^2/yr$	N export from nonagricultural sources into streams	USGS	NAWQA SPARROW
	Point	#/ha	Active facilities with PAHs in wastewater	US EPA	National Pollutant Discharge Elimination System (NPDES)
	Point	#/ha	Density of facilities discharging into surface waters	US EPA	Toxic Release Inventory (TRI)
Human population density and development	8-digit HUC	Proportion	Percent of lacustrine emergent wetland change 1982–1992	USDA–NRCS	National Resources Inventory (NRI)
	Point	km	Distance to nearest Area of Concern	US EPA	Environmental Information Management System (EIMS)
	Census block	$\#/km^2$	Population density	US Census Bureau	US Census
	1:100,000	$\#/km^2$	Total road density	US Census Bureau	Topologically Integrated Geographic Encoding and Referencing (TIGER)
Shoreline modification	1:24,000–1:250,000	Proportion	Artificial (man-made) structures comprising the shoreline	NOAA	Great Lakes Environmental Research Laboratory (GLERL)
	1:24,000–1:250,000	Proportion	Amount of shoreline that is highly protected (70–100%)	NOAA	Great Lakes Environmental Research Laboratory (GLERL)
	1:24,000–1:250,000	Proportion	Amount of shoreline with nonstructural protection	NOAA	Great Lakes Environmental Research Laboratory (GLERL)
Soils	1:250,000	inches/inch	Maximum average available water capacity of soil	USDA–NRCS	State Soil Geographic Database (STATSGO)
	1:250,000	cm	Maximum average depth to bedrock	USDA–NRCS	State Soil Geographic Database (STATSGO)
	1:250,000	Proportion	Area with soils very poorly drained	USDA–NRCS	State Soil Geographic Database (STATSGO)
	1:250,000	Proportion	Area with clay	USDA–NRCS	State Soil Geographic Database (STATSGO)

5. Environmental Strata

Our general strategy for distributing sampling effort across a range of environmental conditions in the basin was to create groups (strata) of segment-sheds having similar environmental profiles, followed by selection of segment-sheds from strata using a randomized procedure (Figure 2). We based our strata on (i) Ecological Provinces (Bailey, 1989), (ii) coastal ecosystem types, and (iii) clusters of segment-sheds generated by the statistical treatment of environmental variables thought to influence potential indicators or ecological condition (Figure 4).Particular coastal ecosystems within a particular Ecological Province define subunits of the Great Lakes basin for which indicators will be developed. Clusters of segment-sheds with similar environmental conditions were used to distribute segment-sheds across the range of environmental variation represented in the GIS data.

5.1. ECOLOGICAL PROVINCES

As part of the National Heirarchical Framework of Ecological Units, the U.S. Great Lakes basin has recently been classified using criteria on the basis of ecological

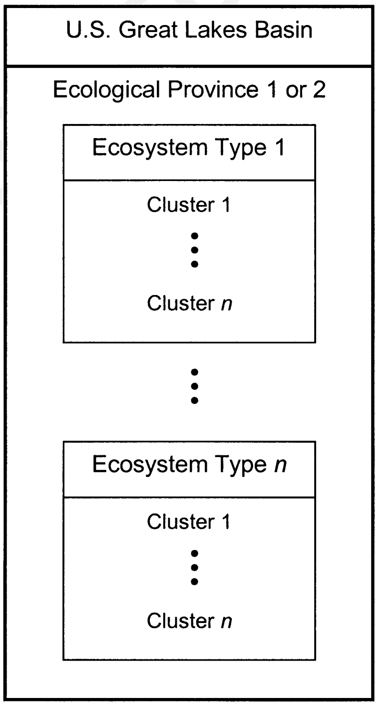

Figure 4. Schematic for environmental stratification. Clusters represent groups of segment-sheds with similar environmental conditions for each coastal ecosystem type in each Province and are strata from which sites were selected.

factors at different geographical scales (Bailey, 1989; Keys *et al*., 1995). The units delimit areas of different ecological capabilities and are being used to facilitate a sound approach to resource planning, management, and research (Cleland *et al*., 1997). Province is the highest level of the hierarchy that segregates the Great Lakes basin into two portions of nearly equal size, the Laurentian Mixed Forest and Eastern Broadleaf Forest Provinces (Figure 1). Preliminary analysis of our environmental data revealed major differences in primary environmental gradients between the Provinces. By using Provinces as environmental strata, we will be able to develop indicators for each Province, as well as for the entire basin. In addition, these strata allowed us to ensure that samples were well distributed, geographically (Stevens and Olsen, 1999). Although Provinces are divided into finer units (e.g., Sections and Subsections), the combination of the large extent of the basin and limitations on the number of samples prevented us from using the finer units as strata.

5.2. COASTAL ECOSYSTEM TYPES

We used our inventory of coastal ecosystem types to construct lists of segment-sheds that potentially contained each type of ecosystem; these lists were used as sets of segment-sheds for which further statistical analyses would identify strata (clusters) (Figure 4). For example, according to our inventory, 187 segment-sheds contained one or more river-influenced wetlands; segment-sheds that did not contain river-influenced wetlands were excluded from further stratification when selecting river-influenced wetland samples. Sampling across the range of environmental variation for each ecosystem type, enables the project subcomponents to develop indicators specific to those ecosystems (e.g., fish indicators of embayment condition), and for integration of indicators across taxonomic groups (e.g., multi-taxonomic indicators of Great Lakes coastal wetland condition).

5.3. CLUSTERS

Conceptually, each individual environmental variable represented a gradient across which we desired to distribute sampling effort. However, because the number of variables was large compared to the number of sites we could select, it was impossible to define strata for each variable. This also was unnecessary, because of the large amount of redundancy in the set of environmental variables. For the purpose of sampling design, we considered the seven categories of environmental variables equally important. That is, we wanted these categories to have equal influence in the development of environmental strata. We used principal components analysis (PCA) on the correlation matrix to remove redundancy and to reduce dimensionality, within each category of environmental variables (Table III)(SAS Institute, 2000). Prior to PCA, two types of transformations were applied to all variables to reduce the influence of outliers. Data that were proportions were subject to the arcsine square-root transformation; all other variables were transformed by first

TABLE III
Cumulative proportion of variance explained by the first five principal components for categories of environmental variables. The number of variables used as input to each PCA is indicated by *n*

			Principal component				
Province	Category	*n*	1	2	3	4	5
Laurentian mixed forest	Agriculture	21	0.72	0.81	0.86	0.90	0.93
	Atm. dep.	11	0.76	0.86	0.93	0.99	1
	Land cover	23	0.23	0.37	0.49	0.57	0.63
	Pop. Dens.	14	0.27	0.49	0.59	0.68	0.74
	Point source	79	0.38	0.47	0.54	0.60	0.66
	Shoreline Mod.	6	0.33	0.52	0.69	0.85	1
	Soils	53	0.24	0.42	0.52	0.57	0.63
Eastern broadleaf forest	Agriculture	21	0.41	0.60	0.72	0.82	0.86
	Atm. dep.	11	0.60	0.85	0.94	0.97	0.99
	Land cover	23	0.23	0.37	0.49	0.58	0.64
	Pop. dens.	14	0.29	0.43	0.55	0.65	0.73
	Point source	79	0.41	0.50	0.57	0.63	0.67
	Shoreline Mod.	6	0.29	0.50	0.68	0.85	1
	Soils	53	0.17	0.31	0.44	0.52	0.57

adding the minimum nonzero value for the variable and then calculating the natural logarithm. PCA can be thought of as rotation of the data so that observations are maximally spread along new axes (Rencher, 1995). The new axes (principal components, PCs) are uncorrelated and represent gradients of environmental variation within each variable category.

To generate the environmental strata, we used nonhierarchical k-means clustering with principal component scores as input variables (PROC FASTCLUS; SAS Institute, 2000). Because of differences between project subcomponents in sample numbers and types (Table I), separate cluster analyses were run for the bird/amphibian subcomponent for the other three subcomponents: diatom/water quality, fish/macroinvertebrate, and wetland vegetation. For simplicity, we describe the process for the latter three subcomponents only. Cluster analyses were carried out separately for segment-sheds containing the five ecosystem types to be sampled by these project subcomponents: three wetland types, high-energy shoreline, and embayments within the two Provinces (Table I), resulting in 10 cluster analyses. Clustering was carried out individually for ecosystem types to ensure that all segment-sheds within clusters contained the appropriate ecosystem, because at least one site from a segment-shed was to be selected from each cluster. We specified 20 clusters because this number was the largest common denominator for the number of sites that would be selected from each ecosystem type among the subcomponents (Table I). Eleven clusters were specified for the Laurentian

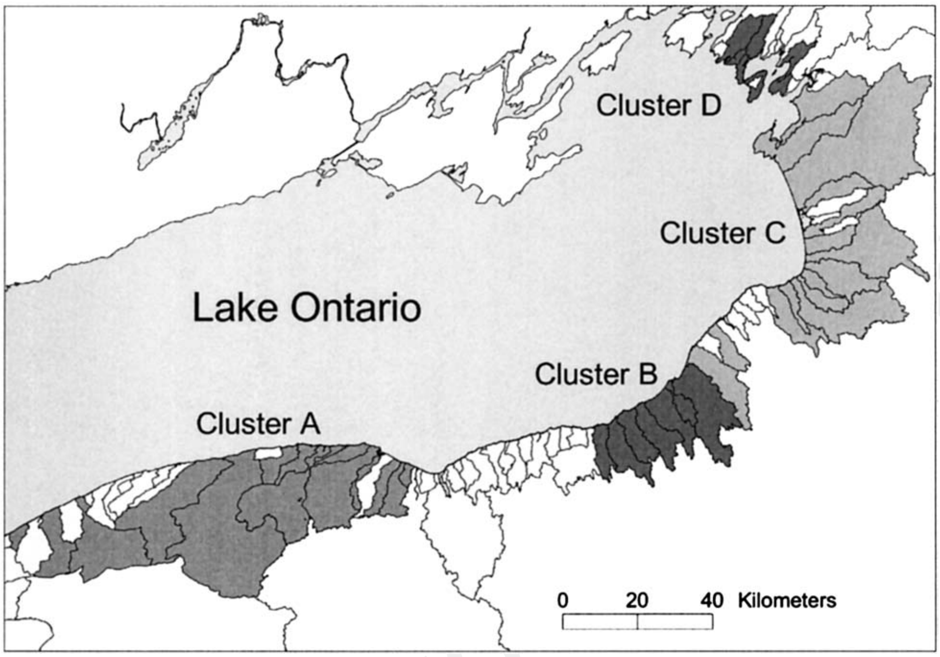

Figure 5. Example clusters for segment-sheds containing river-influenced wetlands in eastern Lake Ontario. Unshaded segment-sheds do not contain this type of wetland. Clusters A–C are in the Eastern Broadleaf Province, Cluster D is in the Laurentian Mixed Forest Province. **Q1**

Mixed Forest Province and nine were specified for the Eastern Broadleaf Province; the ratio 11:9 is equivalent to the ratio of segment-sheds in the two Provinces, respectively.

In FASTCLUS, variables with large variances have more effect on the resulting clusters than those with small variances (SAS Institute, 2000). Prior to clustering, we standardized the principal component scores for the PCs that explained 90% of the variation within each environmental category (Table II). This amounted to 65 PCs for the Laurentian Mixed Forest Province and 72 PCs for the Eastern Broadleaf Province. We then rescaled the scores by multiplying by the square root of the proportion of variance explained for the corresponding component. This had the effect of equalizing the total variance for all categories, while allowing for the PCs with greater original eigenvalues to have greater variance. Thus, each category had equal influence on the clustering overall, but individual PCs from categories had influence relative to the amount of variance they explained. The resulting clusters were strata having segment-sheds with a similar environmental profile, and the clusters were spread across the range of environmental conditions present in the GIS data for each ecosystem type in each Province.

6. Site Selection

Our sampling units were individual coastal ecosystem units as described above (see 4.1., Coastal Ecosystems) rather than entire segment-sheds. The focus of our site selection process was to identify and choose sites appropriate for sampling within segment-sheds (Figure 2). A “site” refers to an individual coastal ecosystem. To span the range of environmental conditions, at least one site was selected from every cluster (Table I). We evaluated segment-sheds using aerial photos and maps in a GIS (see 4.2., Segment-Sheds) to locate individual ecosystem units within segment-sheds and to determine whether the units were accessible. Segment-sheds were evaluated one at a time in random order within a cluster to minimize bias due to any preexisting familiarity with the sites. If a segment-shed did not contain at least one accessible site, the segment-shed was rejected and another segment-shed from the same cluster was evaluated (Figure 2). If a segment-shed was found to contain one or more accessible sites, one site was chosen randomly and included in the sample. Only in a few instances did segment-sheds in fact contain more than one accessible ecosystem unit of the same type, e.g., two river-influenced wetlands in the same segment-shed. This process was repeated until the appropriate number of sites was selected for each cluster (Table I). If a cluster did not contain enough acceptable sites, segment-sheds were evaluated from other clusters having similar environmental profiles as judged by Euclidean distance from the centroid of the original cluster (SAS Institute, 2000). To maximize sampling overlap, the project subcomponents selected sites in a progression, with the bird/amphibian subcomponent selecting sites first followed by the other groups in decreasing order of sample size. During segment-shed evaluations, subcomponents gave priority to sites previously included in the samples of other groups.

A sample is defined as the group of sites selected for each coastal ecosystem type for each subcomponent. For example, the wetland vegetation subcomponent selected one site from each of 20 clusters for each of the 3 wetland ecosystem types (Table I). Thus, this subcomponent had a protected wetland sample, an open wetland sample, and a river-influenced wetland sample, each consisting of 20 sites, for a total of 60 sites altogether.

7. Sample Distribution

Ideally, sites would be distributed widely across every environmental variable used in site selection. To check the success of the sampling design in distributing sites along the gradients, we compared the range of variation present in each sample to the potential range of the variables used in cluster analysis for each of the two ecological Provinces ($n = 65$ variables for the Laurentian Mixed Province;

$n = 72$ variables for Eastern Broadleaf Province). For each variable, the potential range of variation was defined as 100 (percentiles) for each combination of ecosystem type and Province. The range covered by a sample was the difference in percentile scores between the segment-shed having the minimum and maximum values along each variable. For example, Figure 6shows the distribution of the protected wetland samples for three subcomponents along three variables, plotted together with the total possible distribution of segment-sheds containing protected wetland. We used the median percentile range covered by each sample across all the clustering variables for individual Provinces to represent the success of the sample design, with median ranges, nearer to 100, representing greater success.

The degree to which the samples spanned the range of variation was related to sample size and to the evaluation criteria used to accept or exclude sites. The bird/amphibian subcomponent had the most well-distributed samples, with the samples having a median percentile range above 90 for both Provinces (Table IV).This group also had the largest sample size (Table I). Most samples selected by the other subcomponents had a median percentile range over 80, with open-coast wetland samples being a notable exception (Table IV). We had difficulty selecting open-coast wetlands because they are poorly characterized on existing maps (Johnston and Meysembourg, 2002), and segment-sheds that were thought to contain open-coast wetlands prior to cluster analysis were found to be lacking such wetlands during site selection. In addition, areas of the Eastern Broadleaf Province portion of the basin that were formerly open-coast wetlands were often diked, which converted them to protected wetlands.

8. Discussion

Sampling designs for observational studies to detect and understand human-caused changes in biological systems should include explicit consideration of how to distribute sampling effort with respect to important environmental gradients. If the objective is to characterize the relationship between stress and biological response along entire stress gradients (e.g., curve-fitting), then it is necessary for the sites to span the gradients (Karr and Chu, 1999). Alternatively, studies to develop indicators by comparing measures taken at reference versus degraded sites will not be concerned with sampling the middle of stress gradients. Reference versus degraded designs would be least well served by simple random sampling, especially if the population of sites is normally distributed with regard to stress; a random sample would result in most sites in the middle of a stress gradient and few sites at the extremes. Because many present-day landscapes have a long and varied history of human activity, no single measure will adequately describe human influence (Fore, 2003). We have presented a general technique that uses detailed environmental stratification to ensure that sampling effort is distributed across many environmental

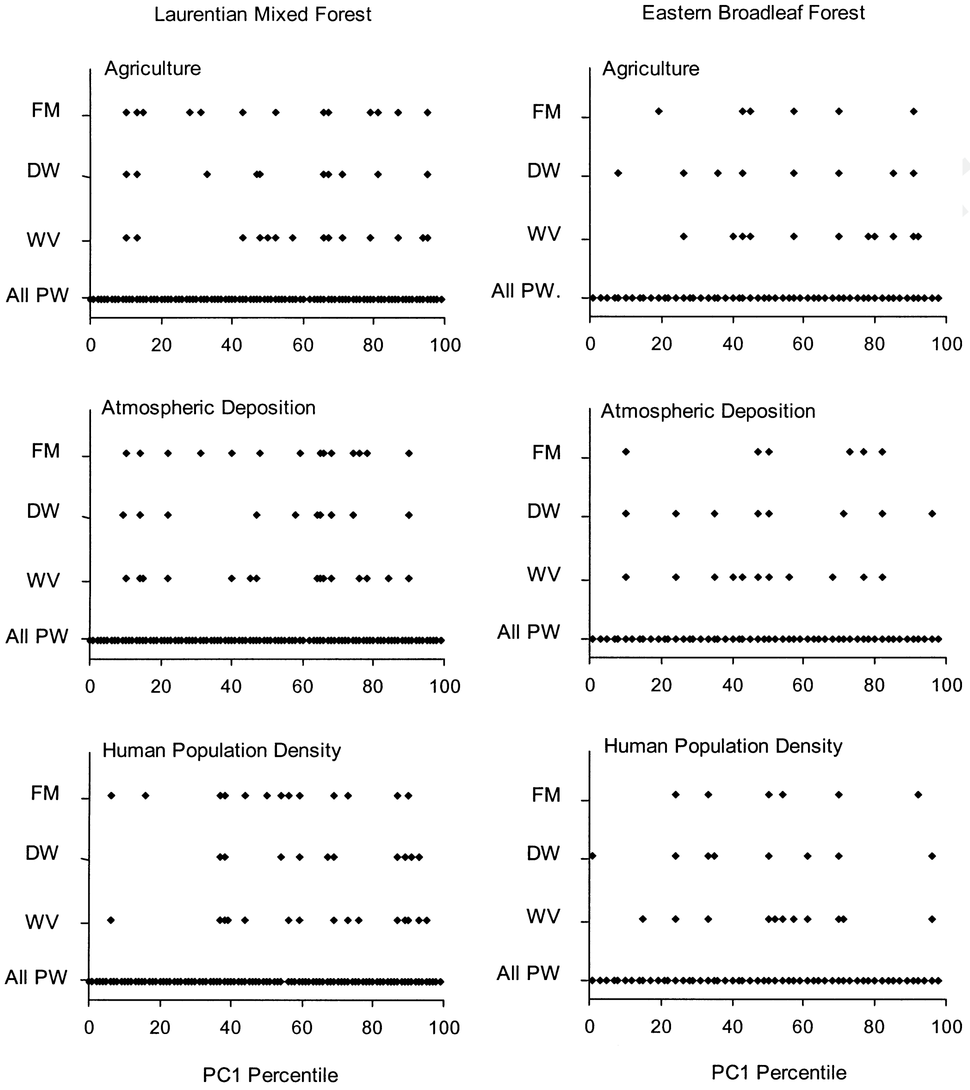

Figure 6. For three project subcomponents, the distribution of protected wetland samples along the first principal component of the agriculture, atmospheric deposition, and human population density varies. Scatter points represent individual segment-sheds. In each plot, the bottom row (All PW) shows values for all segment-sheds containing protected wetlands and represents the total possible range of variation along that component for each Province. Subcomponents: FM = fish/macroinvertebrate, DW = diatoms/water quality, WV = wetland vegetation.

TABLE IV

Median percentile range covered by samples along the set of environmental variables used in cluster analysis for each Ecological Province. Ranges were calculated along 65 variables for the Laurentian Mixed Province and 72 variables for the Eastern Broadleaf Province. Ideally, a sample would cover the entire range for all environmental variables. The nearer the median range is to 100, the more widely a sample is distributed across the range of variation present in the environmental variables

	Laurentian mixed forest				Eastern broadleaf forest			
Coastal ecosystem	Birds and amphibians	Diatoms and water quality	Fish and macro-invertebrates	Wetland vegetation	Birds and amphibians	Diatoms and water quality	Fish and macro-invertebrate	Wetland vegetation
Nearshore uplands	98				93			
Nearshore wetlands	99				98			
Open		74	73	83		55	55	25
Protected		91	89	91		70	81	82
River-influenced		89	90	93		88	84	86
Embayments		88	81			87	69	
High-energy shoreline		91	87			94	90	

gradients. The steps in site selection were to (i) divide the study area into a manageable number of units, (ii) compute environmental variables for the units and remove redundancy with PCA, (iii) cluster the reduced data, and (iv) select sites from clusters according to a set of evaluation criteria. In our design, we specified the number of clusters according to the number of sites that could be chosen by project subcomponents (e.g., sampling intensity was known *a priori*). Sites within clusters are likely to be spatially clumped due to autocorrelation in the clustering variables. Thus, selecting a small number of sites from each cluster will have the effect of minimizing spatial autocorrelation in the sample because neighboring sites would not likely be selected. In cases where sampling intensity can be flexible, cluster analysis can be used to identify how many samples are needed to sufficiently cover the gradients. In terms of cost, it may be beneficial to use the smallest number of clusters that capture most of the environmental variation (Austin and Heyligers, 1991).

The general principle of sampling along environmental gradients is scale-free, but whether the data used to distribute the sites will be appropriate for stress-response characterization depends on the scale at which an indicator is influenced. Fore (2003) showed that for several multimetric biological indexes, the more integrative the measure of anthropogenic disturbance, the greater the responsiveness. Principal components of a set of stress variables often have been used as integrated disturbance measures (Hughes *et al.*, 1998; Norton *et al.*, 2000; O'Connor *et al.*, 2002). Many individual metrics (e.g., wetland plant species abundance) would be expected to be responsive at a finer scale. In cases when data used for site selection are not appropriate for evaluating responsiveness, new data must be obtained from site-based measurements. The amount of publicly available environmental GIS data is impressive; many of the variables we used were available for the entire continental United States. However, for each variable that is obtained, substantial additional effort must be allocated to processing, rescaling, and archiving. One advantage our project had in this regard was that the effort was simultaneously spread across the project subcomponents, which were sharing a single sampling design and common objectives. It is easy to imagine how compiling a database with an exhaustive stressor list or for a large geographic region could become cost prohibitive.

The influence of human activity on Great Lakes coastal ecosystems continues to be of great concern. This is highlighted by recent work that identifies current major human pressures in the Great Lakes, including nutrient inputs, exotic species, contaminants, sedimentation, atmospheric deposition, land use, and human-population growth (Environment Canada and U.S. EPA, 2003). We were able to use knowledge of the important pressures to design a study particular to the management concerns of the Great Lakes coastal region by incorporating into our sampling design data, regarding these primary stresses. In combination with measures of stress collected during field sampling, the GIS data representing stresses at various resolutions can also be used during indicator development to evaluate the scale at which our biological responses are related to human activity. The focus of this paper has

Q2

been a sampling design that is applicable to any geographic region. In addition, the database of environmental variables and the summary of stress gradients are also valuable sources of information regarding land-based human activity in the Great Lakes basin. Much of the previous indicator research in the Great Lakes has focused on estuaries and the blue waters, with few studies focusing on the coastal margins. Our research explicitly considers the basin as a contributor to the condition of the lakes' margins. Such a view will offer insights into long-term protection and restoration of coastal ecosystems from land-based stresses.

Acknowledgements

We are grateful to Connie Host, Paul Meysembourg, Jim Sales, and Gerry Sjerven for assistance in compiling GIS data. Two anonymous referees provided valuable comments on an earlier draft. This research was supported by a grant from the U.S. Environmental Protection Agency's Science to Achieve Results (STAR) Estuarine and Great Lakes (EaGLe) program through funding to the Great Lakes Environmental Indicators (GLEI) Project, U.S. EPA Agreement EPA/R-8286750. This document has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred. This is contribution number xxx of the Center for Water and the Environment, Natural Resources Research Institute, University of Minnesota Duluth.

References

Andreasen, J.K., O'Neill, R.V., Noss, R. and Slosser, N.C.: 2001, 'Considerations for the development of a terrestrial index of ecological integrity', *Ecol. Indicators* **1**, 21–35.

Austin, M.P. and Heyligers, P.C.: 1991, 'Vegetation Survey Design, a New Approach: Gradsect Sampling', in: C.R. Margules and M.P. Austin (eds), *Nature Conservation: Cost Effective Biological Surveys and Data Analysis*, CSIRO, Melbourne, pp. 31–36.

Austin, J.E., Buhl, T.K., Guntenspergen, G.R., Norling, W. and Sklebar, H.T.: 2001, 'Duck populations as indicators of landscape condition in the prairie pothole region', *Environ. Monit. Assess.* **69**, 29–47.

Bailey, R.G.: 1989, 'Explanatory supplement to the ecoregions map of the continents', *Environ. Conserv.* **15**(4), 307–309.

Barber, M.C. (ed.): 1994, *Environ. Monit. Assess. Program Indicator Dev. Strat.*, EPA/620/R-94/022. U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC.

Brazner, J.C.: 1997, 'Regional, habitat, and human development influences on coastal wetland and beach fish assemblages in Green Bay, Lake Michigan', *J. Great Lakes Res.* **23**, 36–51.

Brazner, J.C., Tanner, D.K., Jensen, D.A. and Lemke, A.: 1998, 'Relative abundance and distribution of ruffe (*Gymnocephalus cernuus*) in a Lake Superior coastal wetland fish assemblage', *J. Great Lakes Res.* **24**, 293–303.

Brazner, J.C. and Jensen, D.A.: 1999, 'Zebra mussel [*Dreissena polymorpha* (Pallas)] colonization of rusty crayfish [*Orconectes rusticus* (Girard)] in Green Bay, Lake Michigan', *Am. Midland Nat.* **143**, 250–256.

Cairns, J., Jr., McCormick, P.V. and Niederlehner, B.R.: 1993, 'A proposed framework for developing indicators of ecosystem health', *Hydrobiol.* **263**, 1–44.

Cleland, D.T., Avers, P.E., McNabb, W.H., Jensen, M.E., Bailey, R.G., King, T. and Russel, W.E.: 1997, 'National heirarchical framework of ecological units', in: M. S. Boyce and A. Haney (eds), *Ecosystem Manage. Appl. Sustainable Forest and Wildlife Resour.*, Yale University Press, New Haven, CT, pp. 181–200.

Cochran, W.G.: 1965, 'The planning of observational studies of human populations', *J. R. Stat. Soc., Series A* **128**, 234–266.

Cole, A.: 2002, 'The assessment of herbaceous plant cover in wetlands as an indicator of function', *Ecol. Indicators* **2**, 287–293.

Dale, V.H. and Beyeler, S.C.: 2001, 'Challenges in the development and use of ecological indicators', *Ecol. Indicators* **1**, 3–10.

Detenbeck, N., Johnston, C.A. and Niemi, G.J.: 1990, 'Use of a Geographic Information System to Assess the Effect of Wetlands on Lake Water Quality in the Minneapolis/St. Paul Metropolitan Area', in: *Proceedings of the Minnesota Lake Management Conference*, Brainerd, MN, pp. 81–85.

Detenbeck, N.E., Johnston, C.A. and Niemi, G.J.: 1993, 'Wetland effects on lake water quality in the Minneapolis/St. Paul metropolitan area,' *Landscape Ecol.* **8**, 39–61.

Detenbeck, N.E., Galatowitsch, S.M., Atkinson, J. and Ball, H.: 1999, 'Evaluating perturbations and developing restoration strategies for inland wetlands in the Great Lakes Basin,' *Wetlands* **19**, 789–820.

Environment Canada and U.S. Environmental Protection Agency: 2003, *State of the Great Lakes 2003*, EPA 905-R-03-004.

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

Fore, L.S. and Grafe, C.: 2002, 'Using diatoms to assess the biological condition of large rivers in Idaho (U.S.A.)', *Freshw. Biol.* **47**, 2015–2027.

Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M. and Tyler, D.: 2002, 'The National
Q3 Elevation Dataset', *Photogrammatic Engineering and Remote Sensing* **68**(1), 5–12.

Griffiths, R.W.: 1993, 'Effects of Zebra Mussels (*Dreissena polymorpha*) on Benthic Fauna of Lake St. Clair', in: T.F. Nalepa and D.W. Scholesser (eds), *Zebra Mussels: Biology, Impacts, and Control*, Lewis Publishers, Chelsea, MI, pp. 415–437.

Hartmann, H.C.: 1990, 'Climate change impacts on Laurentian Great Lakes levels,' *Climate Change* **17**, 49–67.

Herdendorf, C.E., Hartley, S.M. and Barnes, M.D. (eds): 1981, *Fish and Wildlife Resources of the Great Lakes Coastal Wetlands Within the United States, Volumes 1–6*. U.S. Fish and Wildlife Service, FWS/OBS-81/02-v1-v6.

Herlihy, A. T., Larsen, D. P., Paulsen, S. G., Urquhart, N. S. and Rosenbaum, B. J.: 2000, 'Designing a spatially balanced, randomized site selection process for regional stream surveys: The EMAP mid-Atlantic pilot study', *Environ. Monit. Assess.* **63**, 95–113.

Holland, A.F. (ed.): 1990, *Near Coastal Plan for 1990, Estuaries,* EPA/600/4-90/033. U.S. Environmental Protection Agency, Office of Research and Development, Narragansett, RI.

Hughes, R.M., Kaufmann, P.R., Herlihy, A.T., Kincaid, T.M., Reynolds, L. and Larsen, D.P.: 1998, 'A process for developing and evaluating indices of fish assemblage integrity', *Can. J. Fisheries Aquatic Sci.* **55**, 1618–1631.

Hunsaker, C.T., Levine, D.A., Timmins, S.P., Jackson, B.L. and O'Neill, R.V.: 1992, 'Landscape Characterization for Assessing Regional Water Quality', in: D.H. McKenzie, D.E.

Hyatt and V.J. McDonald (eds), *Ecological Indicators*, Elsevier, New York, NY, pp. 997–1006.

Hunsaker, C.T. and Carpenter, D.E. (eds): 1990, *Ecological Indicators for the Environmental Monitoring and Assessment program*, EPA/600/3-90/060. Research Triangle Park, NC: U.S. Environmental Protection Agency, Office of Research and Development, Atmospheric Research and Exposure Assessment Laboratory.

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

Johnson, L.B. and Gage, S.H.: 1997, 'Landscape approaches to the analysis of aquatic ecosystems', *Freshw. Biol.* **37**, 113–132.

Johnston, C.A.: 1984, 'Wetlands in the Wisconsin Landscape,' *Wisconsin Nat. Resour.* **8**,4–6.

Johnston, C.A., Detenbeck, N.E. and Niemi, G.J.: 1990, 'The cumulative effects of wetlands on stream water quality and quantity: A landscape approach', *Biogeochem.* **10**, 105–141.

Johnston, C.A. and Meysembourg, P.: 2002, 'Comparison of the Wisconsin and National Wetlands Inventories', *Wetlands* **22**(2), 386–405.

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

Kennish, M.J.: 2002, 'Environmental threats and environmental future of estuaries,' *Environ. Conserv.* **29**, 78–107.

Keough, J.R., Thompson, T.A., Guntenspergen, G.R. and Wilcox, D.A.: 1999, 'Hydrogeomorphic factors and ecosystem responses in coastal wetlands of the Great Lakes,' *Wetlands* **19**(4), 821–834.

Keys, J.E., Jr., Carpenter, C.A., Hooks, S.L., Koeneg, F.G., McNab, W.H., Russell, W.E. and Smith, M.L.: 1995, *Ecological Units of the Eastern United States—First Approximation*, Technical Publication R8-TP 21. Map (scale 1:3,500,000). Atlanta, GA: U.S. Department of Agriculture, Forest Service.

Kunkel, K., Changon, S.A., Croley II, T.E. and Quinn, F.H.: 1998, 'Transposed climates for study of water supply variability on the Laurentian Great Lakes', *Climate Change* **38**, 387–404.

Larsen, D.P., Thornton, K.W., Urquhart, N.S. and Paulsen, S.G.: 1994, 'The role of sample surveys for monitoring the condition of the nation's lakes', *Environ. Monit. Assess.* **32**, 101–134.

Magnuson, J.J., Webster, K.E., Assel, R.A., Bowser, C.J., Dillon, P.J., Eaton, J.G., Evans, H.E., Fee, E.J., Hall, R.I., Mortsch, L.R., Schindler, D.W. and Quinn, F.W.: 1997, 'Potential effects of climate changes on aquatic systems: Laurentian Great Lakes and Precambrian shield region', *Hydrological Processes* **11**, 825–871.

Minc, L.D. and Albert, D.A.: 1998, 'Great Lakes Coastal Wetlands: Abiotic and Floristic Characterization', *Michigan Natural Features Inventory*, Lansing, MI.

Mortsch, L.D.: 1998, 'Assessing the impact of climate change on the Great Lakes shoreline wetlands', *Climate Change* **40**, 391–416.

Mortsch, L.D. and Quinn, F.H.: 1996, 'Climate change scenarios for Great Lakes Basin ecosystem studies', *Limnol. Oceanogr.* **41**, 903–416.

Natural Resource Council (NRC): 2000, *Ecological Indicators for the Nation*, National Academy Press, Washington, D.C.

National Wetlands Inventory (NWI): 1990, *Photointerpretation Conventions for the National Wetlands Inventory*, U.S. Fish and Wildlife Service, National Wetlands Inventory Center, St. Petersburg, FL.

Nichols, J., Bradbury, S. and Swartout, J.: 1999, 'Derivation of wildlife values for mercury', *J. Toxicol. Environ. Health, Part B* **2**, 325–355.

Norton, S.B., Cormier, S.M., Smith, M. and Christian Jones, R.: 2000, 'Can biological assessments discriminate among types of stress? A case study from the Eastern Corn Belt Plains ecoregion', *Environ. Toxicol. Chem.* **19**, 1113–1119.

O'Connell, T.J., Jackson, L.E. and Brooks, R.P.: 1998, 'A bird community index of biotic integrity for the Mid-Atlantic Highlands', *Environ. Monit. Assess.* **51**, 145–156.

O'Connor, R.J., Walls, T.E. and Hughes, R.M.: 2000, 'Using multiple taxonomic groups to index the ecological condition of lakes', *Environ. Monit. Assess.* **61**, 207–228.

Olsen, A. R. and Schreuder, T. H.: 1997, 'Perspectives on large-scale natural resource surveys when cause-effect is a potential concern', *Environ. Ecol. Stat.* **4**, 167–180. **Q4**

Olsen, A.R., Sedransk, J., Edwards, D., Gotway, C.A., Liggett, W., Rathburn, S., Reckhow, K.H. and Young, L.J.: 1999, 'Statistical issues for monitoring ecological and natural resources in the United States', *Environ. Monit. Assess.* **54**, 1–45.

Osenberg, C.W., Schmitt, R.J., Holbrook, S.J., Abu-Sara, K.E. and Flegal, A.R.: 1994, 'Detection of environmental impacts: Natural variability, effect size, and power analysis', *Ecol. Appl.* **4**, 16–30.

Overton, W.S. and Stehman, S.V.: 1995, 'Design implications of anticipated data uses for comprehensive monitoring programmes', *Environ. Ecol. Stat.* **2**, 287–303.

Rencher, A.C.: 1995, *Methods of Multivariate Analysis*, John Wiley & Sons, New York, NY.

Richards, C., Johnson, L.B. and Host, G.E.: 1996, 'Landscape-scale influence on stream habitats and biota', *Can. J. Fish Aquatic Sci.* **53**, 295–311.

Richards, C. and Johnson, L.B.: 1998, 'Landscape Perspectives on Ecological Risk Assessment', in: M.C. Newman and C.L. Strojan (eds), *Risk Assessment: Logic and Measurement*, Ann Arbor Press, Chelsea, MI, pp. 255–274.

Royall, R.M.: 1970. 'On finite population sampling theory under certain linear regression models', *Biometrika* **57**(2), 377–387.

SAS Institute: 2000, *SAS OnlineDoc*®, Version 8. SAS Institute Inc., Cary, NC, USA.

Schreuder, H.T., Gregoire, T.G. and Weyer, J.P.: 2001, 'For what applications can probability and non-probability sampling be used?', *Environ. Monit. Assess.* **66**, 281–291.

Seaber, P.R., Kapinos, F.P. and Knapp, G.L.: 1987, *Hydrologic Unit Maps*, U.S. Geological Survey Water-Supply Paper 2294, 63 p.

Simon, T.P., Jankowski, R. and Morris, C.: 2000, 'Modification of an index of biotic integrity for assessing vernal ponds and small palustrine wetlands using fish, crayfish, and amphibian assemblages along southern Lake Michigan', *Aquatic Ecosystem Health and Management* **3**, 407–418. **Q5**

Skalski, J.R.: 1990, 'A design for long-term status and trends monitoring', *J. Environ. Manage.* **30**, 139–144.

Stevens, Jr., D.L. and Urquhart, N.S.: 2000, 'Response designs and support regions in sampling continuous domains', *Environmetrics* **11**, 13–41.

Stevens, Jr., D.L. and Olsen, A.R.: 1999, 'Spatially restricted surveys over time for aquatic resources', *J. Agric. Biol. Environ. Stat.* **4**, 415–428.

Suter II, G.W., Norton, S.B. and Cormier, S.M.: 2002, 'A methodology for inferring the causes of observed impairments in aquatic ecosystems', *Environ. Toxicol. Chem.* **21**, 1101–1111.

The Nature Conservancy: 1994, *The Conservation of Biological Diversity in the Great Lakes Ecosystem: Issues and Opportunities*, The Nature Conservancy Great Lakes Program. Chicago, IL.

Urquhart, N.S., Overton, W.S. and Birks, D.S.: 1993, 'Chapter 3. Comparing Sampling Designs for Monitoring Ecological Status and Trends: Impact of Temporal Patterns', in Barnett and Turkmann (eds), *Stat. Environ.*, John Wiley & Sons, New York, NY, pp. 71–85.

U.S. Environmental Protection Agency: 1994, *The U.S. EPA Reach File Version 3.0 Alpha Release (RF3-Alpha) Technical Reference*, Office of Water, U.S. Environmental Protection Agency, Washington, D.C.

U.S. Environmental Protection Agency: 1998, ‘Guidelines for ecological risk assessment’, EPA/630/R-95/002Fa, *Federal Register* **63**(93), 26846–26924.

U.S. Environmental Protection Agency: 2002, *A Framework for Assessing and Reporting on Ecological Condition: An SAB Report*, EPA-SAB-EPEC-02-009, EPA Science Advisory Board, Washington, D.C.

U.S. Environmental Protection Agency and Government of Canada: 1995, *The Great Lakes: An Environmental Atlas and Resource*, U.S. EPA, Chicago, IL. 46 pp.

Vitousek, P.M., Aber, J., Howarth, R.W., Likens, G.E., Matson, P.A., Schindler, D.W., Schlesinger, W.H. and Tilman, G.D.: 1997, ‘Human Alteration of the Global Nitrogen Cycle: Causes and Consequences’, *Issues in Ecology 1*, Special publication, Ecological Society of America. 15 pp.

Vogelmann, J.E., Howard, S.M., Yang, L., Larson, C.R., Wylie, B.K. and Van Driel, N.: 2001, ‘Completion of the 1990s National Land Cover Data Set for the Conterminous United States from Landsat Thematic Mapper Data and Ancillary Data Sources’, *Photogrammetric Engineering and Remote Sensing* **67**, 650–652. **Q5**

Weisberg, S.B., Frithsen, J.B., Holland, A.F., Paul, J.F., Scott, K.J., Summers, J.K., Wilson, H.T., Valente, R., Heimbuch, D.G., Gerritsen, J., Schimmel, S.C. and Latimer, R.W.: 1993, *EMAP-Estuaries Virginian Province 1990 Demonstration Project Report*, EPA 620/R-93/006, U.S. Environmental Protection Agency, Environmental Research Laboratory, Narragansett, RI.

Queries

Q1. Author: Figure 5 is not cited in text.
Q2. Au: Pls. check this sentence.
Q3. Au: Pls. provide standard abbreviation.
Q4. Au: Not cited in the text.
Q5. Au: Pls. provide standard abbreviation.