An a priori process for selecting candidate reference lakes for a national survey

Authors: Herlihy, Alan T., and Sobota, Janel Banks

Source: Freshwater Science, 32(2) : 385-396

Published By: Society for Freshwater Science

URL: https://doi.org/10.1899/11-081.1

Freshwater Science, 2013, 32(2):385–396

DOI: 10.1899/11-081.1

An a priori process for selecting candidate reference lakes for a national survey

---

Alan T. Herlihy[1] AND **Janel Banks Sobota**[2]
Department of Fisheries and Wildlife, Oregon State University, Corvallis, Oregon 97331 USA

Todd C. McDonnell[3] AND **Timothy J. Sullivan**[4]
E&S Environmental Chemistry, Inc., P.O. Box 609, Corvallis, Oregon 97339 USA

Sarah Lehmann[5] AND **Ellen Tarquinio**[6]
Office of Water, US Environmental Protection Agency, 1200 Pennsylvania Avenue NW,
Washington, DC 20460 USA

Abstract. One of the biggest challenges when conducting a national-scale assessment of lakes, such as the 2007 US National Lake Assessment (NLA), is finding enough reference lakes to set appropriate expectations for the assessed sites. In the NLA, a random design was used to select lakes for sampling to make unbiased estimates of regional condition. However, such an approach was unlikely to yield enough minimally impacted lakes to use as reference sites, especially in disturbed regions. We developed a 3-stage process to select candidate reference lakes to augment the NLA probability sample in the northeastern USA (Northeast). Screening included a water-chemistry database filter, landuse evaluation, and analysis of aerial photographs. In the Northeast, we assembled a database of 2109 lakes >4 ha in surface area, of which 369 passed the water-chemistry screen. Of these, 220 failed the watershed landuse screen and 60 failed the aerial photograph screen, leaving a set of 89 optimal candidate reference lakes. Twenty of these lakes were sampled as potential reference lakes in the NLA. Based on a wide variety of indicators, NLA field measurements indicated that almost all (85–100%) of the chosen candidate reference lakes had least-disturbed water chemistry, although somewhat fewer had least disturbed physical habitat (74–79%) and biology (68–78%). Nevertheless, our 3-stage screening process was an efficient method for identification of good candidates for reference-lake sampling. The reference-lake selection process used in our study can be done in the office and relatively inexpensively. As such, it is very useful for large-scale regional or national studies encompassing areas too large to census. It also has the advantage of adding a level of consistency and quantification to the reference-site selection process.

Key words: reference condition, reference site, regionalization, biological condition gradient, regional assessment, lake, minimally disturbed, least disturbed.

Human activities have altered to widely varying degrees the physical, chemical, and biological conditions of lakes throughout the world. In lake assessments, investigators typically attempt to evaluate the effects of human activities on aquatic ecosystem structure and function and to describe conditions that would be expected to occur in the absence of human-caused disturbance within the lake watersheds. *Reference condition*, which describes the natural condition that might be expected to occur, is frequently used to provide a benchmark against which current condition can be compared (Reynoldson et al. 1997, Bailey et al. 1998, 2004, Reynoldson and Wright 2000). Data collected from lakes in reference condition are essential in the analysis of lake bioassessment data.

A single uniform approach for identifying and describing lake reference condition does not exist. In some regions, human disturbance is so pervasive that locating any lake that could be considered minimally

[1] E-mail addresses: alan.herlihy@oregonstate.edu
[2] janel.banks@oregonstate.edu
[3] todd.mcdonnell@esenvironmental.com
[4] tim.sullivan@esenvironmental.com
[5] lehman.sarah@epa.gov
[6] tarquinio.ellen@epa.gov

disturbed or close to historical condition is virtually impossible. However, identification of least-disturbed, or best-attainable, condition within a given region or study area should be possible regardless of land use. Stoddard et al. (2006) emphasized that the concept of reference condition can be used to represent a range of possible conditions, including least-disturbed, minimally disturbed, historical, or best-attainable condition. Each implies a different level of naturalness, and each requires a different approach when attempting to classify aquatic ecosystems according to level of disturbance.

It is probably safe to assume that no lake anywhere in the world is completely undisturbed. At a minimum, all lakes are influenced to some degree by atmospheric deposition of pollutants. In addition, most have been affected by introductions of nonnative species, nutrient additions, sedimentation, or other effects of human activities around the lakeshore or within the watershed. In some regions or subregions of the USA, minimally disturbed lakes may not exist. In such cases, candidate reference lakes may be those that exhibit least-disturbed condition. Such lakes represent the best-available conditions, given past and current landuse patterns and human activities. Least-disturbed condition is specified with explicit criteria (Hughes et al. 1986), such as that incorporated into the biological condition gradient (Davies and Jackson 2006), which can also vary across ecological regions (Stoddard et al. 2006). Candidate reference sites can be selected on the basis of existing water-chemistry data and the extent of observable (typically from maps and aerial photographs) human activities within the watershed. Such activities are reflected by the presence of roads or dwellings, certain land uses (urban, agricultural, mining), and other disturbances. We assume that the best candidates for minimally disturbed or least-disturbed conditions are found in watersheds that are largely free of observable human activities.

The 2007 National Lakes Assessment (NLA) was conducted by the US Environmental Protection Agency (EPA) as an ecological assessment of all lakes within the conterminous USA with the primary goal of evaluating ecological condition (USEPA 2009). The NLA was based on a probability sample of 1033 lakes that were used to infer condition for the entire lake population (USEPA 2009). Many aspects of the assessment of this survey data rely on a reference-condition approach (Bailey et al. 2004) to set expectations. Thus, in addition to sampling the probability set of lakes, selecting and sampling a set of minimally disturbed or least-disturbed lakes to serve as candidate reference lakes was an important component of the NLA.

In the USA, no commonly agreed upon list of reference lakes exists to be used for large-scale surveys like the NLA. Thus, selecting potential reference lakes for field sampling in the NLA required a priori selection of candidate reference lakes. In large surveys like the NLA, reference lakes often have been chosen based on lists of best professional judgment (BPJ) sites compiled from multiple sources that were selected based on the goals and objectives of the particular studies. These sites may be reference for other purposes or spatial scales, but may not be reference quality for a national bioassessment survey. BPJ reference lakes are often difficult to describe quantitatively because of the subjective judgment involved. Moreover, opinions about what represents reference condition vary widely depending on individual experience and expectations regarding the type of systems and disturbance regimes encountered. In our experience, selection of candidate reference sites by BPJ has been highly inefficient. In large-scale probability surveys of streams, such as the Mid-Atlantic Highlands Assessment project (Waite et al. 2000) and the national Wadeable Streams Assessment (Herlihy et al. 2008), >½ of the sampled hand-picked BPJ candidate reference streams were, in fact, classified as nonreference after the data were analyzed. The cost of sampling these nonrandom, nonreference sites is essentially wasted in that their data are not used for setting reference expectation or for making statistical inference to regional condition. Thus, strategies to decrease the percentage of a priori-selected candidate reference sites that are subsequently shown to be nonreference would be very useful for reducing survey costs. Here, we describe our efforts to develop a standardized process with a higher success rate for the a priori selection of candidate reference lakes when conducting large regional surveys.

Methods

We lacked the resources to develop and implement the full reference-lake screening process throughout the entire USA, so we piloted our approach in 1 region of the country, the Northeast (Fig. 1). In the NLA, reference lakes in other parts of the country were selected by BPJ and a much more limited screening. The major reasons for selecting the Northeast for piloting were the wealth of available water-chemistry data and the known diversity of lake ecological condition. The Northeast is very diverse in terms of topography and degree of human disturbance. Almost the entire region was glaciated, and the region contains a high density of lakes. The coastal area is relatively flat and densely populated

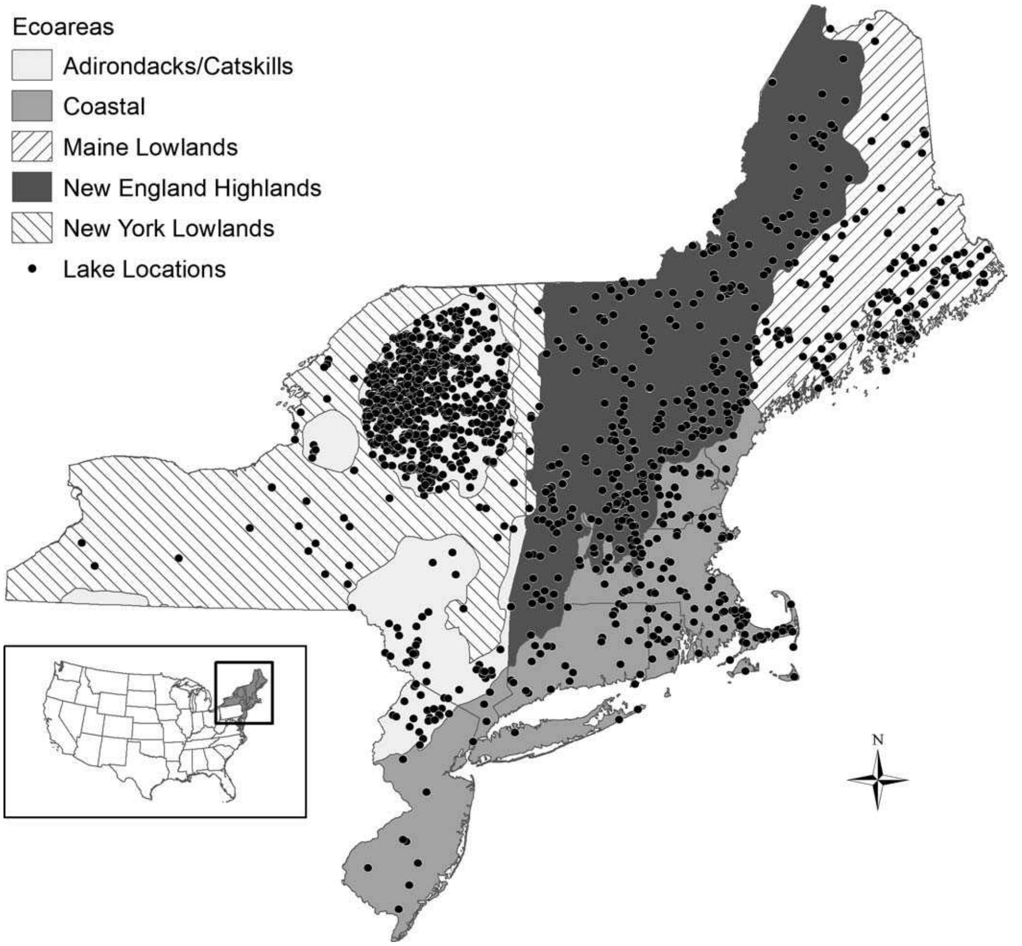

FIG. 1. Map showing the ecoarea classification of the Northeast and the locations of all lakes in the stage-1 water-quality database before screening. We defined 5 ecoareas based on Omernik (1987) Level III ecoregions. The New York Lowlands ecoarea consists of the Omernik Eastern Great Lakes and Hudson Lowlands and the Northern Appalachian Plateau and Uplands ecoregions. The Coastal ecoarea consists of the Omernik Northeast Coastal Zone, Atlantic Coastal Pine Barrens, Mid-Atlantic Coastal Plain, and Northern Piedmont ecoregions. The Maine Lowlands ecoarea is the same as the Omernik Laurentian Plains and Hills ecoregion. The Omernik Northeastern Highlands ecoregion was divided into 2 ecoareas, the Adirondacks/Catskills in New York and the New England Highlands in the other states. The small portions of the North Central Appalachians and the Ridge and Valley ecoregions in the Northeast were included in the Adirondacks/Catskills ecoarea.

and includes the New York City and Boston metropolitan areas. The northern portion of the region is mountainous (e.g., Adirondack, Green, and White mountains) reaching elevations of 1900 m. Because of its topography, this mountainous area is much less populated and, therefore, less affected by human disturbance than the rest of the region.

Within the NLA, we used a 3-stage screening process to select candidate reference lakes in the Northeast (Fig. 2). First, we assembled an initial list of candidate lakes by compiling all existing water-chemistry data and eliminating sites that failed a set of water-quality screening criteria (Stage 1). Second, we evaluated and further reduced the candidate list by examining a geographic information system (GIS) of watershed landuse and road-location data to eliminate lakes with significant human activity within the watershed (Stage 2). Third, we examined aerial photographs of lake shorelines to identify lakes having the least shoreline disturbance (Stage 3). We considered lakes that passed all 3 screening stages as candidate reference lakes for the NLA. Last, we examined how well this process worked by evaluating the field-documented ecological condition of 20 of the final candidate reference lakes sampled as part of the NLA.

Classification

A classification scheme was necessary for setting screening criteria and finding candidate reference

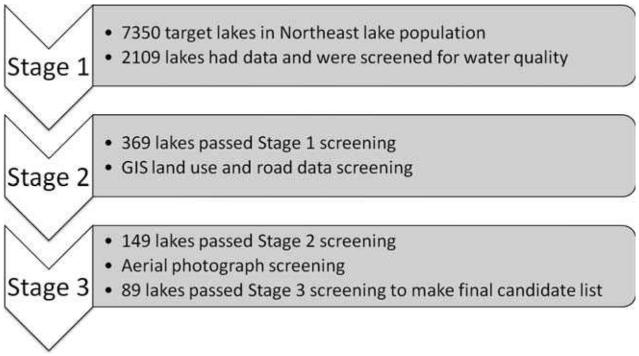

FIG. 2. Flow chart summarizing the 3-stage reference-lake screening process. GIS = geographical information system.

lakes representative of the entire regional lake population. We used ecoregions to help account for environmental variability. Ecoregion classification also works well to discriminate among degrees of human impact. In the Northeast, areas of high human perturbation tend to follow ecoregional boundaries (e.g., major cities are in the Coastal Zone ecoregion). We also used lake area as a classification variable to ensure that candidate reference lakes were selected across the lake size spectrum. For our screening process, we classified all lakes in the Northeast into 5 aggregated ecoregions (*ecoareas*) and 3 size classes. We used the level III ecoregions developed by Omernik (1987) to define ecoareas (Fig. 1). We selected lake-area size classes for the stratification as: ≥4–10 ha, ≥10–50 ha, and ≥50 ha. We used a 4-ha minimum surface area cutoff because 4 ha was the minimum lake size in the NLA target population. The combination of ecoarea and lake size classification resulted in 15 lake classes.

Screening stages

We eliminated any lake that failed any 1 of the 3 filters described below from consideration.

Stage 1: water-chemistry screening.—We compiled available lake water-chemistry data throughout the Northeast into one database. We defined a set of water-quality filters for each ecoarea based on concentrations of nutrients, acid neutralizing capacity (ANC), dissolved organic C (DOC), SO_4^{2-}, and Cl^-. We assembled lake water-chemistry data from major synoptic survey databases developed since 1985. We included in the database only those lakes having location information (latitude and longitude), and anion (SO_4^{2-}, Cl^-, NO_3^-), nutrient (total N, total P), and acid–base (ANC, pH, DOC) water-quality data.

The Eastern Lakes Survey (Landers et al. 1988), Adirondack Lakes Survey (Sullivan et al. 1990), and the Environmental Monitoring and Assessment Program Northeast Lake Pilot study (Whittier and Paulsen 1992) provided the vast majority of available information on lakes in the Northeast (2566 lakes). Acadia National Park in Maine, the University of Maine, and state agencies in Connecticut, New Hampshire, and Vermont contributed data for an additional 257 lakes to our final database. Among all these lakes, 2109 were ≥4 ha and were considered as initial candidates for screening. Statistical estimates from a previous probability survey in the region (Whittier et al. 2002) predict 7350 lakes ≥4 ha in the Northeast. Thus, our initial screening database contained ~29% of all lakes in the region.

We applied a water-chemistry screening procedure to all lakes in the compiled regional database in an effort to delete from consideration those lakes that showed evidence of human disturbance based solely on available lake water chemistry. We excluded lakes from further consideration if the concentration of any 1 of the disturbance-sensitive ions exceeded the range of expected values representative of relatively undisturbed conditions (Table 1). Chemical screening criteria varied by ecoarea but not by size class. We based our rationale for screening criteria on biogeochemical principles. We excluded lakes having both ANC $\leq$ 50 μeq/L and DOC $<$ 6 mg/L as they are probably affected by acidic deposition. Lakes with ANC $<$ 0 μeq/L are considered to be acidic and those with summer ANC between 0 and 50 μeq/L often become

TABLE 1. Water-quality criteria values for excluding candidate reference lakes. ANC = acid neutralizing capacity, DOC = dissolved organic C.

Ecoarea	Total P (μg/L)	NO_3^- (μeq/L)	Cl^- (μeq/L)	SO_4^{2-} (μeq/L)	ANC (μeq/L) and DOC (mg/L)
Adirondacks/Catskills	>10	>5	>20	>200	≤50 and <6
Coastal	>15	>5	>400	>200	≤50 and <6
Maine Lowlands	>10	>5	>400	>200	≤50 and <6
New England Highlands	>10	>5	>20	>200	≤50 and <6
New York Lowlands	>20	>5	>100	>300	≤50 and <6

acidic during spring snowmelt. We used the DOC requirement to distinguish naturally organically acidic lakes (DOC commonly $>$ 6 mg/L) from inorganically acidic lakes likely to have been affected by acidic deposition (Baker et al. 1991). We used Cl^- and SO_4^{2-} concentrations as general indicators of human disturbance. We based the screening criteria for these analytes on expected background concentrations in the Northeast (Table 1). Concentrations above the screening criteria are suggestive of anthropogenic activities in the watershed causing S or Cl^- inputs (e.g., mining, road salt, septic tanks). We set the Cl^- criteria higher in the coastal ecoareas because of natural atmospheric deposition of sea salt. In the New York Lowlands, a heavily disturbed region, we had to raise the Cl^- criteria from 20 to 100 μeq/L to obtain any final candidate reference lakes in the region. Thus, lakes in this ecoarea should be considered least-disturbed and not necessarily minimally disturbed. For nutrient screening, we used total P and NO_3^- as metrics to set ecoarea-specific levels (Table 1) above which we suspected anthropogenic nutrient enrichment.

Stage 2: watershed disturbance screening.—We used 1 arc second (~30-m) digital elevation models (DEMs) from the US Geological Survey (USGS) National Elevation Dataset (NED; http://seamless.usgs.gov/website/seamless/viewer.php) to delineate watershed boundaries. We processed the DEMs with ArcHydro Tools in the ArcGIS 9.1 environment (Environmental Systems Research Institute, Redlands, California). Terrain processing of the DEMs resulted in several newly generated grid data layers representing topographic and flow characteristics across the landscape. The 2 new layers that were central to watershed processing were the Flow Direction grid and the Flow Accumulation grid. The Flow Direction grid represents the direction that water will flow from any given 30 × 30-m cell, based on the elevations of surrounding cells. We used this layer to help understand the general direction of water flow in each lake watershed. We used the Flow Accumulation grid as the primary basis for locating the lake outlet. Higher values of flow accumulation reflect larger drainage area. We identified which flow accumulation value represented the lake (watershed) outlet by overlaying the hydrography layer on the Flow Accumulation grid. After we selected the lake outlet, the watershed was automatically delineated with ArcHydro Tools. If the initial watershed boundary did not coincide with the hydrography (lake polygons) or information found on underlying digital 1:24,000 USGS topographic maps, we generated new boundaries by selecting a new Flow Accumulation grid cell as the watershed outlet or by manually adjusting the watershed boundary via examination of the topographic maps.

After the watershed delineation, we acquired 30-m-resolution 1992 National Land Cover Data (NLCD; http://gisdata.usgs.net/website/MRLC/viewer.php) for each state in the Northeast. We used the watershed boundaries to clip the land-cover data to provide coverage for each watershed. We generated statistics (area and % cover) of land-cover types for each watershed. The NLCD data distinguishes 21 different land-cover types, which we grouped into 8 classes for our analysis (water, urban, barren/transitional, mining, forest/shrub, grass, agricultural, and wetlands). We tabulated % cover for the 3 land-cover types that were suggestive of human disturbance: urban, mining, and agricultural land use.

We calculated road length (km) and density (km/km^2) throughout each watershed for use as further screening criteria, under the assumption that human disturbance is correlated with human access represented by the presence of roads. We obtained road data layers representing the most detailed linework for the study area from individual state GIS data download locations (Table 2). For Maine, we merged 2 road layers to obtain the best available coverage for the state.

We did a secondary analysis to extract only the roads located within a 100-m buffer from the lakeshore. We also used this 100-m buffer zone for aerial photograph interpretation (see below). We conducted lakeshore buffer analyses because we assumed that human disturbance close to the lakeshore would be more likely to affect the chemistry and biology of the lake than would human disturbance elsewhere in the watershed. We clipped the same statewide roads data (Table 2) to the 100-m buffer surrounding each lake to provide the basis for calculating the length and density of roads within the lakeshore buffer zone.

Stage 3: aerial photograph screening.—In the final stage of the screening process, we used recent (2003–2005) aerial photographs of each lake to identify or confirm presumed disturbance. We obtained aerial photographs from multiple sources on a state-by-state basis (Table 2). We established guidelines for aerial photograph interpretation (Table 3) to identify various types of development, forestry activity, or the presence of previously unidentified roads within the 100-m lakeshore buffer. We assigned ranks to these disturbance features for each lake based on a visual approximation of the % buffer area affected by the feature (Table 4). We used this information to generate the final candidate lake list.

Table 2. Sources for state geographic information system (GIS) and aerial-photograph screening data.

Data set	State	Web download	Scale or year of photo
Road	Connecticut	http://www.dep.state.ct.us/gis/data/data.asp	1:24,000
	Massachusetts	http://mass.gov/mgis/laylist.htm	1:5,000
	Maine	http://www.maine.gov/megis/	1:24,000
	New Jersey	http://www.state.nj.us/transportation/gis	1:24,000
	New Hampshire	http://www.granit.sr.unh.edu/	1:24,000/1:25,000
	New York	http://www.nysgis.state.ny.us	Accurate to orthophotos
	Rhode Island	http://www.edc.uri.edu	1:5000
	Vermont	http://www.vcgi.org	Accurate to orthophotos
Aerial photographs	Connecticut	http://datagateway.nrcs.usda.gov/	2005
	Massachusetts	http://datagateway.nrcs.usda.gov/	2004
	Maine	http://www.maine.gov/megis/	2003–2005
	New Jersey	None required	—
	New Hampshire	http://datagateway.nrcs.usda.gov/	2004
	New York	http://www.nysgis.state.ny.us/	2003–2005
	Rhode Island	http://datagateway.nrcs.usda.gov/	2006
	Vermont	http://datagateway.nrcs.usda.gov/	2004

Evaluation of reference-lake quality

Twenty of our 89 final candidate reference lakes were field-sampled at the same time and the same way that probability NLA lakes were sampled. Data collection and condition analyses were the same as for the probability NLA lakes (USEPA 2009). The decision regarding which of the 20 final candidate reference lakes were sampled was driven by available funding and logistical constraints imposed by work-crew schedules, driving time, and locations of probability NLA lakes in the region. Because of available funding, 19 of the 20 lakes were in New England. Details of field sampling, laboratory methods, and assessment procedures were published by the USEPA (2007). We used these NLA results to assess the condition of our 20 final candidate reference lakes to see if they were, in fact, good reference lakes. In particular, we evaluated the condition indicators described below.

Table 3. Aerial photograph interpretation guidelines for evaluating lake shoreline disturbance features.

Disturbance	Features to identify
Residential development	Residences, maintained lawns, construction
Agricultural development	Cropland, pasture, barns, livestock, orchards, poultry operations, fencelines
Recreational development	Campgrounds, parks, golf courses, tennis courts, ski areas
Industrial development	Mines/quarries, industrial facilities, commercial facilities
Forestry	Recent logging, forest regeneration
Water development	Dams, water level fluctuations
Additional roads	Present/not present

Water quality.—A depth-integrated water-chemistry sample was collected from 0–2 m depth at the deepest part of each lake. Details on sampling, processing, and nutrient condition classification are described in Herlihy et al. (2013). Samples were also collected for microcystin and cyanobacterial analyses (see USEPA 2007, 2009 for details).

Physical habitat.—Field crews sampled habitat condition around the lake shoreline along 10 equidistant transects. Crews measured littoral-zone cover, riparian-zone vegetative cover, and riparian-zone disturbance.

Lake biota.—Field crews collected littoral macroinvertebrates at each of the 10 physical habitat transects with a sweep net. Phytoplankton and zooplankton were collected with tow nets at the water-quality station at the deepest part of each lake. A macroinvertebrate index of biotic integrity (IBI) was developed for the NLA, and IBI scores were classified as indicating good, fair, or poor condition. A combined

Table 4. Ranks applied to aerial photograph interpretation guidelines.

Description	Rank
No visual evidence of disturbance	0
Disturbance feature occurs, but appears to affect only a small percentage of the lakeshore area (<10%)	1
Disturbance feature appears to affect 10–25% of lakeshore	2
Disturbance feature appears to affect >25% of lakeshore	3

TABLE 5. Number of northeastern lakes in the candidate pool before and after water-chemistry screening (Stage 1).

		After screening			
		Lake size (ha)			
Ecoarea	Initial total	4–10	10–50	>50	Total
Adirondacks/Catskills	1179	58	65	31	154
Coastal	209	5	19	14	38
Maine Lowlands	195	10	14	27	51
New England Highlands	462	31	47	32	110
New York Lowlands	64	5	8	3	16
Total	2109	109	153	107	369

predictive observed/expected (O/E) model for phytoplankton and zooplankton species was also developed for the NLA (USEPA 2009). O/E models produce a score indicating reference taxa loss as a ratio of number of observed reference taxa (O) divided by expected number of reference taxa (E). For both the IBI and O/E measures, expectations were derived from least-disturbed site data.

Sediment diatoms.—A sediment core was collected from the deepest part of each lake, and sediment diatom assemblages were analyzed from the top and bottom sections of the core. Transfer functions were developed to predict current (core top) and historical (core bottom) lake-water total P and total N concentrations. Only lakes that had cores of sufficient length to ensure that core bottom assemblages were representative of historical conditions were used in our analyses (USEPA 2009).

Results

Effects of applying screening criteria

Stage 1: water chemistry screening.—The distribution of the 2109 lakes in the compiled dataset was uneven across the Northeast region (Fig. 1). Some of the unevenness was the result of natural differences in lake density among ecoareas, but much of it was a consequence of availability of data. For example, the Adirondack Lakes Survey was very extensive and included most of the lakes in the Adirondack portion of the Adirondacks/Catskills ecoarea. Thus, an especially high density of lakes occurred in our initial data set in the Adirondacks (Fig. 1).

After applying the screening criteria described in Table 1 to the 2109-lake dataset, 369 lakes remained for stage 2 analyses (Fig. 2, Table 5). Most lakes were eliminated by the total P (47%) and the Cl^- (33%) screens. Eliminations by these screens were not uniform across ecoareas. The P screen eliminated 57% of the lakes in the Adirondacks/Catskills but only 30–40% of the lakes in other ecoareas, whereas the Cl^- screen eliminated only 20% of the Adirondack/Catskill lakes but ~50% of the lakes in other ecoareas. The acidic deposition screen eliminated another 26% of the lakes in the Adirondacks/Catskills and New England Highlands. Only a few lakes failed the NO_3^- (7%) and SO_4^{2-} (4%) screens. Note that these totals do not add to 100% because some lakes would have been removed by several screening criteria.

Most (72%) of the 369 lakes with relatively undisturbed water chemistry were in highland areas in New England and the Adirondack and Catskill mountains in New York (Table 5). Even with the relaxed water-quality criteria for the New York Lowlands (Table 1), only 16 candidate reference lakes remained in the pool for that ecoarea after stage-1 screening.

Stage 2: watershed disturbance screening.—We calculated an overall GIS watershed-disturbance score for each site by summing 5 metrics: the proportion of the lake watershed in agriculture, urban, and mines/quarry land uses and watershed and lake-perimeter buffer road density (km roads/km^2 area). Overall disturbance scores varied widely among ecoareas. Scores for the Adirondack/Catskills and New England Uplands ranged from 0–0.31, where 0 represents no evidence of disturbance and higher numbers reflect greater disturbance. In the Coastal and New York Lowlands ecoareas, no lake scored 0 and maximum scores were 9.2 and 11.3, respectively. Scores in the Maine Lowlands ranged from 0–6.5.

We ranked the 369 remaining candidate lakes within each lake size × ecoarea class by overall GIS watershed disturbance score. The minimum target number of lakes to maintain within each class was 10 lakes. The most desirable lakes were those with no roads or land-cover types indicative of human disturbance in their associated watershed (0 disturbance score). For several classes, >10 lakes had these least-disturbed characteristics. We retained all lakes with 0 disturbance scores on the potential candidate

Table 6. Number of northeastern lakes remaining in the candidate pool before and after geographic information system (GIS) screening for watershed land cover and roads (Stage 2).

		After screening			
		Lake size (ha)			
Ecoarea	Initial total	4–10	10–50	>50	Total
Adirondacks/Catskills	154	23	11	10	44
Coastal	38	5	10	10	25
Maine Lowlands	51	10	10	10	30
New England Highlands	110	12	12	10	34
New York Lowlands	16	5	8	3	16
Total	369	55	51	43	149

list because no data were available at this point in the screening with which to discriminate disturbance characteristics among them. If a class did not contain ≥10 lakes with 0 disturbance score, we retained the 10 least-disturbed lakes in each class (based on disturbance score) on the potential candidate list, and we eliminated the remaining lakes. If a lake class had <10 remaining candidate lakes, we retained all of the remaining lakes. In the New York Lowlands, we had to pass on all of the stage-1 screened lakes to stage-3 screening. The candidate lakes in this ecoarea should be considered least-disturbed, compared with the other ecoareas where we considered the candidates to be minimally disturbed. The stage-2 screening by land cover and roads reduced the number of lakes on the list from 369 to 149 (Fig. 2, Table 6).

Stage 3: aerial photograph screening.—We screened the 149 candidate lakes remaining after application of the water-quality and GIS landuse screens by examining aerial photographs. We calculated a combined photo score by summing individual scores (each ranging from 0–3; Table 4) for each of the 7 categories listed in Table 3. Therefore, each lake could have a photo score of 0–21 with a 0 score indicating that no human disturbances were evident in the photographs. Of the 149 stage-2 screened lakes, 23 had photo scores of ≥3. These lakes were mainly in the Coastal and New York Lowlands ecoareas (Table 7). Seventy-five lakes had photo scores of 0. We considered these lakes our best set of candidate reference lakes and to be minimally disturbed. Only one 0-score lake occurred in each of 2 ecoareas, the Coastal and the New York Lowlands. Thus, these ecoareas were not well represented in the candidate list. Therefore, in these 2 ecoareas, we added the 14 lakes with photo scores of 1 to the 0-score lakes to compile the final potential reference lake list. We consider these 14 lakes to be least-disturbed reference lakes for their respective ecoareas. The final candidate reference lake list for NLA sampling in the Northeast consisted of 89 lakes (Fig. 2).

Evaluation of reference-lake quality

Twenty candidate reference lakes were sampled in the Northeast and had NLA condition assessment results. They included 9 lakes from the Maine Lowlands, 7 from the New England Highlands, 1 from the Adirondacks/Catskills, and 3 from the Coastal ecoarea. Only 1 of our candidate reference lakes was sampled in the Adirondacks/Catskills ecoarea because it was assumed that enough minimally impacted lakes would be found in the random NLA probability sampling for this ecoarea. All of the sampled candidate reference lakes had photo disturbance scores of 0 except for the 3 Coastal lakes, which

Table 7. Number of candidate reference lakes in each aerial-photograph disturbance category in stage-3 screening. A 0 score indicates no disturbances evident on photograph.

	Disturbance Score				
Ecoarea	0	1	2	3+	Total
Adirondacks/Catskills	37	3	3	1	44
Coastal	1	10	3	11	25
Maine Lowlands	15	7	4	4	30
New England Highlands	21	6	7	0	34
New York Lowlands	1	4	4	7	16
Total	75	30	21	23	149

TABLE 8. National Lake Assessment (NLA) lake-condition scores for the 20 candidate reference lakes sampled in the Northeast during the NLA. IBI = Index of Biotic Integrity, O/E = ratio of observed to expected reference-condition taxa.

Condition indicator	*n*[a]	% of candidate lakes
Biology		
Good macroinvertebrate IBI class	19	68%
Plankton O/E score > 0.85	18	78%
Physical habitat		
Least-disturbed riparian vegetation class	19	79%
Least-disturbed littoral vegetation cover class	19	74%
Least-disturbed riparian-zone condition class	19	79%
Water quality		
Least-disturbed total P class	20	85%
Least-disturbed total N class	20	90%
Least-disturbed chlorophyll *a* class	20	85%
Least-disturbed acidic deposition class	20	100%
Low-risk microcystin class	20	100%
Low-risk Cyanobacteria class	20	100%

[a] Sample size; not all lakes were sampled for physical habitat and biology

had photo disturbance scores of 1. The sampled candidate reference lakes covered a wide range of lake sizes (median lake area = 31 ha, range: 6.8–625 ha).

The vast majority of the sampled candidate reference lakes were assessed by the NLA as in least-disturbed or good condition for all indicators (Table 8). All of the lakes were classified in the low-risk category for microcystin, Cyanobacteria, and acidic deposition. A total of 85–90% of the 20 lakes were in the least-disturbed class for nutrients, and none of them were in the most-disturbed class. Sixty-eight percent of the lakes had good IBI scores and 78% had planktonic O/E scores >0.85 ($\geq 85\%$ of the expected plankton reference taxa were present in the lake). Least-disturbed habitat conditions were found at 74–79% of the 19 lakes that had physical-habitat measurements collected by the NLA (Table 8). Overall, 6 of the 20 lakes had least-disturbed conditions for all 11 assessment indicators listed in Table 8 and another 3 lakes had only 1 indicator not in least-disturbed condition. The 6 least-disturbed lakes were in the Maine Lowlands (3 lakes), New England Highlands (2 lakes), and Adirondack Mountains (1 lake).

Sediment-core data were used to infer current (core top) and past (core bottom) total P and N concentrations from sediment diatom assemblages in the sampled candidate reference lakes. Eighteen of 20 candidate reference lakes had cores of sufficient length and quality to be used in our analysis of historical condition. Diatom-inferred nutrient concentrations showed very little change over time, a result indicating that these lakes had not undergone significant anthropogenic nutrient additions in recent history. The median change in total P from core bottom to top was +0.6 μg/L, and the median change in total N was +37 μg/L (Table 9). Many lakes had diatom-inferred decreases in nutrient concentrations from historical conditions.

Current water quality in the sampled candidate reference lakes (Table 9) generally reflected the criteria used in the stage-1 screening (Table 1). All lakes had NO_3^-, SO_4^{2-}, and Cl^-concentrations below the screening criteria. In contrast, total P concentration was above the various stage-1 screening criterion in 9 lakes. Of these, 3 lakes had total P >20 μg/L (maximum = 35 μg/L). This result suggests that the screening criterion for total P (10–20 μg/L; Table 1) may be too low.

Discussion

Use of reference-site data is a fundamental requirement for most bioassessment surveys. However, the extensive nature of anthropogenic disturbance often has made finding reference sites a difficult process. This problem can increase the cost of sampling. Many lakes that are presumed in reference condition in fact are not, thus necessitating sampling more lakes than can actually be used in subsequent data analyses. The difficulty tends to increase with the scale of the survey. In more localized surveys, a large proportion of the study population can be censused or intensively examined for reference suitability. For example, in southeastern Arkansas, Justus (2010) considered all lakes with available water-quality data and used field reconnaissance and intensive sampling to identify a reference lake as the one with the least impairment in each of their 4 lake classes. In Portugal, Chaves et al. (2006) identified candidate lakes from maps and used field visits and validation with biological data to identify reference sites in the 6700-km^2 Mondego River basin. However, as the scale of study increases and becomes continent-wide, this level of intensive effort is not practical, and less-intensive screening is required.

No commonly accepted method exists to select reference sites for large-scale surveys. Other research-

TABLE 9. Water-quality statistics (sample size, minimum, 25th percentile, median, 75th percentile, and maximum) for the 20 candidate reference lakes sampled in the Northeast during the National Lake Assessment. Historical change was calculated from sediment diatom-inferred values using sediment core tops (current) and bottoms (past). The value is top minus bottom so positive values mean higher concentrations currently than in the past.

Metric	*n*[a]	Minimum	25th	Median	75th	Maximum
Water quality						
Total P (µg/L)	20	1.0	7.5	11.0	14.0	35.0
Total N (µg/L)	20	129	176	260	429	807
NO_3^- (µeq/L)	20	0.5	0.5	0.5	1.1	1.5
Chlorophyll *a* (µg/L)	20	0.7	1.7	3.0	6.4	31
SO_4^{2-} (µeq/L)	20	21.6	41.9	48.3	74.7	91.9
Cl^- (µeq/L)	20	5.7	8.4	19.9	29.2	275
Historical change[a]						
Total P (µg/L)	18	−5.8	−1.5	0.6	3.3	9.7
Total N (µg/L)	18	−97.0	−18.0	37.0	100	173

[a] Sediment cores not taken at all 20 sites

ers have tried a variety of multitiered screening approaches to select reference sites. Collier et al. (2007; New Zealand) used GIS to identify all stream segments with >85% unmodified upstream catchment vegetation cover and no roads or upstream impacts to set the initial candidate pool and then refined the list by systematic evaluation of aerial photographs. They also screened for accessibility and the presence of pest fish to identify the best sites in each of their stream classes. Yates and Bailey (2010) used a multistep process to identify stream reference sites in southwestern Ontario. First, they used a GIS to identify all 600- to 3000-ha watersheds that lacked urban land cover and to classify the watersheds into 6 groups. Then they used principal components analysis on GIS watershed disturbance metrics to rank each watershed in terms of degree of disturbance. A percentage of the least-disturbed watersheds was selected from each group to serve as reference for that group. Across Europe, a continent-wide reference-lake database has been assembled for studying total P and chlorophyll *a* (Cardoso et al. 2007, Carvalho et al. 2008, Poikane et al. 2010). Lakes were selected by individual nations following a common protocol that included a long list of nutrient-stress criteria. Some countries also used paleolimnological information and many countries used expert judgment in the review of the final reference lists.

Our stage-1 screening process involved compilation and filtering of existing water-quality data. We took full advantage of the wealth of existing lake data in the Northeast to narrow the search for reference conditions, but this approach would not be helpful or possible in regions of the world that lack considerable existing data. However, our stage-2 and -3 screening processes would still be useful for identifying potential reference sites even in places without large existing databases. In many regions of North America and Europe, existing water-quality data are plentiful, but considerable work might be needed to compile them into a consistent format. We opted to use water-quality information as a first screen because water quality is a more direct measure of actual human impact to the lake than are remote measures, such as land cover or aerial photographs. Using these data also allowed us to develop a robust short list for the more intensive GIS and aerial photograph screening stages. We could not have undertaken the photograph screening on all 7000-plus lakes in the Northeast given our existing resources.

Lists of candidate reference lakes often are compiled based on BPJ. This approach may not be problematic in smaller surveys where all judgments are carried out by the same person or team and are internally consistent. However, in surveys over large regions, BPJ often involves putting together lists from many people. Reference condition means different things to different people in different ecoareas, and BPJ lists can be influenced by unknown biases and inconsistent definitions of reference condition. Our experience with past national surveys is that a large proportion of BPJ-selected reference sites are non-reference when analyzed in a consistent manner based on field data (Herlihy et al. 2008).

We used the condition assessments made during the NLA to assess the quality of the candidate reference lakes we chose in the Northeast. Almost all (85–100%) of the chosen lakes had least-disturbed water quality. Somewhat fewer had least-disturbed physical habitat (74–79%) and biology (68–78%). This result was not surprising given that the screening was based on existing water-quality data, but not on data reflecting direct physical habitat or biological condition. Overall, 6 (30%) of the 20 candidate reference lakes sampled were considered least-disturbed for all 11 indicators shown in Table 8. For comparison, only 4 of the 89 (4.5%) probability-survey lakes in the Northeast were considered least-disturbed for each of the same 11 indicators. Thus, the reference screening process we used did a good job of identifying least-disturbed lakes, but it is far from perfect.

Paleolimnological analyses were used to assess historical nutrient concentrations at 18 of our 20 candidate reference lakes. Results showed that almost all of these lakes have not experienced large increases in nutrient concentrations over time. Three-fourths of the lakes had estimated historical total P increases <3.3 μg/L and total N increases <100 μg/L (Table 9). Other investigators have used paleolimnology to assess reference status. In Ireland, Leira et al. (2006) found that 68% of candidate reference lakes showed important deviations from reference condition resulting from acidification and nutrient enrichment. Similar analyses by Bennion et al. (2004) in Scotland showed that 19 of 26 lochs had historical increases in total P of >5 μg/L. They concluded that minimally impacted waters may be difficult to find in their study area under current conditions. Assessors of lakes have an advantage over assessors of streams in that paleolimnological tools are more readily available to document historical changes in lake chemistry conditions. Nevertheless, deciding how much historical change in nutrient concentration is acceptable in a reference site is difficult. Andersen et al. (2004) noted that this decision is ultimately a matter of policy, but has a large effect on the percentage of sites assessed as currently in good, moderate, or poor ecological condition.

The distinction between least-disturbed and minimally disturbed reference condition (Stoddard et al. 2006) is important. We consider our candidate reference lakes in the New England Highlands, Adirondacks/Catskills, and Maine Lowlands to be minimally disturbed. We were able to find large numbers of lakes in these ecoareas that passed water-quality screens and that had 0 human-disturbance scores. On the other hand, lakes in the New York Lowlands and Coastal ecoareas should be considered least disturbed. We had to relax the water-quality criteria and accept some lakes with low levels of disturbance to find final candidate reference lakes in those areas. Yates and Bailey (2010) used the least-disturbed concept to identify reference streams in Ontario. They used a human-activity gradient score to select the top percentage of sites in each of their groups to consider as reference. They decided the exact percentage of sites to use in each group by a statistical process to maximize the difference between reference- and test-site medians. On the extreme end of the least-disturbed continuum in Denmark, Baattrup-Pedersen et al. (2009) found that upon examination, none of the 128 a priori-selected reference streams fulfilled all reference criteria, and only 3 passed when the criteria were less strict. Baattrup-Pedersen et al. (2009) did not recommend relaxing criteria but concluded that a need exists for alternative methods to establish reference condition in Danish streams.

We applied the 3-stage reference-lake identification process successfully to identify reference lakes in the Northeast for use in the NLA. The process is based on both field-collected and remote-sensed data and can be applied in the office with relatively little expensive. As such, the process appears to be useful for large-scale regional or national studies encompassing areas too big to census. The approach has the advantages of cost savings associated with more accurate a priori identification of minimally or least-impacted conditions and adds consistency, objectivity, and quantification to the reference-site selection process.

Acknowledgements

This work was funded by cooperative agreement X-832547-01 between Oregon State University and the US EPA Office of Water. We thank Hilary Snook, Jim Kurtenbach, Scott Kishbaugh, Curt Stager, Barry Baldigo, Lisa Wahle, Robert Estabrook, and Neil Kamman for providing lake data and thoughts about reference condition in the Northeast. Gordon McPherson, Sam Mackey, and Deian Moore provided aerial photograph interpretation. We also thank those involved with the National Lakes Assessment for their insights and data sharing.

Literature Cited

ANDERSEN, J. H., S. J. CONLEY, AND S. HEDAL. 2004. Palaeoecology, reference conditions and classification of ecological status: the EU Water Framework Directive in practice. Marine Pollution Bulletin 49:283–290.

BAATTRUP-PEDERSEN, A., E. A. KRISTENSEN, J. JORGENSEN, J. SKRIVER, B. KRONVANG, H. E. ANDERSEN, C. C. HOFFMAN, AND L. M. K. LARSEN. 2009. Can a priori defined reference criteria be used to select reference sites in Danish streams? Implications for implementing the Water Framework Directive. Journal of Environmental Monitoring 11:344–352.

BAILEY, R. C., M. G. KENNEDY, M. Z. DERVISH, AND R. M. TAYLOR. 1998. Biological assessment of freshwater ecosystems using a reference condition approach: comparing predicted and actual benthic invertebrate communities in Yukon streams. Freshwater Biology 39: 765–774.

BAILEY, R. C., R. H. NORRIS, AND T. B. REYNOLDSON. 2004. Bioassessment of freshwater ecosystems: using the reference condition approach. Kluwer Academic Publishers, New York.

BAKER, L. A., A. T. HERLIHY, P. R. KAUFMANN, AND J. M. EILERS. 1991. Acidic lakes and streams in the United States: the role of acidic deposition. Science 252:1151–1154.

BENNION, H., J. FLUIN, AND G. L. SIMPSON. 2004. Assessing eutrophication and reference conditions for Scottish freshwater lochs using subfossil diatoms. Journal of Applied Ecology 41:124–138.

CARDOSO, A. C., A. SOLIMINI, G. PREMAZZI, L. CARVALHO, A. LYCHE, AND S. REKOLAINEN. 2007. Phosphorus reference concentrations in European lakes. Hydrobiologia 584:3–12.

CARVALHO, L., A. SOLIMINI, G. PHILLIPS, M. VAN DEN BERG, O.-P. PIETILÄINEN, A. LYCHE SOLHEIM, S. POIKANE, AND U. MISCHKE. 2008. Chlorophyll reference conditions for European lake types used for intercalibration of ecological status. Aquatic Ecology 42:203–211.

CHAVES, M. L., J. L. COSTA, P. CHAINHO, M. J. COSTA, AND N. PRAT. 2006. Selection and validation of reference sites in small river basins. Hydrobiologia 573:133–154.

COLLIER, K. J., A. HAIGH, AND J. KELLY. 2007. Coupling GIS and multivariate approaches to reference site selection for wadeable stream monitoring. Environmental Monitoring and Assessment 127:29–45.

DAVIES, S. P., AND S. K. JACKSON. 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. Freshwater Biology 16: 1251–1266.

HERLIHY, A. T., N. C. KAMMAN, J. C. SIFNEOS, D. CHARLES, M. ENACHE, AND R. J. STEVENSON. 2013. Developing nutrient criteria for lakes in the conterminous United States using multiple approaches. Freshwater Science 31:367–384.

HERLIHY, A. T., S. G. PAULSEN, J. VAN SICKLE, J. L. STODDARD, C. P. HAWKINS, AND L. L. YUAN. 2008. Striving for consistency in a national assessment: the challenges of applying a reference condition approach at a continental scale. Journal of the North American Benthological Society 27:860–877.

HUGHES, R. M., D. P. LARSEN, AND J. M. OMERNIK. 1986. Regional reference sites: a method for assessing stream potentials. Environmental Management 10:629–635.

JUSTUS, B. 2010. Water quality of least-impaired lakes in eastern and southern Arkansas. Environmental Monitoring and Assessment 168:363–383.

LANDERS, D. H., W. S. OVERTON, R. A. LINTHURST, AND D. F. BRAKKE. 1988. Eastern Lake Survey: regional estimates of lake chemistry. Environmental Science and Technology 22:128–135.

LEIRA, M., P. JORDAN, D. TAYLOR, C. DALTON, H. BENNION, N. ROSE, AND K. IRVINE. 2006. Assessing the ecological status of candidate reference lakes in Ireland using palaeolimnology. Journal of Applied Ecology 43:816–827.

OMERNIK, J. M. 1987. Ecoregions of the conterminous United States. Annals of the Association of American Geographers 77:118–125.

POIKÄNE, S., M. H. ALVES, C. ARGILLIER, M. VAN DEN BERG, F. BUZZI, E. HOEHN, C. DE HOYOS, I. KAROTTKI, C. LAPLACE-TREYTURE, A. L. SOLHEIM, J. ORTIZ-CASAS, I. OTT, G. PHILLIPS, A. PILKE, J. PÁDUA, S. REMEC-REKAR, U. RIEDMULLER, J. SCHAUMBURG, M. L. SERRANO, H. SOSZKA, D. TIERNEY, G. URBANIC, AND G. WOLFRAM. 2010. Defining chlorophyll-*a* reference conditions in European lakes. Environmental Management 45:1286–1298.

REYNOLDSON, T. B., R. H. NORRIS, V. H. RESH, K. E. DAY, AND D. M. ROSENBERG. 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. Journal of the North American Benthological Society 16:833–852.

REYNOLDSON, T. B., AND J. F. WRIGHT. 2000. The reference condition: problems and solutions. Pages 293–303 *in* J. F. Wright, D. W. Sutcliffe, and M. T. Furse (editors). Assessing the biological quality of fresh waters: RIVPACS and other techniques. Freshwater Biological Association, Ambleside, UK.

STODDARD, J. L., D. P. LARSEN, C. P. HAWKINS, R. K. JOHNSON, AND R. H. NORRIS. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. Ecological Applications 16:1267–1276.

SULLIVAN, T. J., D. L. KUGLER, M. J. SMALL, C. B. JOHNSON, D. H. LANDERS, B. J. ROSENBAUM, W. S. OVERTON, W. A. KRETSER, AND J. GALLAGHER. 1990. Variation in Adirondack, New York lakewater chemistry as a function of surface area. Journal of the American Water Resources Association 26:167–176.

USEPA (US ENVIRONMENTAL PROTECTION AGENCY). 2007. Survey of the Nation's Lakes. Field operations manual. EPA 841-B-07-004. Office of Water, US Environmental Protection Agency, Washington, DC.

USEPA (US ENVIRONMENTAL PROTECTION AGENCY). 2009. National Lakes Assessment: a collaborative survey of the Nation's lakes. EPA 841-R-09-001. Office of Water and Office of Research and Development, US Environmental Protection Agency, Washington, DC.

WAITE, I. R., A. T. HERLIHY, D. P. LARSEN, AND D. J. KLEMM. 2000. Comparing strengths of geographic and nongeographic classifications of stream macroinvertebrates in the mid-Atlantic Highlands, USA. Journal of the North American Benthological Society 19:429–441.

WHITTIER, T. R., AND S. G. PAULSEN. 1992. The surface water component of the Environmental Monitoring and Assessment Program (EMAP): an overview. Journal of Aquatic Ecosystem Health 1:119–126.

WHITTIER, T. R., S. G. PAULSEN, D. P. LARSEN, S. A. PETERSON, A. T. HERLIHY, AND P. R. KAUFMANN. 2002. Indicators of ecological stress and their extent in the population of northeastern lakes: a regional-scale assessment. BioScience 52:235–247.

YATES, A. G., AND R. C. BAILEY. 2010. Selecting objectively defined reference sites for stream bioassessment programs. Environmental Monitoring and Assessment 170:129–140.

Received: 30 June 2011
Accepted: 17 August 2012