# Calibration of the biological condition gradient in Minnesota streams: A quantitative expert-based decision system

**Article** *in* Freshwater Science · March 2017

DOI: 10.1086/691712



Jeroen Gerritsen
Tetra Tech


Raymond William Bouchard, Jr.
Minnesota Pollution Control Agency


Lei Zheng
Tetra Tech


Erik W. Leppo
Tetra Tech

# Calibration of the biological condition gradient in Minnesota streams: a quantitative expert-based decision system

**Jeroen Gerritsen[1,4], R. William Bouchard Jr[2,5], Lei Zheng[1,6], Erik W. Leppo[1,7], and Chris O. Yoder[3,8]**

[1]Tetra Tech, Inc., 10711 Red Run Blvd., Suite 105, Owings Mills, Maryland 21117 USA
[2]Minnesota Pollution Control Agency, Environmental Analysis and Outcomes Division, 520 Lafayette Road North, Saint Paul, Minnesota 55155-4194 USA
[3]Midwest Biodiversity Institute and Center for Applied Bioassessment and Biocriteria, P.O. Box 21561, Columbus, Ohio 43221-0561 USA

**Abstract:** The Biological Condition Gradient (BCG) is a conceptual model that describes changes in aquatic communities with increasing levels of anthropogenic stress. The gradient represented by the BCG has been divided into 6 levels of condition that biologists consider readily discernible in most areas of North America. We developed quantitative BCG models for 7 warm-water stream types in Minnesota for both fish and macroinvertebrates. Panels of aquatic biologists calibrated the general BCG model to Minnesota streams by assigning test samples (271 macroinvertebrate and 288 fish samples) to BCG Levels 1 to 6. From the panelists' descriptions of their criteria for assigning sites to levels, a set of quantitative operational rules was developed for performing the same task. We developed a decision model based on fuzzy-set theory to account for discontinuities and to identify when BCG assignments might be intermediate between adjacent levels. This model captures the consensus professional judgment of the panel and uses panel-derived rules. Decisions based on the quantitative model for macroinvertebrates exactly matched 77% of the panel decisions, 89% within ½ BCG level, and 100% within 1 BCG level. Decisions based on the quantitative fish model exactly matched 70% of the panel decisions, 86% within ½ BCG level, and 99% within 1 BCG level. The BCG provides a tool to interpret aquatic biological condition along a gradient of naturalness and is consistent across stream types and political boundaries. It includes documentation of baselines to prevent inadvertent shifting, and the BCG logic rules are transparent, a desirable property for communicating condition, management goals, and water-quality criteria.
**Key words:** Biological Condition Gradient, decision model, fuzzy logic, expert system, Minnesota, benthic macroinvertebrates, fish, water quality management, streams

In many nations, policies developed to protect and maintain water quality include the concepts of biological and ecological quality, which are assessed on the basis of the ecological structure and function of living aquatic communities. The US Clean Water Act of 1972 (CWA) has the long-term objective of restoration and protection of chemical, physical, and biological integrity (US Code title 33, §1251 [a]; USEPA 2011). In the European Union (EU), the Water Framework Directive (WFD) has the similar objective of restoration and maintenance of 'good' or better ecological quality (e.g., Hering et al. 2010, EU Commission 2015). Both the US Environmental Protection Agency (EPA) and the EU have made efforts to define what was meant by 'biological integrity' (USA) and 'high', 'good', 'fair', 'poor', and 'bad' condition (EU). In the USA, biological integrity has come to mean "The ability of an aquatic ecosystem to support and maintain a balanced, integrated and adaptive community of organisms having a species composition, diversity, and functional organization comparable to that of natural habitats within a region" (Frey 1977, Karr and Dudley 1981). In the EU, high ecological quality is defined as the ecological condition occurring under "no or very low human pressure" and is accepted as the reference condition (EU Commission 2015). Good through bad condition are

E-mail addresses: [4]jeroen.gerritsen@tetratech.com; [5]will.bouchard@state.mn.us; [6]lei.zheng@tetratech.com; [7]erik.leppo@tetratech.com; [8]cyoder@mwbinst.com

DOI: 10.1086/691712. Received 3 May 2016; Accepted 2 January 2017; Published online 14 March 2017.
Freshwater Science. 2017. 36(2):427–451.  

defined as successively increasing deviation from high or reference status (Hering et al. 2010). Both systems use natural condition with no or minimal human influence as a benchmark.

To meet the goals of the CWA and WFD, ecologically consistent interpretations of biological condition are needed to allow definition of thresholds of condition for assessment, restoration, and management. The definitions must be specific, well-defined, and must allow for waters of different natural quality and different desired uses. In the USA, the EPA developed a conceptual model—the Biological Condition Gradient (BCG)—that describes ecological changes from pristine to severely degraded that occur in flowing waters with increased anthropogenic degradation (Davies and Jackson 2006). The BCG was designed to provide a way to map different indicators on a common scale of biological condition to facilitate comparisons among programs and across jurisdictional boundaries. The original BCG is a conceptual, narrative model that describes how biological attributes of aquatic ecosystems change along a gradient of increasing anthropogenic stress (Fig. 1) and provides a framework for understanding current conditions relative to natural, undisturbed conditions (Davies and Jackson 2006, USEPA 2016).

US states, EU member states, and academics and environmental agencies worldwide have developed technical approaches and indexes to assess the biological condition of water bodies. In recent years, most approaches have been variations on the multimetric Index of Biotic Integrity (IBI; Karr et al. 1986, Whittier et al. 2007, Pont et al. 2009) or multivariate interpolations of reference-site species composition (River Invertebrate Prediction and Classification System; RIVPACS; e.g., Hawkins et al. 2000, Simpson and Norris 2000, Wright 2000). These indexes rely on empirical, present-day reference conditions quantified from existing reference sites to anchor their measurement systems. They require 'minimally disturbed' reference sites that are representative of biological integrity (Stoddard et al. 2006). However, in practice, most reference site data sets consist of 'least-disturbed' sites, which are the best remaining sites. The distinction between minimally disturbed and least-disturbed is important: minimally disturbed denotes fully natural biological conditions indistinguishable from pre-industrial or pre-European settlement, whereas least-disturbed denotes an upper quantile of contemporary conditions (Stoddard et al. 2006). Most indexes are built from a statistically adequate sample of least-disturbed (best available) reference sites, so that 1 or 2 minimally disturbed (near-pristine) sites in a reference data set may be treated as statistical outliers and may have little influence on index scoring. In the situation where no reference sites meet minimally disturbed criteria, the best score of this index would be similar to the moderately disturbed reference sites and could be substantially degraded from the natural condition. This situation is an example of the 'shifting baseline syndrome', such that the ideal reference or condition changes over generations as memory of previous baselines is lost (e.g., Pauly 1995, Dayton et al. 1998).

Part of the BCG process is to build a description of a fixed baseline based on either minimally disturbed conditions (Stoddard et al. 2006) or a fixed, agreed-upon point in time. The initial description is based on professional judgment, but as the BCG approach becomes accepted, the professional judgment should be replaced or enhanced

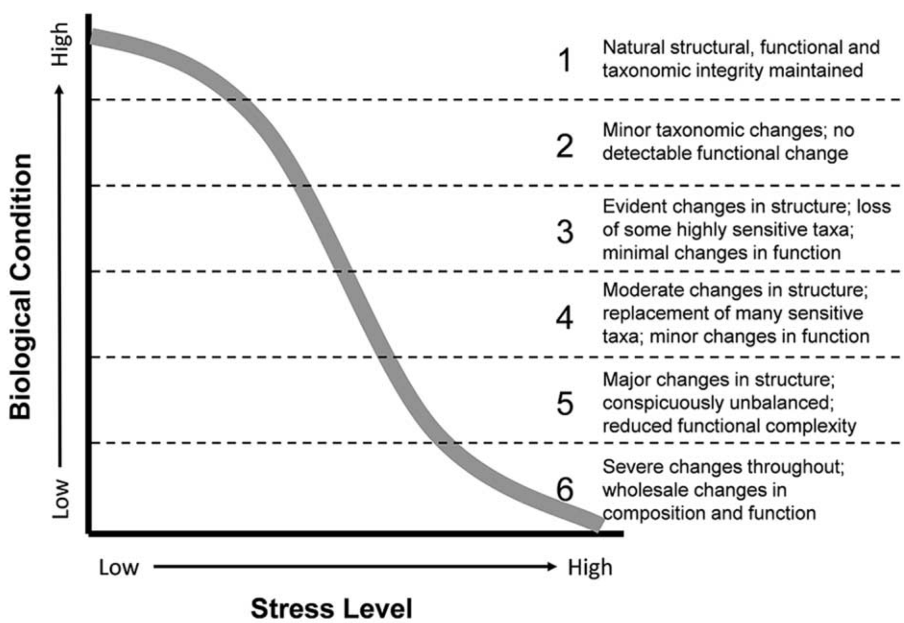


Figure 1. Graphic representation of the Biological Condition Gradient conceptual model (modified from Davies and Jackson 2006 with permission from John Wiley and Sons).

with documented information: historical descriptions, paleo investigations, museum records, and information from documented minimally disturbed sites. The description of minimally disturbed is necessarily incomplete, but its documentation is a defense against future inadvertent baseline shifts. Careful use of the BCG would identify a natural or historic baseline that could be used to guard against shifting baseline syndrome. For regions or situations where all information on natural baseline is irretrievably lost, the BCG could assist in identifying an 'Anthropocene baseline' for restoration and management (Kopf et al. 2015).

The quantitative BCG development was published by the USEPA (2016) based on case studies from the preceding decade. The methods have matured and experience gained has shown that a quantitative BCG has several desirable properties for use in water-quality management:

1. *Universal interpretive scale based on biologically meaningful changes* The original intent of the BCG was to create a scale with uniform interpretation across political and administrative jurisdictions (Davies and Jackson 2006). This intent was in response to the risk that use of different biological indexes and thresholds might result in contrary interpretations among states, wherein one state might call a cross-border stream impaired, but a neighboring state might not.
2. *Documented defense against shifting baselines* BCG values and thresholds are designed to defend against shifting baselines by including a description of undisturbed conditions. Any index or assessment method can include a documented baseline, but many indexes have been built empirically with data from 'least-stressed' reference sites (Stoddard et al. 2006). The BCG is independent of sometimes arbitrary percentiles of empirical reference populations. In the USA, management criteria consisting of the $50^{th}$, $25^{th}$, $10^{th}$, $5^{th}$, and $0^{th}$ percentiles of reference distributions have all been proposed by states and advocacy groups.
3. *A transparent decision system with stated rules* The quantitative BCG method consists of documented decision rules and, therefore, is transparent. Rules can be changed, but changes are conscious and deliberate and cannot result from additions or deletions in a database. The decision system provides a bridge between ecological science and value-based management. BCG levels can be adopted directly as management goals, restoration goals, or regulatory (protective) criteria.
4. *Flexibility* A quantitative BCG model can be used as a stand-alone assessment index or cross-walked to other existing indexes to provide ecological interpretation and identify management thresholds (Bouchard et al. 2016).

Here, we explain the calibration of a quantitative assessment model in the framework of the BCG. We use as an example the development of the model for warm-water streams and rivers of the state of Minnesota, USA, for benthic macroinvertebrate and fish assemblages (original report: Gerritsen et al. 2013).

## METHODS

### BCG primer

Biologists from across the USA developed the BCG conceptual model and agreed that a similar sequence of biological alterations occurs in streams in response to increasing stress, even in different geographic regions (Davies and Jackson 2006). The BCG is divided into 6 levels of biological condition along the stressor–response curve. Levels range from observable biological conditions found at no or low levels of stress (Level 1) to those found at the highest levels of stress (Level 6) (Fig. 1, Table 1). The 6 levels of the BCG are convergent with the 5 ecological status conditions defined in the EU WFD. The BCG levels were described in greater detail by Davies and Jackson (2006).

The BCG uses 10 attributes of aquatic ecosystems that change in response to increasing levels of stress along the gradient to describe the 6 levels (Table 2). The attributes include aspects of community structure, organism condition, ecosystem function, spatial and temporal attributes of stream size, and connectivity and are used as indicators of condition. The BCG was developed originally based on forested streams of eastern North America as examples (Davies and Jackson 2006), but the model has been applied to other regions and water bodies by calibrating it to local conditions on the basis of specific expertise and local data. Several US states, tribes, and territories are calibrating BCG-based indexes based on the first 7 attributes that characterize the biotic community, primarily tolerance to stressors, presence/absence of native and nonnative species, and organism condition (Table 2). BCG models have been developed for streams, lakes, estuaries, and coral reefs and biological assemblages including fish, benthic macroinvertebrates, and diatoms (summarized by USEPA 2016; Gerritsen and Leppo 2005, Stamp and Gerritsen 2012, Hausmann et al. 2016, Santavy et al., in press).

### Approach

Our approach for BCG model development is based on professional judgment and development of consensus. Professional consensus has a long pedigree in the medical field, including the National Institutes of Health (NIH) Consensus Development Conferences to recommend best practices for diagnosis and treatment of diseases (http://consensus.nih.gov/). The NIH consensus meetings were a "hybrid of . . . judicial decision-making, scientific conferences and the town hall meeting" (Nair et al. 2011). Other researchers, institutes, and countries also develop medical consensus statements using NIH methods (Nair et al. 2011).

Table 1. Descriptions of Biological Condition Gradient levels (modified from Davies and Jackson 2006).

| BCG level | Description |
|---|---|
| Level 1: Natural or native condition | Native structural, functional, and taxonomic integrity is preserved; ecosystem function is preserved within the range of natural variability. Level 1 represents biological conditions as they existed (or may still exist) in the absence of measurable effects of stressors. |
| Level 2: Minimal changes in structure of the biotic community and minimal changes in ecosystem function | Virtually all native taxa are maintained with some changes in biomass or abundance; ecosystem functions are fully maintained within the range of natural variability. Level 2 represents the earliest changes in densities, species composition, and biomass that occur as a result of slight elevation in stressors (such as increased temperature regime or nutrient enrichment). |
| Level 3: Evident changes in structure of the biotic community and minimal changes in ecosystem function | Evident changes in structure caused by loss of some highly sensitive native taxa; shifts in relative abundance of taxa but sensitive-to-ubiquitous taxa are common and abundant; ecosystem functions are fully maintained through redundant attributes of the system. Level 3 represents readily observable changes that, e.g., can occur in response to organic enrichment or increased temperature. |
| Level 4: Moderate changes in structure of the biotic community with minimal changes in ecosystem function | Moderate changes in structure caused by replacement of some intermediate-sensitive taxa by more tolerant taxa, but reproducing populations of some sensitive taxa are maintained; overall balanced distribution of all expected major groups; ecosystem functions largely maintained through redundant attributes. |
| Level 5: Major changes in structure of the biotic community and moderate changes in ecosystem function. | Sensitive taxa are markedly diminished; conspicuously unbalanced distribution of major groups from those expected; organism condition shows signs of physiological stress; ecosystem function shows reduced complexity and redundancy; increased build-up or export of unused materials. Changes in ecosystem function (as indicated by marked changes in foodweb structure and guilds) are critical in distinguishing between Levels 4 and 5. |
| Level 6: Severe changes in structure of the biotic community and major loss of ecosystem function | Extreme changes in structure; wholesale changes in taxonomic composition; extreme alterations from normal densities and distributions; organism condition is often poor; ecosystem functions are severely altered. Level 6 systems are taxonomically depauperate (low diversity or reduced number of organisms) compared to the other levels. |

Experts define BCG levels in the context of the conceptual model (Davies and Jackson 2006). They determine the attributes and the changes in those attributes that characterize distinct BCG levels and signal shifts to a different level (Tables 1, 2). The BCG consensus approach asks the experts to make judgments on the biological significance of changes in the attributes. Thus, a fundamental assumption of this approach is that consensus professional judgment is the best current estimate of biological condition. The outcome of the process is a multiple-attribute decision model that mimics the consensus decisions based on a set of quantitative rules. The logic train of the decision model and the experts' documented reasoning create a transparent decision system for review, modification, and water-quality management.

Index calibration begins with the assembly and analysis of biological monitoring data and identification of stress–response relationships for individual taxa. During one or more calibration workshops, experts familiar with local conditions and biota use the data to develop narrative decision rules for assigning sites to a BCG level. Panelists assign relevant taxa to BCG attributes (Table 2). Next, they examine biological data from selected sites, describe the native aquatic assemblages under natural conditions, and assign the samples to Levels 1 to 6 of the BCG. The intent is to achieve consensus and to identify rules that experts use to make their assignments. Experts' opinions are elicited and documented to assist in quantitative rule development.

Over the long term, reconvening the same group of experts for every new sample is impractical. Thus, use of a quantitative BCG in routine monitoring and assessment requires a way to automate the consensus expert judgment. The decision criteria are codified into a quantitative decision model, which is a transparent, formal, and testable method for documenting and validating expert knowledge.

For over a decade, the Minnesota Pollution Control Agency (MPCA) has been using fish and benthic macroinvertebrate assemblage data to assess water resource quality. Until recently, biological indexes in Minnesota were developed for individual drainage basins (e.g., Niemela et al. 1999). The MPCA used data from 2285 fish and 1502 macroinvertebrate samples to develop statewide fish and macroinvertebrate IBIs following the approach published by Whittier et al. (2007). Descriptions of these IBIs can be found in MPCA (2014b, c). The BCG calibration we describe here relies heavily on the knowledge and experience gained from Minnesota's IBI developments, and addresses MPCA's ob-

Table 2. Attributes used to characterize the Biological Condition Gradient (BCG) (modified from Davies and Jackson 2006).

| Attribute | Description |
|---|---|
| | Attributes I–V: Native structure and composition |
| I. Historically documented, sensitive, long-lived, or regionally endemic taxa | Taxa known to have been supported according to historical, museum, or archeological records, or taxa with restricted distribution (occurring only in a locale as opposed to a region), often because of unique life-history requirements (e.g., sturgeon, American Eel, pupfish, unionid mussel species) |
| II. Highly sensitive (typically uncommon) taxa | Taxa that are highly sensitive to pollution or anthropogenic disturbance; tend to occur in low numbers, and many are specialists for habitats and food type; the first to disappear with disturbance or pollution (e.g., most stoneflies, Brook Trout [in the eastern USA], Brook Lamprey) |
| III. Intermediate sensitive and common taxa | Common taxa that are ubiquitous and abundant in relatively undisturbed conditions but are sensitive to anthropogenic disturbance/pollution; have a broader range of tolerance than attribute II taxa and can be found at reduced density and richness in moderately disturbed sites (e.g., many mayflies, many darter fish species) |
| IV. Taxa of intermediate tolerance | Ubiquitous and common taxa that can be found under almost any conditions, from undisturbed to highly stressed sites; broadly tolerant but often decline under extreme conditions (e.g., filter-feeding caddisflies, many midges, many minnow species) |
| V. Highly tolerant taxa | Taxa that typically are uncommon and of low abundance in undisturbed conditions but increase in abundance in disturbed sites; opportunistic species able to exploit resources in disturbed sites; the last survivors (e.g., tubificid worms, Black Bullhead) |
| VI. Nonnative or intentionally introduced species | Any species not native to the ecosystem (e.g., Asiatic clam, Zebra Mussel, carp, European Brown Trout); in addition, many fish native to one part of North America introduced elsewhere |
| VII. Organism condition | Anomalies of the organisms; indicators of individual health (e.g., deformities, lesions, tumors) |
| VIII. Ecosystem function | Processes performed by ecosystems, including primary and secondary production, respiration, nutrient cycling, decomposition, their proportion/dominance, and what components of the system carry the dominant functions (e.g., shift of lakes and estuaries to phytoplankton production and microbial decomposition under disturbance and eutrophication) |
| IX. Spatial and temporal extent of detrimental effects | The spatial and temporal extent of cumulative adverse effects of stressors (e.g., groundwater pumping in Kansas led to change in fish composition from fluvial-dependent to sunfish) |
| X. Ecosystem connectivity | Access or linkage (in space/time) to materials, locations, and conditions required for maintenance of interacting populations of aquatic life; the opposite of fragmentation (e.g., levees restrict connections between flowing water and floodplain nutrient sinks, dams impede fish migration, spawning) |

jective to develop statewide biological criteria for streams within Minnesota.

Aquatic biologists familiar with Minnesota streams met as a work group to develop the ecological attributes and rules for assigning sites to levels. Their expertise included aquatic ecology, benthic macroinvertebrate sampling and monitoring, water quality, and fisheries biology. We summarize here the results of BCG calibration for warm-water streams in Minnesota (Gerritsen et al. 2013). A 2$^{nd}$ multi-state and multi-tribal effort to develop a BCG calibration for cold water streams of the Upper Midwest was reported by Gerritsen and Stamp (2013).

## Data

When the models were developed, the MPCA had collected >3800 fish and >2800 macroinvertebrate samples from warm-water streams (1996–2011). Minnesota's biological assessment program was assessed in 2015 (USEPA 2013) and was deemed sufficient to support development and implementation of biological monitoring tools (MBI 2015).

A fish sampling reach is defined as 35× mean stream width. This length is sufficient to capture a representative and repeatable sample of the fish assemblage in a stream segment (Lyons 1992, MPCA 2014d). Sampling is conducted during daylight hours in the summer index period (mid-June–mid-September). Streams are sampled during or near base flow because floods or droughts can affect fish assemblage structure and sampling efficiency. All habitat types within the sampling reach are sampled in approximate proportion to their occurrence to capture fish ≥25 mm in total length. Four electrofishing methods are used: backpack electrofisher in small headwater streams; towed stream electrofisher in larger wadeable streams; mini-boom electrofisher (2-person jon boat) in small, nonwadeable streams;

and a boat-mounted boom electrofisher in large streams and rivers. For detailed fish sampling methods see MPCA (MPCA 2014d). Fish sampling is repeated at 10% of the sample reaches during the index period to estimate measurement error.

A multihabitat method is used to obtain a representative sample of the macroinvertebrate assemblage of a reach. Habitats sampled include hard bottom (riffle/cobble/boulder), aquatic macrophytes (submerged/emergent vegetation), undercut banks (undercut banks/overhanging vegetation), snags (snags/rootwads), and leaf packs. Twenty D-frame dipnet (500-μm mesh) sweeps are divided equally among the dominant, productive habitats present in the reach. Each sweep covers ~0.09 $m^2$ of substrate for a total area sampled of ~1.8 $m^2$. Collections are randomly subsampled to a target subsample of 300 individuals and identified to genus. Macroinvertebrate collection standard operating procedures (SOPs) were described fully by the MPCA (MPCA 2014e). Macroinvertebrate sampling is repeated at 10% of the sample reaches on the same day to estimate measurement error.

Measurement error (sample variability) was not estimated as part of this project, but Minnesota's sampling and analysis methods are comparable to those used by EPA in national aquatic surveys (e.g., Stoddard et al. 2008). Other studies of similar methods have shown variability of indexes to be low and consistent for repeated samples within and among years (e.g., Hose et al. 2004, Barbour and Gerritsen 2006, Huttunen et al. 2012).

## Classification

Classification of aquatic habitats is necessary to account for natural variability so that the experts can place a stream in context of its setting. Panelists involved in some early attempts to develop a quantitative BCG struggled in the absence of a classification scheme understood by the panel and appropriate for the data set (USEPA 2016). Most panels have preferred a primarily typological classification (e.g., ecoregions), but continuous classifiers, such as catchment area, stream gradient, and elevation, have been used successfully.

The MPCA developed a classification system for natural stream communities to support the development of typological IBI models (MPCA 2014b, c). The stream types were based on distributions of species among classification variables that are not influenced by anthropogenic effects. The classification system for warm-water streams was developed with the same data set used to develop the IBIs and consisted >2200 fish and 1500 macroinvertebrate samples collected from 1996 through 2008. Biological communities and predictive variables were identified with the aid of several tools including: hierarchical cluster analysis, nonmetric multidimensional scaling, and Mean Similarity Analysis (Van Sickle 1998, Van Sickle and Hughes 2000). This process resulted in 7 warm-water stream types each for the fish and the benthic macroinvertebrate communities based on: 1) ecoregion, 2) sampling method, 3) drainage area, and 4) stream gradient (Table 3). Fish and macroinvertebrate stream types follow a similar regional pattern, but they do not match. For example, invertebrate high-gradient and low-gradient habitats may occur in both wadeable and headwater streams as defined for fish sampling. Geographic delineations included northern or southern Minnesota and forest or prairie. The remaining classes were defined by sampling method (e.g., high-gradient vs low-gradient for macroinvertebrates).

## Preliminary analysis: stress–response and BCG attributes

The MPCA developed a disturbance index called the Human Disturbance Score (HDS) based on the degree of human activity in the upstream watershed and at the reach level for biological monitoring sites (Bouchard et al. 2016, MPCA 2016). The HDS includes 8 primary metrics, which consist of measures of watershed land use, stream alteration, riparian condition, and known permitted discharges.

Table 3. Final Minnesota Pollution Control Agency (MPCA) classifications of warm-water stream types for fish and macroinvertebrates, and number of samples with valid data in each. The 2 river classes correspond between fish and macroinvertebrates, but the wadeable stream classes do not correspond.

| Fish stream type | | Macroinvertebrate stream type | |
|---|---|---|---|
| Name | *N* | Name | *N* |
| Northern rivers | 358 | Northern forest rivers | 125 |
| Southern rivers | 525 | Prairie and southern forest rivers | 155 |
| Northern streams | 523 | Northern forest streams, high-gradient | 271 |
| Northern headwaters | 706 | Northern forest streams, low-gradient | 425 |
| Southern streams | 665 | Southern streams, high-gradient | 445 |
| Southern headwaters | 638 | Southern forest streams, low-gradient | 396 |
| Low-gradient streams | 313 | Prairie streams, low-gradient | 617 |

HDS scores can range from 1 (heavily altered watersheds) to 81 (nearly pristine watersheds).

### Stress-response models

BCG composition attributes II through V (Table 2) are familiar tolerance designations (e.g., Merritt et al. 2008) applied in many IBI and multimetric indexes. Published tolerance values are often 'received wisdom' originally estimated from different regions (Carlisle et al. 2007), so we augmented the published values with analysis of the MPCA data to estimate tolerances from the local data. We used general linear models (GLMs) to estimate the probability of observing a particular taxon across the HDS score. The optimum of the model (maximum probability) yielded the tolerance value. We plotted the capture probabilities over the range of the disturbance gradient (Figs 2–5).

### Assign taxa to attributes

Assignments of taxa to attributes relied on a combination of the empirical data analysis (Figs 2, 3A, B, 4A, B, 5A, B), published values, and professional experience of the expert panels (Tables 4, 5). HDS is not a perfect measurement of stressors in a stream reach because it is a general predictor of disturbance. It provided an a priori general stressor gradient that is associated with taxon abundance and probability of occurrence to assist the panel in assigning the BCG attributes. The use of empirical data, published tolerances, ecological theory, and professional experience minimizes the effect of noise in the HDS during BCG development.

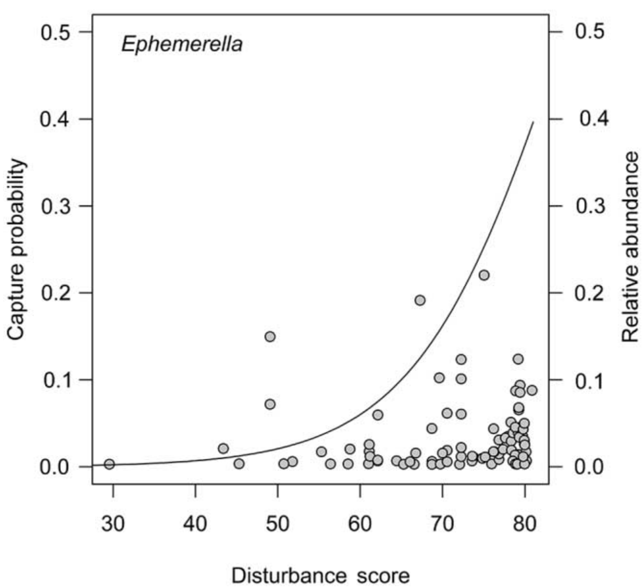


Figure 2. Disturbance score and *Ephemerella* occurrence in stream samples. Circles show observations and relative abundance of *Ephemerella* (right axis); curve shows probability of occurrence (left axis; maximum likelihood). *Ephemerella* was assigned to Biological Condition Gradient (BCG) attribute II (highly sensitive taxa), as shown by its high abundance and high probability of occurrence in minimally disturbed sites (disturbance score 81). See Table 2 for BCG attributes.

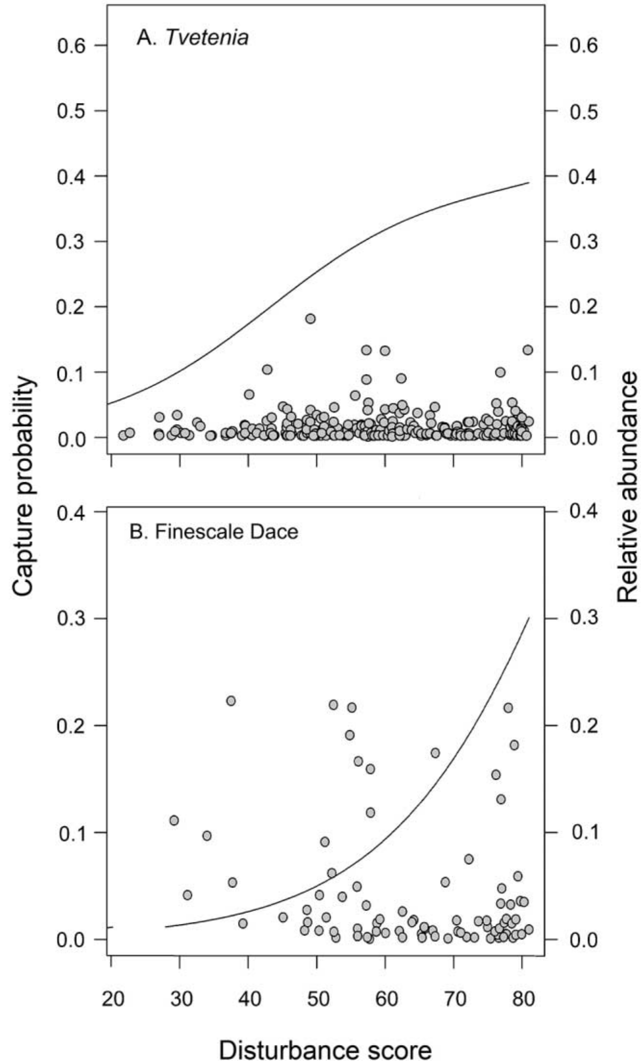


Figure 3. Examples of Biological Condition Gradient (BCG) attribute III taxa *Tvetenia* (A) and Finescale Dace (B). These species occur throughout the disturbance gradient, but with higher probability in better sites. Final attribute assignment was based on these plots and on professional judgment of the panel. See Table 2 for BCG attributes.

For taxa with a sufficient sample, the capture probabilities and, to a lesser extent, the observed abundances followed the expectations given by the attribute descriptions (Table 2, Figs 2, 3A, B, 4A, B, 5A, B). In cases of disagreement, the panel relied on consensus professional opinion unless contradicted by an overwhelming response in the data analysis.

The fish panel identified 2 additional subclasses of the attributes 'tolerant species' and 'nonnative species'. They identified highly tolerant native species (attribute Va) as the last survivors in a degraded stream and divided the

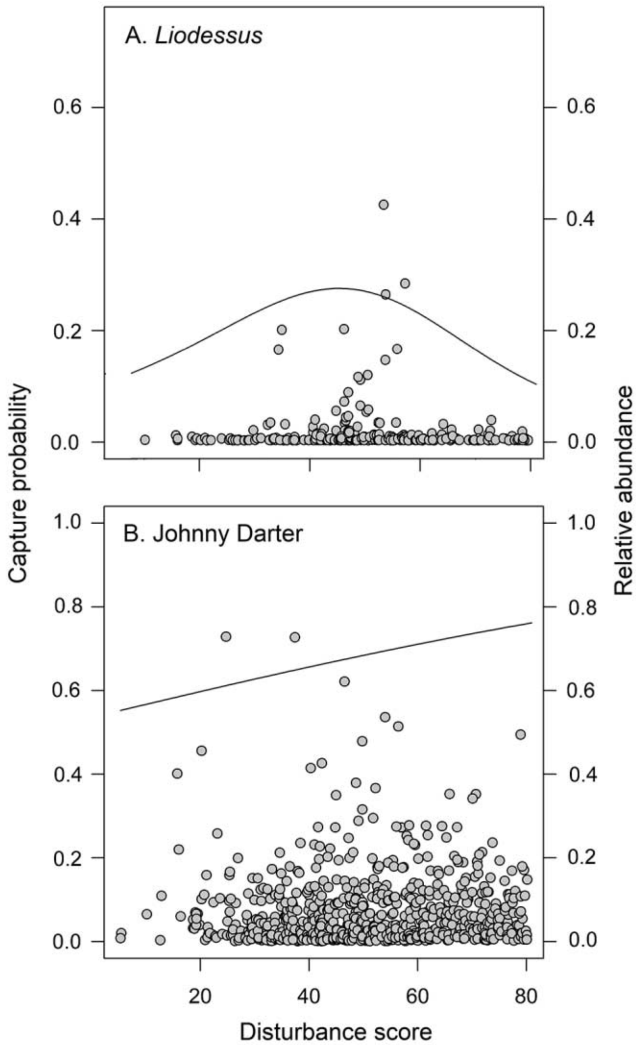


Figure 4. Examples of intermediate tolerant, Biological Condition Gradient (BCG) attribute IV taxa *Liodessus* (A) and Johnny Darter (B). These species occur throughout the disturbance gradient with roughly equal probability throughout or with a peak in the middle of the disturbance range. See Table 2 for BCG attributes.

nonnative group into sensitive nonnative species (attribute VI, e.g., nonnative salmonids) and tolerant nonnatives (attribute VIa; e.g., Common Carp, Ruffe; Table 5).

## Assign sites to BCG levels

The panels examined data from selected monitoring sites and assigned the sites to levels of the BCG based on the taxa present in the sample and the generic descriptions of BCG levels (Table 1). The data included lists of taxa and abundances, BCG attribute groups assigned to the taxa, summary metrics, and limited site information, such as stream type and ecoregion, sampling method, and substrate. Stream location, water quality, and MPCA's disturbance score were not revealed to panel members because doing so might have biased assignments. Panel members discussed the species composition, what they expected to see for each level of the BCG, and then assigned samples to BCG levels. The work groups examined macroinvertebrate data from 271 samples (7 stream types), and fish data from 288 samples (7 stream types).

## Quantitative description

In the discussions of BCG assignments, facilitators elicited panelist's reasoning for their decision; e.g., "I expect to see more stonefly taxa in a BCG Level-2 site." The reasoning formed the basis to formalize the expert knowledge by codifying level descriptions into a set of rules (e.g., Droesen 1996).

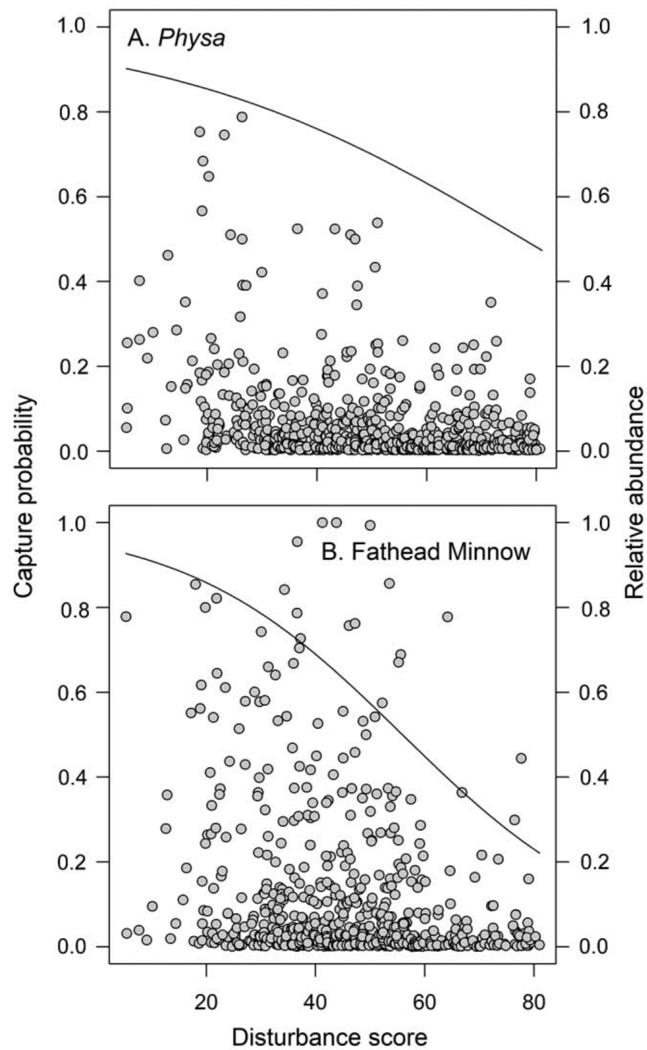


Figure 5. Examples of tolerant taxa, *Physa* (Biological Condition Gradient [BCG] attribute V; tolerant) (A) and Fathead Minnow (BCG attribute Va; highly tolerant) (B). These species occur throughout the disturbance gradient, but with higher probability of occurrence and higher abundances in more stressed sites. See Table 2 for BCG attributes.

Table 4. Examples of macroinvertebrate taxa by Biological Condition Gradient (BCG) attribute group. Assignment to attribute varied between habitats (high-gradient and low-gradient) for some taxa, so number of taxa represents the range of the number of genera assigned to the attribute group among stream types.

| BCG attribute | Number of taxa | Example taxa |
|---|---|---|
| I Endemic, rare | 1 | *Goera* |
| II Highly Sensitive | 29–41 | *Stempellina*, *Heleniella*, *Ephemerella*, *Paraleuctra*, *Ophiogomphus*, *Parapsyche*, *Diplectron*,, *Lepidostoma*, *Dolophilodes*, *Rhyacophila* |
| III Intermediate Sensitive | 107–148 | *Diamesa*, *Tvetenia*, *Hexatoma*, *Plauditus*, *Parapoynx*, *Isoperla*, *Boyeria*, *Amphinemura*, *Pycnopsyche*, *Brachycentrus*, *Limnephilus* |
| IV Intermediate Tolerant | 201–231 | Dytiscidae, Ceratopogonidae, *Polypedilum*, *Limonia*, *Perlesta*, *Heptagenia*, *Libellula*, *Hydropsyche*, *Sphaerium*, *Planorbella* |
| V Tolerant | 25–41 | Erpobdellidae, *Cricotopus*, *Pseudocloeon*, Corixidae, *Enallagma*, *Caecidotea*, Physidae |
| VI Nonnative | 1 | *Corbicula* |
| x Unassigned | 20 | Family identifications or unusual taxa; *Chaoborus*, *Zavrelia*, *Didymops*, Nemata |

Rule development required discussion and documentation of BCG-level assignment decisions and the reasoning behind the decisions. During this discussion, we recorded: 1) each participant's decision ('vote') for the site; 2) the critical or most important information for the decision, e.g., the number of taxa of a certain attribute, the abundance of an attribute, the presence of indicator taxa; and 3) confounding or conflicting information and how the conflict was resolved for the eventual decision.

After initial site assignment and rule development, we estimated descriptive statistics of the attributes and other biological indicators for each BCG level determined by the panel. These descriptions assisted in review of the rules and their iteration for testing and refinement. The first 2 panel sessions were in-person, 3-d workshops, and subsequent panel sessions were by webinar. The initial panel decisions comprised a preliminary set of decision rules. We quantified the rules in Excel® (versions 2003–2013; Microsoft, Redmond, Washington) workbooks, and calculated BCG level assignments for each sample. We evaluated model performance by comparing model-assigned BCG levels to the panel assignments. Following the initial development phase, the panel tested the draft rules with new data to ensure that new sites were assessed in the same way. Any remaining ambiguities and inconsistencies from the first iterations were resolved.

Table 5. Examples of fish taxa by Biological Condition Gradient (BCG) attribute group. Assignment to attribute varied among stream types for some species, so number of taxa represents the range of the number of species assigned to the attribute group among 7 stream types.

| BCG attribute | Number species | Example species |
|---|---|---|
| I Endemic, rare | 1–9 | Blue Sucker, Crystal Darter, Gilt Darter, Greater Redhorse, Lake Sturgeon, Pugnose Shiner, River Redhorse, Shovelnose Sturgeon, Topeka Shiner |
| II Highly sensitive | 6–17 | American Brook Lamprey, Blackchin Shiner, Brook Trout, Southern Brook Lamprey, Western Sand Darter |
| III Intermediate sensitive | 15–35 | Blacknose Shiner, Burbot, Golden Redhorse, Hornyhead Chub, Shorthead Redhorse, Smallmouth Bass |
| IV Intermediate tolerant | 26–43 | Common Shiner, Gizzard Shad, Johnny Darter, Northern Pike, Spotfin Shiner, White Sucker[a] |
| V Tolerant | 5–18 | Creek Chub, Brassy Minnow, Brook Stickleback, Central Stoneroller, Sand Shiner |
| Va Highly tolerant | 7–8 | Bigmouth Shiner, Bluntnose Minnow, Fathead Minnow, Green Sunfish |
| VI Sensitive nonnative | 3 | Brown Trout, Rainbow Trout, Chinook Salmon |
| VIa Tolerant nonnative | 4 | Common Carp, Goldfish, Ruffe, Threespine Stickleback |
| x unassigned | 2 | Unidentified fish, hybrids |

[a] White Sucker is identified tolerant (attribute V) in wadeable streams only.

## BCG inference models

The decision models calculated BCG levels directly from the quantified rules by applying fuzzy logic (Zadeh 1965, 2008). Instead of a statistical prediction of expert judgment, this approach directly and transparently converts the expert consensus to automated site assessment. Fuzzy logic is "a precise logic of imprecision and approximate reasoning" (Zadeh 2008). It is directly applicable to environmental assessment and has been used extensively in engineering and environmental applications worldwide (e.g., Castella and Speight 1996, Ibelings et al. 2003, Demicco and Klir 2004, Cheung et al. 2005, Joss et al. 2008).

Fuzzy logic and set theory allows degrees of truth, in contrast to binary truth in classical logic and set theory. For example, one can compare how classical set theory and fuzzy-set theory treat classification of sediment, where sand is defined as particles ≤2.0 mm diameter and gravel is >2.0 mm (Klir 2004). In classical 'crisp' set theory, a particle with diameter = 2.00 mm is classified as sand, and one with diameter = 2.01 mm is classified as gravel. In fuzzy-set theory, both particles have nearly equal membership in both classes (Klir 2004). Measurement error as small as 0.005 mm greatly increases the uncertainty of classification in classical set theory, but in fuzzy-set theory a particle near the boundary would have nearly equal membership in both sets (sand and gravel). Thus, fuzzy sets retain the understanding and knowledge of measurements close to a set boundary, which is lost in classical sets. For further explanation of fuzzy logic, see Klir (2004) or any online tutorial.

To develop the fuzzy inference model, each linguistic variable (e.g., high taxon richness) is defined quantitatively as a fuzzy set (e.g., Klir 2004). A fuzzy set has a membership function in the range of 0 to 1 that determines whether an object is in the set or not in the set. Example membership functions of different sets of taxon richness are shown in Fig. 6A, B. We used piecewise linear functions (i.e., functions consisting of line segments) to assign membership values. If the number of taxa is less than or equal to the lower threshold it has membership of 0, if the number of taxa is greater than or equal to an upper threshold it has membership of 1, and if the number of taxa is between the thresholds, the membership is assigned using a linear interpolation between the lower and upper thresholds. For example, a sample with 30 total taxa would have a membership of ~0.5 in the set 'Moderate number of taxa' and a membership of 0.5 in the set 'High number of taxa' (Fig. 6A).

Assigning membership on the basis of fuzzy-set theory is different from doing so on the basis of classical set theory. Suppose 2 rules determine whether a water body is BCG Level 3: 1) the number of total taxa is high and 2) the number of sensitive taxa is moderate or higher (shaded areas in Fig. 6A, B). If both rules must be true, they are combined with the Boolean AND operator. In fuzzy-set theory, the Boolean AND operator is equivalent to the

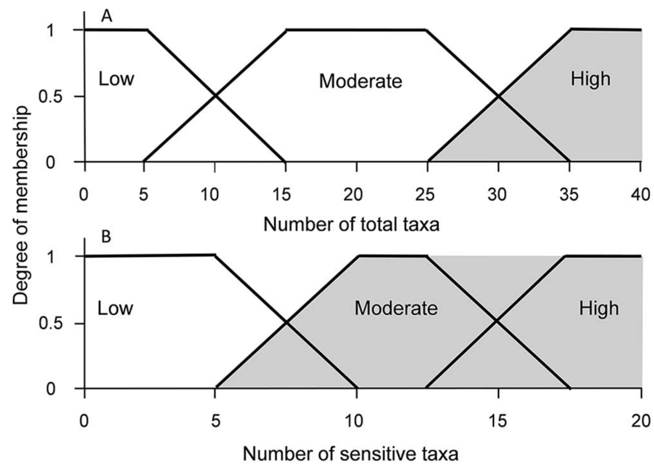


Figure 6. Fuzzy-set membership functions assigning linguistic values to defined ranges for total taxa (A) and sensitive taxa (B). Shaded regions correspond to example rules for Biological Condition Gradient Level 3: "number of total taxa is high" and "number of sensitive taxa is moderate or greater."

minimum membership given by each rule: Level 3 = MIN (total taxa is high, sensitive taxa is moderate or higher). For 32 total taxa and 7 sensitive taxa, fuzzy membership in total taxa is high = 0.6 (Fig. 6A), and fuzzy membership in sensitive taxa is low-moderate to moderate = 0.4 (Fig. 6B). Membership of BCG Level 3 is then 0.4, indicating that the site is "somewhat like Level 3 sites, but not overwhelmingly"; i.e., it is borderline. In the fuzzy-set case, a single additional sensitive taxon raises the membership in BCG Level 3 from 0.4 to 0.6, indicating it is somewhat more like Level 3, but still borderline. In classical set theory, the boundaries of the categories in Fig. 6A, B would be vertical lines. A sample with 7 sensitive taxa would be deemed not in Level 3, but a sample with 8 sensitive taxa would be deemed in Level 3.

If the 2 rules are combined with an OR operator, then either can be true for a site to meet BCG Level 3. In words, we would say, "BCG Level is 3 if total taxa are high OR sensitive taxa are moderate or higher." Classical set theory now yields a value of 'true' if total taxa = 32 and sensitive taxa = 7 (total taxa > 32, therefore, it is true). Fuzzy-set theory yields a membership of 0.6 (maximum of 0.4 and 0.6). In practice, the OR operator is specified only occasionally, when the panel wishes to set up alternative criteria for a certain decision.

In the decision model, rules work as a logical cascade from BCG Level 1 to Level 6. A sample is first tested against the BCG Level 1 rules. If a required rule fails, then the level fails, and the assessment moves down to BCG Level 2, and so on. Required rules are combined with AND operators (i.e., all must be true), and alternate rules are combined with OR operators. Membership in any BCG Level ranges from 0 to 1, and the model requires all membership values to sum to 1. The highest membership is taken as the nominal level, although memberships within 0.2 of each other are considered ties. For example, if the membership of BCG Level 2 is 0.5 and Level 3 is 0.4, then the site is considered to be intermediate between Levels 2 and 3. The output of the model is the nominal BCG level and its membership value and the 2nd (runner-up) BCG level and its membership value.

Because MPCA intended to use the BCG to develop meaningful thresholds for its IBI indexes, the BCG scores were compared to IBI scores from all available biological visits. This analysis consisted of examining box plots and outliers (e.g., sites with high IBI scores, but BCG scores indicating an altered community). The intent of this analysis was not to identify individual visits and bring them in alignment with BCG expectations, but to identify groups of similar communities that were not part of the calibration or test data sets and might require changes to both BCG and IBI models. This effort was parsimonious because too much modification to the models could lead to over-fitting or altering the model from the intent of the panel.

## RESULTS

### Stress-response relationships and BCG taxa attributes

We examined stress-response scatterplots and estimated maximum likelihood models for taxon occurrence for all taxa with >20 occurrences in the data set (Figs 2, 3A, B, 4A, B, 5A, B, S1, S2). HDS scores were not evenly distributed with relatively few sites with scores <40 (highly altered). An apparent reduction in point density at low-disturbance scores reflects the fact that few sites in the database had such low scores and not necessarily the response of the taxa. The capture probability curve takes the distribution of disturbance scores into account and shows which taxa are tolerant or thrive under disturbed conditions (Figs 2, 3A, B, 4A, B, 5A, B, S1, S2).

Scatterplots that combined abundances of individual taxa on the disturbance gradient with the maximum likelihood models were deemed to be the most useful for identifying attribute groups (Tables 4, 5, Figs 2, 3A, B, 4A, B, 5A, B). Fish species were assigned to attributes separately for each of the 7 fish stream types, and macroinvertebrates were assigned separately to 2 groups: high-gradient and low-gradient streams. Only a few taxa differed in assigned attribute among stream types.

Fish experts identified 2 additional subattributes related to highly tolerant taxa (Table 5). An additional very tolerant classification was created (attribute Va). Separation of the highly tolerant attribute Va fish from the merely tolerant attribute V fish was based on the collective professional experience and judgment of the fish panel. The nonnative fish taxa attribute (VI) was similarly divided into sensitive nonnative salmonids (attribute VI; e.g., Brown Trout and Rainbow Trout) and highly tolerant nonsalmonid, nonnative species (attribute VIa; e.g., Ruffe, Sea Lamprey, Common Carp).

In total, 133 fish taxa and 516 macroinvertebrate taxa were assigned to BCG taxonomic attributes (Tables S1, S2). An additional 53 fish species occurred in MDNR's species list, but were absent from the stream data set and were left unclassified, and 10 fish taxa in the data were left unclassified (family- or genus-level identifications or hybrids considered uninformative). Twenty invertebrate taxa were left unassigned because participants thought information on the taxa was insufficient, or they were relatively unusual in the data set.

### Site assignments to BCG levels

The panel was able to reach a majority opinion on the BCG level assignments for all sites reviewed. Some sites required discussion and resolution of disagreement on which of 2 adjacent BCG levels to assign the site. These sites were considered intermediate, with characteristics of both adjacent BCG levels.

The panels were able to distinguish 6 BCG levels (BCG Levels 1–6), but sites that fit Levels 1 (nearly pristine) and 6 (extreme degradation) were rare. The fish panel identified 9 BCG Level 1 sites, but the macroinvertebrate panel identified none. In general, macroinvertebrate experts felt that BCG Level 1 and Level 2 sites were not distinguishable based on macroinvertebrate data only, in part because rare and endemic taxa are poorly identified, their historic distributions are poorly known, and macroinvertebrate sampling methods are inefficient at finding rare and endemic species. Further examination may be necessary to decide whether any sites meet criteria for minimally disturbed (Stoddard et al. 2006). The macroinvertebrate panel identified 9 and the fish panel identified 8 BCG Level 6 samples.

### Attributes and BCG levels

We derived metrics (e.g., taxon richness, % taxa, % individuals, dominance) based on BCG attributes and taxonomic groupings (see examples in Figs 7A–F, 8A–F, 9A–D, 10A–E). These box plots were used to help with the selection of metrics for initial model development and for panel review of metrics and rules during subsequent iterations. We developed the BCG using only taxonomic information (attributes I–VI; USEPA 2016) because MPCA's monitoring program does not require collection of information on the other attributes. If available, information from attributes VII–X could be incorporated into the BCG models to improve their performance.

### BCG rule development

Panelists followed the descriptions of the BCG levels (Table 1) and gave their reasoning during the deliberations for assigning sites to levels. Rules and reasoning of the panel, whether quantitative or qualitative, were compared to data summaries of the panel decisions (Figs 7A–F,

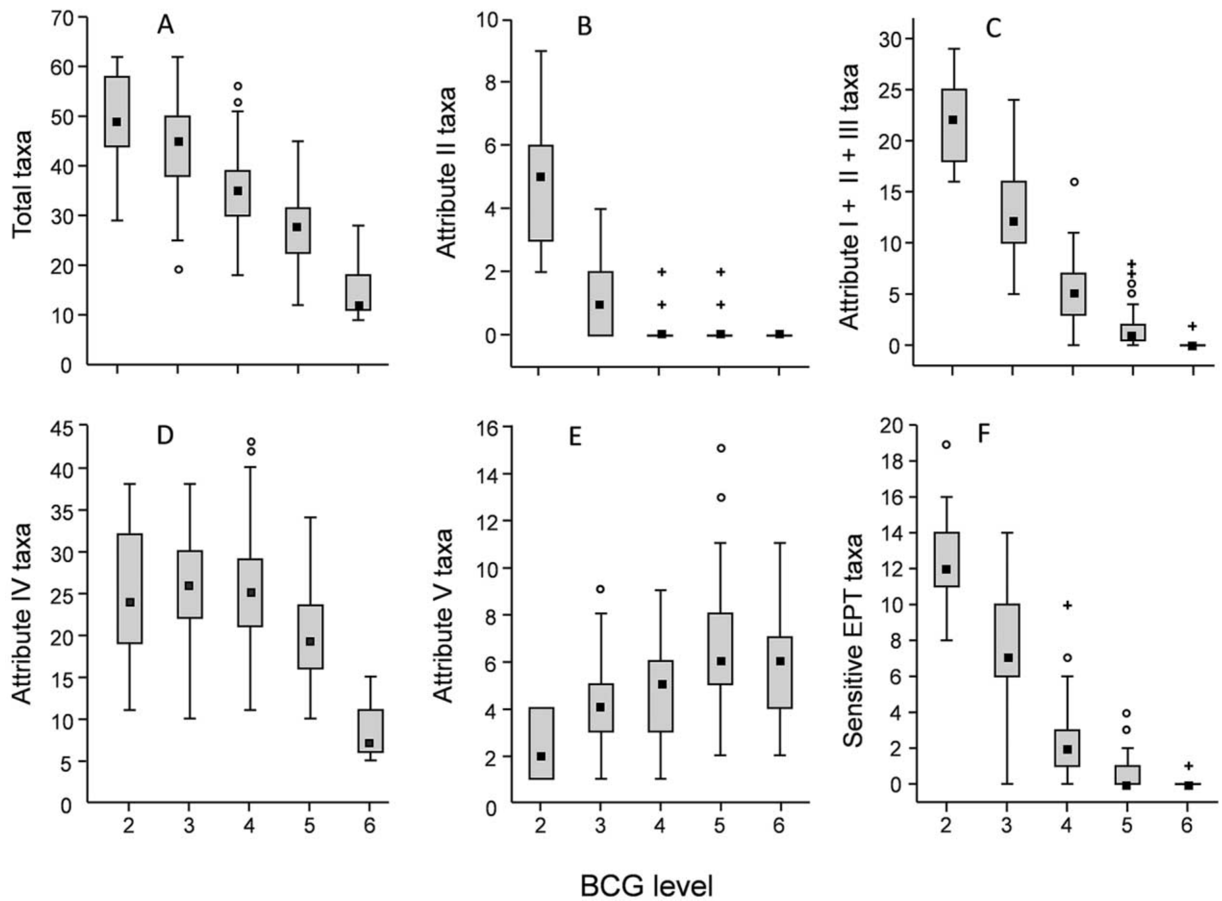


Figure 7. Box-and-whisker plots for the total (A), attribute IV (B), attributes I, II, and III (C), attribute IV (D), attribute V (E), and sensitive Ephemeroptera, Plecoptera, Trichoptera (EPT) (F) number of benthic macroinvertebrate taxa by Biological Condition Gradient (BCG) level. Squares in boxes are medians, boxes are interquartile range (IQR), whiskers are to 1.5 × IQR, circles are outliers up to 3 IQR, and crosses show extreme values > 3 IQR.

8A–F, 9A–D, 10A–E). For example, if the panel identified a moderate number of sensitive taxa for BCG Level 3, then we examined the number of sensitive taxa in samples the panel assigned to BCG Level 3. We then selected a reasonable minimum of the distribution of sensitive taxa in BCG Level 3, say the minimum or a $10^{th}$ quantile, as the decision threshold. This process was repeated for all rules and attributes identified by the panel as being important to their decisions. Sample sizes for the highest and lowest levels (BCG Levels 1, 2, and 6) were small, and required increased professional judgment from the panel to develop rules.

For a particular attribute or metric, the threshold identified by the panel typically was the 50% membership value in a fuzzy membership function. For example, if the panel identified ">10" sensitive taxa as a requirement for BCG Level 3 (Fig. 7A–F), then 10 taxa would correspond to 50% membership, 5 taxa might correspond to 0% membership, and 15 taxa to 100%. Because number of taxa is always a whole number, this membership function is not continuous. Some rules are non-fuzzy: if a rule requires "≥1" or "presence," then presence receives a membership of 100% and absence receives 0%. Final rules for all 14 assessed stream types are in Tables S3–S8. We include 2 sets of rules here for illustration: riffle–run invertebrate samples (Table 6) and wadeable stream fish samples (Table 7).

Panelists preferred to use taxon richness within the sensitive attributes as the most important criteria for setting site BCG level assignments. Thus, the number of sensitive taxa was most often used to distinguish BCG Level 2 from Level 3 sites. BCG Level 2 should have several highly sensitive taxa (attribute II), but their richness may be reduced or absent in BCG Level 3. All of the BCG Level 1 fish samples had ≥2 attribute I taxa (rare or endemic taxa). Higher BCG levels (1–3) all required some minimum relative abundance or relative richness of sensitive taxa (attributes I–III). In addition, for a site to be considered in Level 1 to Level 3, participants often placed upper limits on the abundance and dominance of tolerant taxa, especially attributes V and Va (for fish). Going further down the gradient, BCG Level 4 typically had a fairly low minimum requirement for sensi-

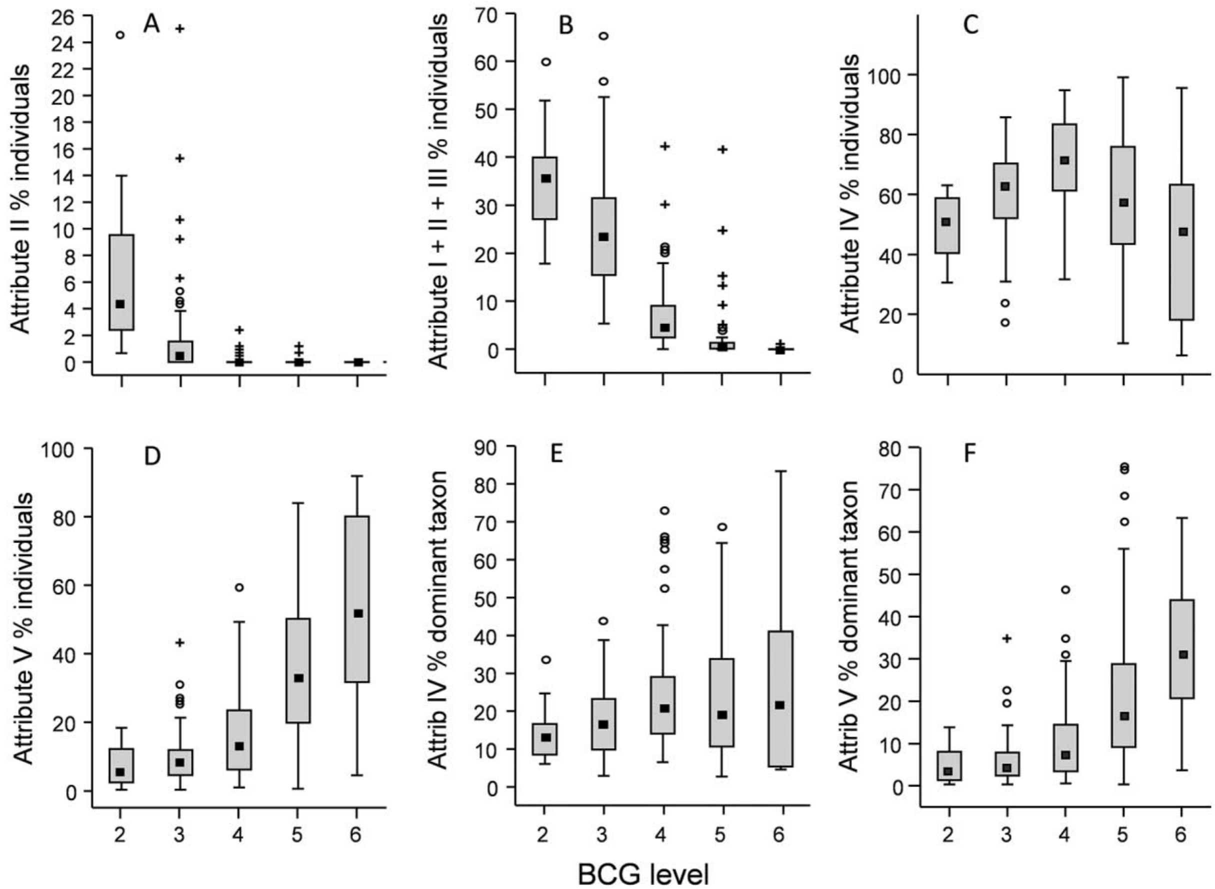


Figure 8. Box-and-whisker plots for the % attribute II (A), attributes I, II, and III (B), attribute IV (C), attribute V (D) individuals and % dominance of attribute IV (E) and attribute V (F) genera of benthic macroinvertebrates by Biological Condition Gradient (BCG) level. See Fig. 7 for explanation of plots.

tive taxa (attribute III), sufficient to show they had not disappeared. BCG Level 5 usually had only requirements of minimum overall richness, and often a maximum dominance (not to be exceeded) of a tolerant taxon. Failure of Level 5 rules result in an assessment of Level 6. The decision patterns described here are consistent with those developed in other states and regions by other panels for invertebrates and fish (see case studies in USEPA 2016).

Rules (Tables 6, 7, S3–S8) were expressed as an inequality, a midpoint, and a range: e.g., ≥20 (15–25). The first number is the midpoint, and the range is in parentheses, where the range describes the linear fuzzy membership function as it increases from 0 to 1 for '≥' and decreases from 1 to 0 for '≤'. Thus, for a rule expressed as ≥20% (15–25), the given membership is 0 at a metric value ≤15%; rises linearly to 1 at a metric value of 25%; and remains 1 for values >25%. The membership is 0.5 at the midpoint of 20%.

Some rule sets included alternatives; i.e., 2 or 3 alternative rules may exist for a certain BCG level (e.g., BCG Level 3 in Table 6, Levels 4 and 5 in Table 7). At least one of the alternatives must be true for the site to be assigned to that level. Alternatives usually reflected a trade-off specified by the panel. For example, a high number of total taxa could offset a low proportion of sensitive taxa, and vice versa. Rules *within* each alternative are joined by AND operators, and the 2 or 3 alternatives are then joined by OR operators to assign level.

## Model performance

To evaluate the performance of the quantitative decision model, we assessed the number of samples where the BCG decision model's nominal level exactly matched the panel's median (exact match) and the number of samples where the model predicted a BCG level that differed from the median expert opinion (mismatch samples). For the mismatched samples, we examined the size of the difference between the BCG level assignments.

The model output is in terms of relative membership (0–100%) of a site among BCG levels, where memberships of all levels must sum to 100%. Model output could yield ties between adjacent levels, or a majority could be assigned to 1 level over ≥1 other levels. As with the quanti-

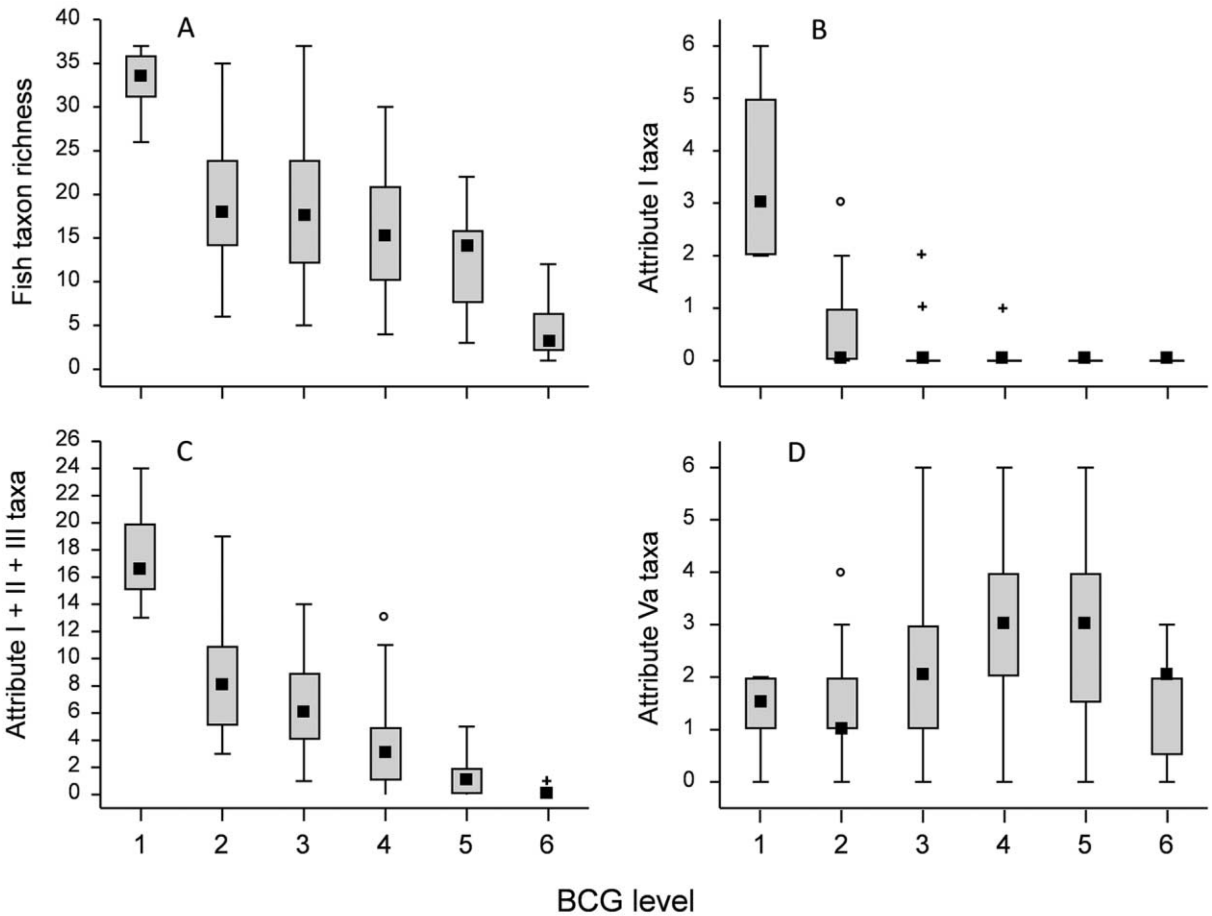


Figure 9. Box-and-whisker plots for the total (A), attribute I (B), attributes I, II, and III (C), and attribute Va (D) number of fish taxa by Biological Condition Gradient (BCG) level. See Fig. 7 for explanation of plots.

tative model, panelists' site ratings could be split among BCG levels.

To estimate concurrence between the quantitative model and the panel, we assigned scores as clear majority or ties and near-ties based on the panelists' votes and the model membership outcomes. We assigned ties and near-ties where either the model or the panel was divided. For model ties, nearly equal membership was present in 2 BCG levels (e.g., membership of 0.4–0.6 in BCG Level 2 and membership of 0.6–0.4 in BCG Level 3). Panelist ties were site ratings where a single vote could have flipped the decision (e.g., 4–4 or 5–4 decisions).

If either the BCG model assigned a tie that did not match with the panelist consensus, or vice-versa, we assigned a difference of ½ BCG level. For example, if the model assignment was a BCG Level 2–3 tie and panelist consensus was BCG Level 2, the model was considered to be off by ½ BCG level; more specifically, the model rating was a ½ BCG level worse than the panelists' consensus. To avoid cutting the differences too finely, we considered mismatches by units of only ½ BCG level. These units were: match (i.e., both panel and model a clear majority for the same level or the same tie); ≤½ level (i.e., panel and model mismatch by ≤½ BCG level); ≤1 level (i.e., panel and model mismatch ½ but ≤1 BCG level); and so on.

Model performance is summarized in Tables 8 and 9, which show the number and % model assessments compared to panel assessments. The panel did not consider a ½-level mismatch with their consensus to be a meaningfully different assessment, and a ½ level was similar to the spread in ratings among panel members. Thus, the panel was unwilling to adjust ratings or to modify rules for small mismatches. On average, the macroinvertebrate models were 89% accurate in replicating the panel assessments within ½ BCG level, and the fish models were 86% accurate. The fish model had 2 mismatches >1 BCG level.

We compared BCG model performance on all sites to IBI models, which had been developed independently. Neither IBI model nor BCG model was regarded as objective truth. Rather, the comparison was used to identify situations where, in the expert opinion of the panel, either or both models might need modifications. Overall, the IBI and BCG models corresponded to each other, but interquartile ranges did overlap between adjacent BCG levels

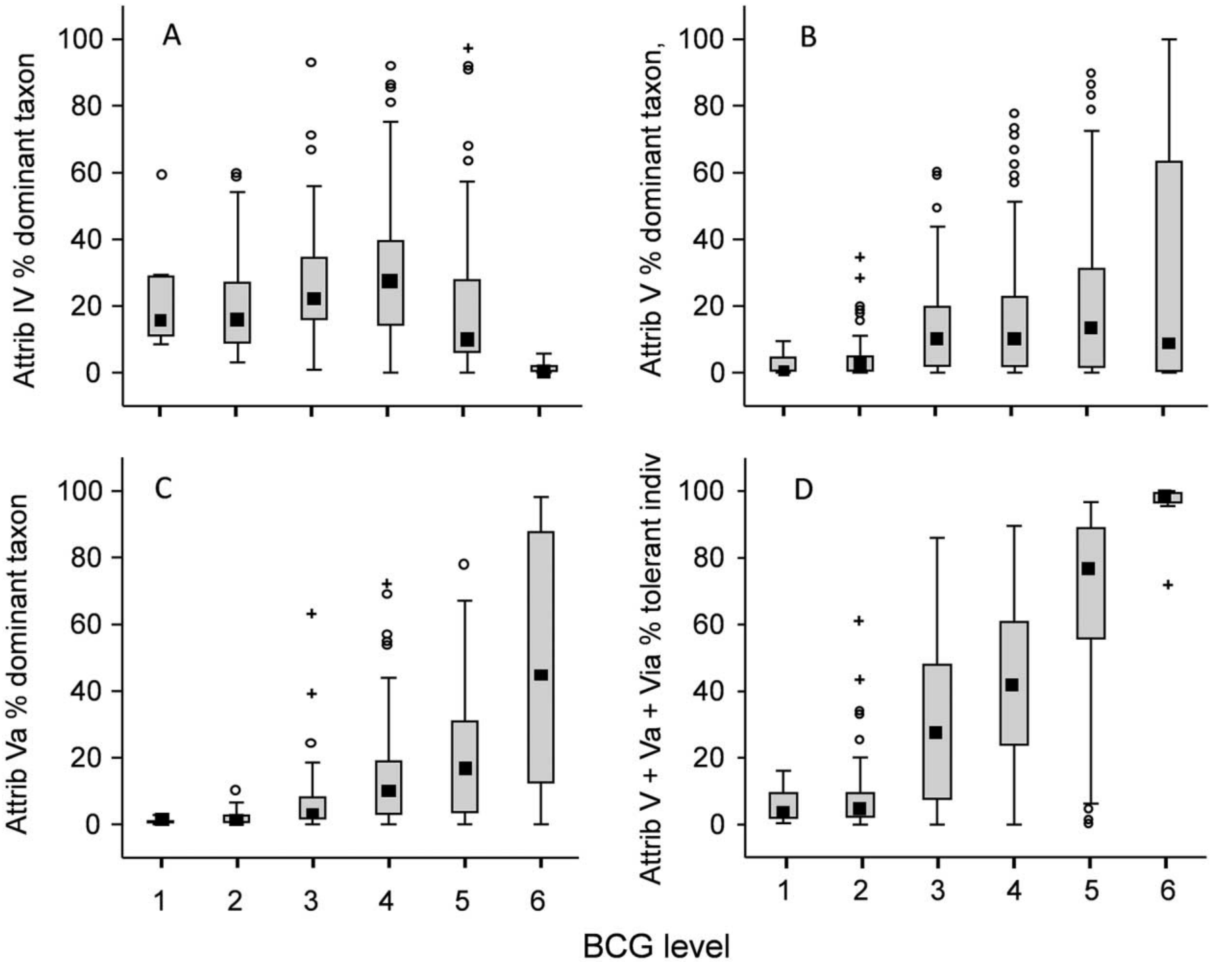


Figure 10. Box-and-whisker plots for the % dominance of attribute II (A), attribute V (B), and attribute Va (C) fish, and % tolerant (attributes V, Va, and VIa) individuals (D) by Biological Condition Gradient (BCG) level. See Fig. 7 for explanation of plots.

(Figs 11A–G, 12A–G). In some stream types, the distribution of IBI scores for BCG Level 6 appeared anomalous. Differences between the 2 models often were the result of differences in the scoring approaches. For example, with the IBIs, a biological sample might score extremely poorly for a single metric, but because the final score is a sum of multiple metric scores, the final score could still be high or intermediate if other metrics score high, a phenomenon known as "eclipsing" (Suter 1993).

The exercise also identified situations where the panel thought the BCG rules were too stringent, and the rules subsequently were relaxed. These changes included modifying the thresholds for some metric criteria or, in some cases, addition of alternate criteria (e.g., BCG Level 3 in Table 6). The alternate criteria provide multiple paths to a higher BCG level score for a sample and account for the diversity of healthy communities within a stream type. The rule changes improved the applicability of the BCG models beyond the population of the sites used in the model development and testing efforts. This exercise also indicated where changes should be made to the IBI models. The process identified a small number of samples with poorly scoring biological communities in relatively undisturbed watersheds. These streams were often wetland-influenced streams, and new IBI and BCG models are needed to measure biological condition appropriately for this type of stream.

### Fish-invertebrate assemblage comparison

An issue of interest to managers is whether fish and macroinvertebrate assemblages yield the same results, and whether both must be monitored. We examined BCG assessments by the panel for a set of sites with both fish and benthic macroinvertebrate samples that were sampled in the same calendar year (typically within 1–3 mo). The maximum difference found was 3 BCG levels in 2 rivers where the fish were rated Level 2 but the invertebrates were rated Level 5. Both assemblages were rated at ≤1 BCG

Table 6. Decision rules for macroinvertebrate assemblages in high-gradient streams (riffle–run habitat). Rules show the midpoint and ranges (in parentheses) of fuzzy membership functions (see Fig. 6). $N$ is the number of sites at the indicated Biological Condition Gradient (BCG) level and stream type in the calibration data set. 'Alt' designation in rules identifies alternative rule sets for a particular stream type and BCG Level (see text for details). EPT = Ephemeroptera, Plecoptera, Trichoptera; '= Alt 1' indicates the rule is the same as given under Alt 1 for this metric; n/a = not applicable.

| | Northern forest streams, high-gradient | | Southern streams, high-gradient | |
|---|---|---|---|---|
| Metric | Alt 1 | Alt 2 | Alt 1 | Alt 2 |
| BCG Level 2 | $N = 2$ | | $N = 0$[a] | |
| Total taxa | ≥40 (35–45) | n/a | ≥40 (35–45) | n/a |
| Attribute I+II taxa | >3 (2–5) | n/a | >3 (2–5) | n/a |
| Attribute I+II+III % taxa | ≥50% (45–55) | n/a | ≥50% (45–55) | n/a |
| Attribute I+II+III % individuals | ≥30% (25–35) | n/a | ≥30% (25–35) | n/a |
| Attribute V % individuals | ≤10% (7–13) | n/a | ≤10% (7–13) | n/a |
| Sensitive EPT taxa | >11 (9–14) | n/a | >11 (9–14) | n/a |
| BCG Level 3 | $N = 17$ | | $N = 8$ | |
| Total taxa | ≥30 (25–35) | ≥45 (40–50) | ≥30 (25–35) | ≥45 (40–50) |
| Attribute I+II+III % taxa | ≥20% (15–25) | ≥15% (10–20) | ≥20% (15–25) | ≥10% (7–13) |
| Attribute I+II+III % individuals | ≥10% (7–13) | ≥5% (3–7) | ≥15% (10–20) | ≥5% (3–7) |
| Attribute IV dominance | ≤25% (20–30) | = Alt 1 | n/a | n/a |
| Attribute V % individuals | n/a | n/a | ≤20% (15–25) | = Alt 1 |
| Attribute V dominance | ≤35% (30–40) | = Alt 1 | ≤10% (7–13) | = Alt 1 |
| Sensitive EPT taxa | >3 (2–5) | = Alt 1 | >3 (2–5) | = Alt 1 |
| BCG Level 4 | $N = 9$ | | $N = 19$ | |
| Total taxa | ≥20 (16–24) | n/a | ≥20 (16–24) | ≥30 (25–35) |
| Attribute I+II+III % taxa | ≥10% (7–13) | n/a | ≥5% (3–7) | Present |
| Attribute I+II+III % individuals | Present | n/a | ≥5% (3–7) | Present |
| Attribute V % individuals | ≤25% (30–40) | n/a | ≤25% (30–40) | ≤40% (35–45) |
| Attribute V dominance | ≤25% (20–30) | n/a | ≤20% (15–25)13 | = Alt 1 |
| Sensitive EPT | Present | n/a | Present | = Alt 1 |
| BCG Level 5 | $N = 2$ | | $N = 20$ | |
| Total taxa | >13 (11–16) | ≥20 (16–24) | >13 (11–16) | ≥20 (16–24) |
| Attribute II+III+IV % taxa | n/a | n/a | n/a | ≥50% (45–55) |
| Attribute V % taxa | ≤40% (35–45) | ≤50% (45–55) | ≤40% (35–45) | n/a |
| Attribute V dominance | ≤60% (55–65) | = Alt 1 | ≤60% (55–65) | n/a |
| BCG Level 6 | $N = 0$ | | $N = 0$ | |

[a] BCG rules for southern streams, high-gradient Level 2 provisionally set to same criteria as northern forest streams, high-gradient

level apart in 83% of the sample sets (Table 10). The macroinvertebrate panel was slightly more stringent than the fish panel: no invertebrate samples were rated BCG Level 1, and slightly fewer Levels 2 and 3 ratings were given by the macroinvertebrate panel than by the fish panel. More large differences of ≥2 levels (Table 10) occurred at river than at wadeable stream sites. Fish and invertebrates were rated at ≥2 BCG levels apart in 40% of large river sites (non-wadeable; drainage area > 1300 $km^2$) but in only 9% of wadeable sites. The 2 assemblages respond to different stressors, so we would not expect a perfect correlation between ratings based on macroinvertebrates and on fish. Both assemblages are sampled and assessed because of their different responses.

## DISCUSSION

Recent developments of environmental assessment using professional judgment have shown that experts are highly concordant in their ratings of marine benthic macroinvertebrates (Weisberg et al. 2008, Teixeira et al. 2010), marine sediment quality (Bay et al. 2007, Bay and Weisberg 2010), and fecal contamination (Cao et al. 2013). In the pilot BCG studies (USEPA 2016), aquatic biologists have

Table 7. Decision rules for fish assemblages in wadeable streams. Rules show the midpoint and ranges (in parentheses) of fuzzy membership functions (see Fig. 6). *N* is the number of sites at the indicated Biological Condition Gradient (BCG) level and stream type in the calibration data set. 'Alt' designation in rules identifies alternative rule sets for a particular stream type and BCG level (see text for details). '= Alt 1' indicates the rule is the same as given under Alt 1 for this metric; n/a = not applicable.

| | Southern streams | | | Northern streams | |
|---|---|---|---|---|---|
| Metric | Alt 1 | Alt 2 | Alt 3 | Alt 1 | Alt 2 |
| BCG Level 1 | | $N = 0$[a] | | $N = 0$[a] | |
| Total taxa | ≥30 (25–35) | n/a | n/a | ≥30 (25–35) | n/a |
| Attribute I endemic taxa | Present | n/a | n/a | Present | n/a |
| Attribute I+II taxa | >3 (2–5) | n/a | n/a | >3 (2–5) | n/a |
| Attribute I+II+III % taxa | ≥50% (45–55) | n/a | n/a | ≥50% (45–55) | n/a |
| Attribute I+II+III % individuals | ≥30% (25–35) | n/a | n/a | ≥30% (25–35) | n/a |
| Tolerant % individuals (V+Va+VIa) | ≤5% (3–7%) | n/a | n/a | ≤5% (3–7%) | n/a |
| BCG Level 2 | | $N = 1$ | | $N = 8$ | |
| Total taxa | ≥20 (16–24) | n/a | n/a | >13 (11–16) | n/a |
| Attribute I+II+III total taxa | ≥8 (6–10) | n/a | n/a | n/a | n/a |
| Attribute I+II+III % taxa | ≥40% (35–45) | n/a | n/a | ≥30% (25–35) | n/a |
| Attribute I+II+III % individuals | ≥10% (7–13) | n/a | n/a | ≥10% (7–13) | n/a |
| Attribute Va or VIa dominance | n/a | n/a | n/a | ≤10% (7–13) | n/a |
| Tolerant % individuals (V+Va+VIa) | n/a | n/a | n/a | ≤35% (30–40) | n/a |
| Highly tolerant % individuals (Va+VIa) | ≤20% (15–25) | n/a | n/a | n/a | n/a |
| BCG Level 3 | | $N = 4$ | | $N = 10$ | |
| Total taxa | >13 (11–16) | n/a | n/a | >13 (11–16) | n/a |
| Attribute I+II+III % taxa | ≥10% (7–13) | n/a | n/a | ≥25% (20–30) | n/a |
| Attribute I+II+III % individuals | ≥5% (3–7) | n/a | n/a | ≥5% (3–7) | n/a |
| Attribute Va or VIa dominance | ≤20% (15–25) | n/a | n/a | ≤10% (7–13) | n/a |
| Highly tolerant % individuals (Va+VIa) | ≤40% (35–45) | n/a | n/a | ≤20% (15–25) | n/a |
| BCG Level 4 | | $N = 10$ | | $N = 15$ | |
| Total taxa | ≥8 (6–10) | ≥20 (16–24) | n/a | ≥8 (6–10) | = Alt 1 |
| Attribute I+II+III % taxa | Present | n/a | n/a | ≥5% (3–7) | n/a |
| Attribute 1+ 2+3 % individuals | ≥0.5% (0–1) | n/a | n/a | Present | n/a |
| I+II+III+IV % individuals | n/a | n/a | n/a | n/a | ≥70% (65–75) |
| Attribute I+II+III+IV % taxa | n/a | n/a | n/a | n/a | ≥50% (45–55) |
| Attribute Va or VIa dominance | ≤50% (45–55) | = Alt 1 | n/a | ≤30% (25–35) | ≤20% (15–25) |
| Tolerant % individuals (V+Va+VIa) | ≤70% (65–75) | = Alt 1 | n/a | n/a | n/a |
| Highly tolerant % individuals (Va+VIa) | ≤60% (55–65) | = Alt 1 | n/a | ≤60% (55–65) | n/a |
| BCG Level 5 | | $N = 18$ | | $N = 4$ | |
| Total taxa | ≥5 (3–7) | >13 (11–16) | ≥20 (16–24) | ≥3 (1–5) | n/a |
| Attribute I+II+III % taxa | n/a | Present | n/a | n/a | n/a |
| Attribute I+II+III+IV % taxa | ≥10% (7–13) | n/a | ≥20% (15–25) | ≥15% (10–20) | n/a |
| Attribute Va or VIa dominance | ≤50% (45–55) | n/a | n/a | <65–75% | n/a |
| Highly tolerant % individuals (Va+VIa) | ≤70% (65–75) | n/a | n/a | n/a | n/a |
| BCG Level 6 | | $N = 2$ | | $N = 0$ | |

[a] BCG rules for Level 1 provisionally set to same criteria as Prairie Rivers (Table S4).

come to very tight consensus on the descriptions of individual levels of the BCG and on the BCG level assigned to individual sites. The Minnesota BCG reported here confirms the concordance among experts.

The conceptual model of the BCG was derived from experience of working aquatic ecologists from across the US (Davies and Jackson 2006). Development of the quantitative BCG requires quantitative mapping of biological infor-

Table 8. Performance of Biological Condition Gradient (BCG) quantitative macroinvertebrate models. ‘Better’ and ‘worse’ indicate model assessment of stream condition compared to panel (e.g., ‘better’ if model assessed BCG Level 2, but panel assessed BCG Level 3). $N$ = number of comparisons in category, % = % of comparisons in category.

| | | Quantitative model performance | | | | | |
|---|---|---|---|---|---|---|---|
| Invertebrate stream type | Type | 1 better | 0.5 better | Match | 0.5 worse | 1 worse | Total |
| Northern forest rivers | $N$ | 2 | 2 | 26 | 2 | 5 | 37 |
| | % | 5% | 5% | 70% | 5% | 14% | |
| Prairie and southern rivers | $N$ | 0 | 4 | 21 | 3 | 1 | 29 |
| | % | 0% | 14% | 72% | 10% | 3% | |
| Northern forest high-gradient | $N$ | 1 | 1 | 27 | 3 | 5 | 37 |
| | % | 3% | 3% | 73% | 8% | 14% | |
| Northern forest low-gradient | $N$ | 2 | 1 | 28 | 2 | 2 | 35 |
| | % | 6% | 3% | 80% | 6% | 6% | |
| Southern high-gradient | $N$ | 1 | 2 | 35 | 5 | 2 | 45 |
| | % | 2% | 4% | 78% | 11% | 4% | |
| Southern forest low-gradient | $N$ | 2 | 1 | 29 | 3 | 1 | 36 |
| | % | 6% | 3% | 81% | 8% | 3% | |
| Prairie low-gradient | $N$ | 3 | 2 | 44 | 0 | 3 | 52 |
| | % | 6% | 4% | 85% | 0 | 6% | |
| Total | $N$ | 11 | 13 | 210 | 18 | 19 | 271 |
| | % | 4% | 5% | 77% | 7% | 7% | |

mation into the conceptual and theoretical model. The BCG is calibrated using a data set, but also requires ecological considerations with wide expert agreement from biologists familiar with the resources. The result is intended to be more general than a regression analysis of biological response to stressors. The BCG uses universal attributes (in this application, only the taxonomic attributes I–VI) that are intended to apply in all regions. Specifics of the attributes (taxon membership, attribute groups indicating good, fair, poor, etc.) do vary across regions and stream types, but the attributes themselves and their importance are consistent. The BCG requires descriptions of the levels from pristine to degraded. Documentation of the rationale for making BCG level determinations (i.e., the rules) provides the foundation for building a robust quantitative model and ensures that future information and discoveries can be related back to the level descriptions.

The approach requires substantial time and effort from the expert panel, but does it also require a rich database? We think the BCG calibration itself can be done with a smaller data set. Stress–response analysis benefits from a large database because we generally require a minimum of 20 occurrences of a taxon to develop the stress–response model. Other sources of tolerances for attribute assignments in the absence of stress–response analysis include existing literature and panelists’ experience with the taxa. Early BCG calibrations were successful with 50 to 100 sites assessed by the panel, and stress–response was not used in those efforts (e.g., case studies in USEPA 2016). As a general rule, ≥30 sites in each stream type and perhaps as few as 20 is sufficient for rule development.

In a critique of ecosystem health and indexes, Suter (1993) pointed out technical weakness of common indexes. Weaknesses include: 1) *ambiguity*: one cannot tell why an index value is high or low (although individual metric values will reveal it); 2) *eclipsing*: a high metric value balances a low metric value, with a resulting inappropriate score (site is better or worse than its score indicates); and 3) *arbitrary combining functions*: most multimetric indexes (and observed/expected taxon ratios) are the sum of the component metrics (or component reference taxa), with no weighting or other combining function, nor consideration of why or why not to do so (Suter 1993). Eclipsing is one consequence of arbitrary equal weighting and summing. In the BCG rule-based method, weighting and combining functions are stated and not arbitrary. For example, a rule for a BCG level may require a certain number of sensitive taxa. If a site has too few sensitive taxa, it will be rated at a lower level because the sensitive taxa rule failed. Rules prevent ambiguity (i.e., we know why it failed), eclipsing (i.e., a high value in another attribute or metric does not change the decision, unless a rule specifically allows it), and the combining function for the rules is not arbitrary (i.e., transparent and established by the panel).

We do not suggest that the BCG is a panacea for all current issues in bioassessment. It has distinct disadvantages in development and acceptance. For example, the BCG is labor-intensive to develop, requires a panel of experts who are knowledgeable about local water bodies and biota. It can-

Table 9. Performance of Biological Condition Gradient (BCG) quantitative fish models. 'Better' and 'worse' indicate model assessment of stream condition compared to panel (e.g., 'better' if model assessed BCG Level 2, but panel assessed BCG Level 3). $N$ = number of comparisons in category, % = % of comparisons in category.

| | | Quantitative model performance | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Fish stream type | Type | 1.5 better | 1 better | 0.5 better | Match | 0.5 worse | 1 worse | 1.5 worse | 2 worse | Total |
| Northern rivers | N | 0 | 4 | 2 | 36 | 4 | 1 | 0 | 0 | 47 |
| | % | 0% | 9% | 4% | 77% | 9% | 2% | 0 | 0% | |
| Southern rivers | N | 0 | 5 | 4 | 52 | 10 | 4 | 0 | 0 | 75 |
| | % | 0% | 7% | 5% | 69% | 13% | 5% | 0 | 0% | |
| Northern streams | N | 1 | 1 | 3 | 22 | 8 | 1 | 0 | 1 | 37 |
| | % | 3% | 3% | 8% | 59% | 22% | 3% | 0 | 3% | |
| Northern headwaters | N | 0 | 2 | 4 | 19 | 2 | 3 | 0 | 0 | 30 |
| | % | 0% | 7% | 13% | 63% | 7% | 10% | 0 | 0% | |
| Southern streams | N | 0 | 4 | 1 | 23 | 2 | 5 | 0 | 0 | 35 |
| | % | 0% | 11% | 3% | 66% | 6% | 14% | 0 | 0% | |
| Southern headwaters | N | 0 | | 1 | 25 | 3 | 3 | 0 | 0 | 32 |
| | % | 0% | 0% | 3% | 78% | 9% | 9% | 0 | 0% | |
| Low-gradient streams | N | 0 | 2 | 1 | 26 | 1 | 2 | 0 | 0 | 32 |
| | % | 0% | 6% | 3% | 81% | 3% | 6% | 0 | 0% | |
| Total | N | 1 | 18 | 16 | 203 | 30 | 19 | 0 | 1 | 288 |
| | % | 0.3% | 6% | 6% | 70% | 10% | 7% | 0 | 0% | |

not be developed and calibrated by an individual with a data set and a computer. Broad acceptance of the BCG may be problematic. Many scientists and managers sometimes implicitly assume that continuous, quantitative models are somehow better than expert consensus. We contend that this assumption is untested, and may be an unfounded personal bias.

## Decision analysis

To develop the fuzzy decision analysis system, we needed a set of rules to which we could apply fuzzy logic. The greatest single strength of the fuzzy-model approach may be development of a set of transparent rules that can, in principle, be followed by anyone making a decision on a site. Fuzzy-model rules may seem exotic to those not familiar with the approach, but they are fully laid out and are not hidden in a statistical model or in artificial machine learning. Experts can describe the classes of the BCG in a very general way, but without the specific rules and their combination, their decisions cannot be replicated and the rules cannot be modified effectively as new knowledge is gained.

The fuzzy-rule model replicates expert judgment by direct application of rules. It is only as good as the rules themselves. Experts also make errors, so an iterative process is required for rule development to correct inconsistencies, elicit hidden rules, or recalibrate incorrect membership functions. The fuzzy model does not require a statistical model to predict the expert panel decisions. If one accepts the expert consensus and rules of the BCG, then a fuzzy-model approach is the best way we know to automate it.

The rules have no requirement for linearity or monotonicity of metrics or attributes. For example, a linguistic rule that captures subsidy–stress (e.g., Odum et al. 1979) is permissible, such as "If taxon richness is high and abundance is high, then BCG level is $\leq$3." Moderate taxon richness may indicate very good conditions and fair or poor conditions and could be problematic in monotonic applications of taxon richness to condition. Most biotic indexes (e.g., IBI and RIVPACS models; e.g., Barbour et al. 1999) require monotonic responses of component metrics.

Like the BCG, a fuzzy-decision analysis approach has a disadvantage in acceptance. For example, an unfounded linguistic bias exists among American English-speakers against the term "fuzzy" in any scientific context. This bias has resulted in slower acceptance of fuzzy logic systems in English-speaking countries, especially in the USA, than elsewhere because the word 'fuzzy' has colloquial meaning in the USA (fuzzy thinking, warm and fuzzy). Prominent English-speaking scientists revealed their linguistic bias when criticizing fuzzy theory (see quotes in Zadeh 2008). In continental Europe and Japan, fuzzy logic systems are widely used in engineering and decision analysis, including ecological applications (e.g., Ibelings et al. 2003), because

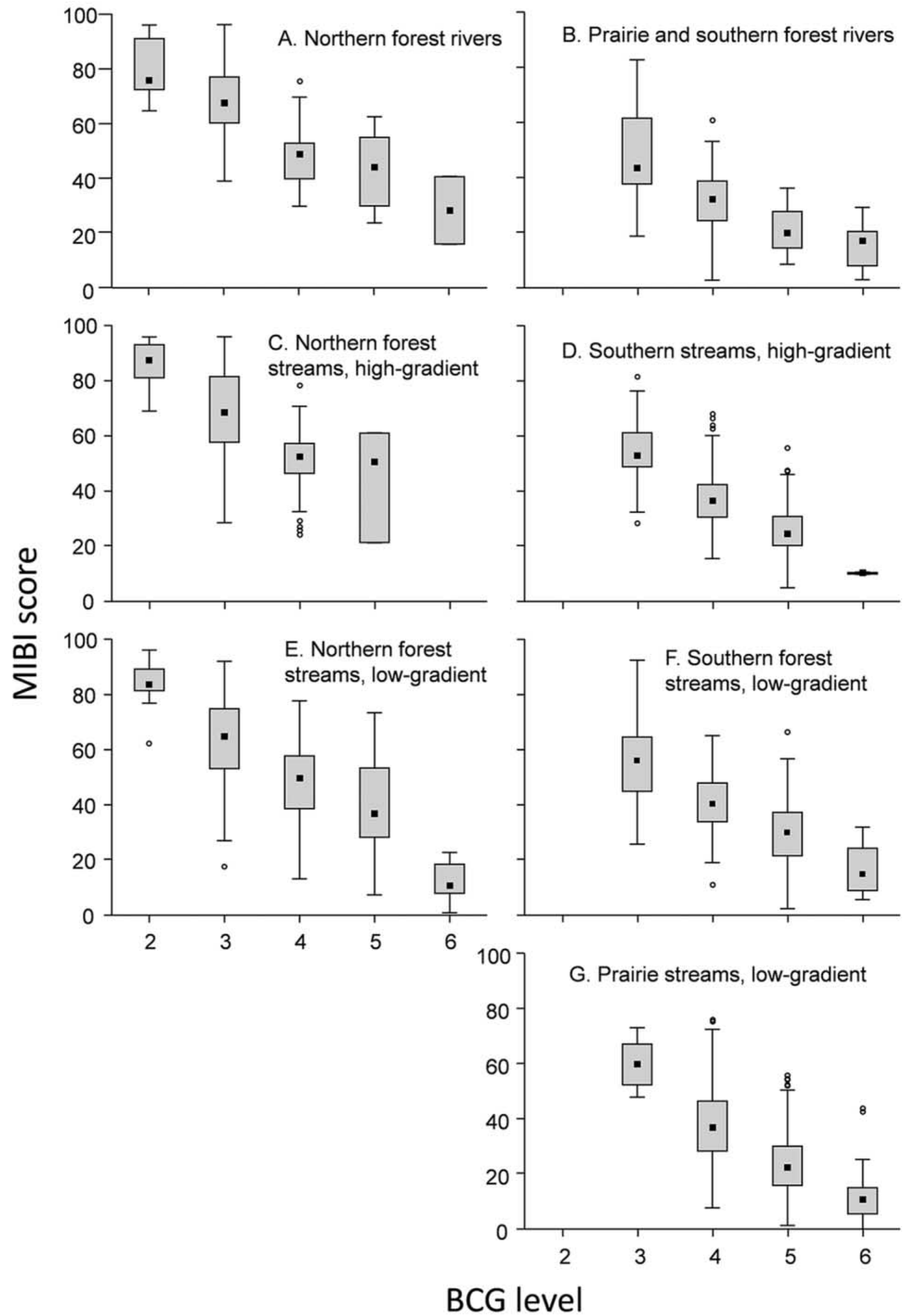


Figure 11. Frequency distributions of Macroinvertebrate Index of Biological Integrity (MIBI) scores by Biological Condition Gradient (BCG) level for northern forest rivers (A), prairie rivers (B), high- (C) and low-gradient (E) gradient northern forest streams, high- (D) and low-gradient (F) southern forest streams, and low-gradient prairie streams (G) in Minnesota at sites sampled from 1996–2011. See Fig. 7 for explanation of plots.

of greater economy of development with respect to nonlinear responses, and because the English word fuzzy has no colloquial connotations in other languages.

We measure things on continuous scales (e.g., pH) or as whole numbers (e.g., counts of taxa), but most interpretations and decisions are binary or categorical. Management and public communication require assessments such as 'no impact', 'slight impact', or 'severe impact'; or decisions such as 'no action' or 'reduce phosphorus by 50%'. Statements such as '5.8 mg/L $O_2$' or '29 insect species' are neither decisions nor interpretations. Fuzzy-decision systems are an explicit and transparent bridge between continuous

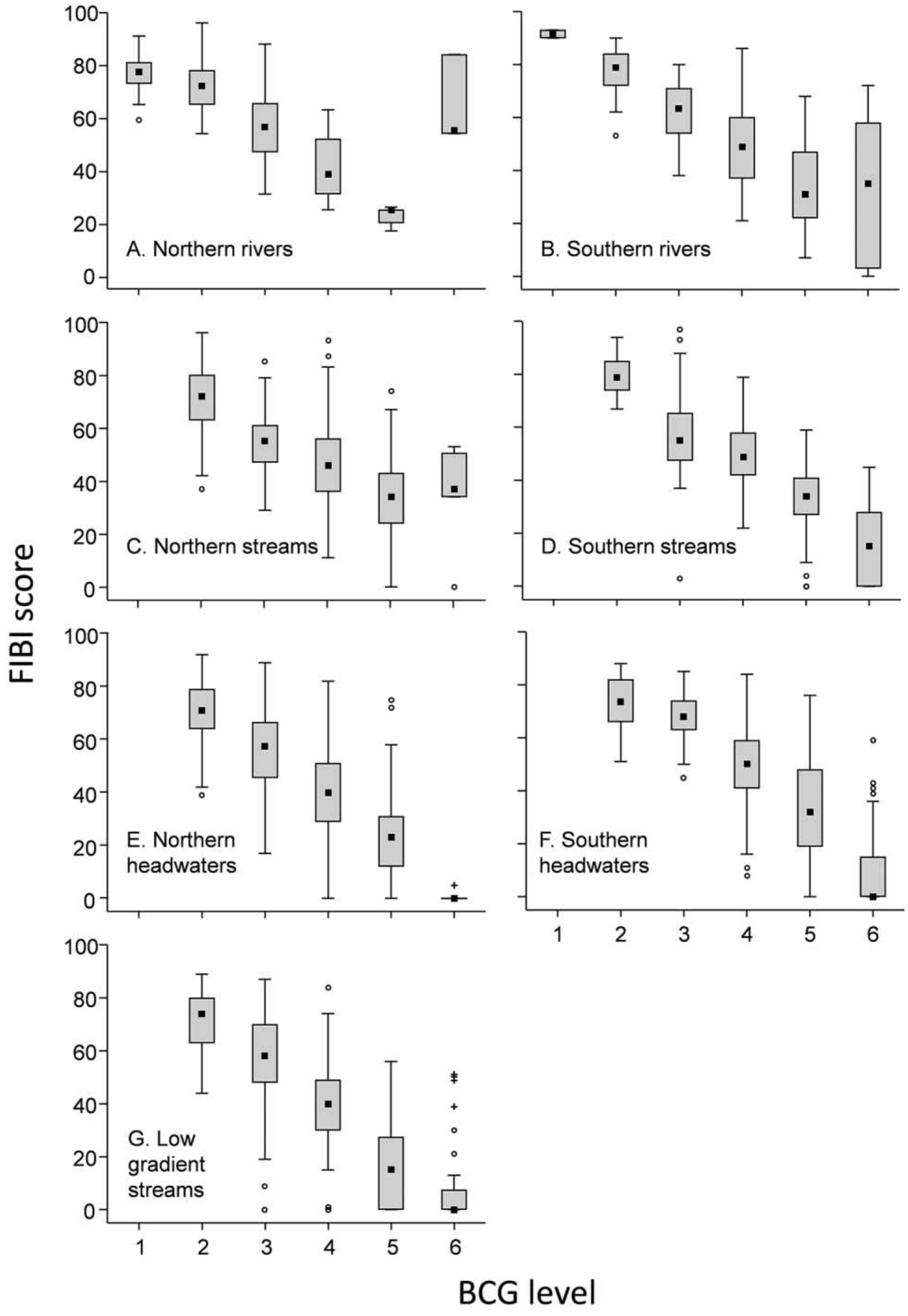


Figure 12. Frequency distributions of Fish Index of Biological Integrity (FIBI) scores by Biological Condition Gradient (BCG) level for northern rivers (A), southern rivers (B), northern streams (C), southern streams (D), northern headwaters (E), southern headwaters (F), and low-gradient streams (G) in Minnesota at sites sampled from 1996–2011. See Fig. 7 for explanation of plots.

measurements and interpretation and management decisions that are categorical (Silvert 2000).

### BCG and multimetric IBIs

The BCG and IBI (Figs 10, 11) results were similar, which was not unexpected because they use the same field data sets. Moreover, the fundamental concept of both is that aquatic systems deviate from natural conditions (embodying biological integrity) with increasing anthropogenic stress. However, the development process differed for the models and, therefore, they do not produce identical results. The BCG concept and methods address some issues that multimetric IBI models cannot. The BCG categorizes biological communities in terms of naturalness, whereas the full range of multimetric IBI scores may reflect only avail-

Table 10. Differences in assessment of Biological Condition Gradient (BCG) level for fish and invertebrate communities at sites where both assemblages were sampled in the same calendar year and assessed by the panels ($N = 76$).

| Variable | Fish more natural | | | Same | Invertebrates more natural | | |
|---|---|---|---|---|---|---|---|
| BCG difference | 3 | 2 | 1 | 0 | 1 | 2 | 3 |
| *N* | 2 | 9 | 22 | 25 | 16 | 2 | 0 |

able conditions. The BCG weights metrics and rules according to the panel's judgments, whereas multimetric IBI indexes weight all metrics equally in the total score. The BCG allows for nonlinear or modal responses in the attributes whereas multimetric IBI metrics are monotonic.

### Management: aquatic life uses

The Minnesota BCG models are promising as a basis for developing decision criteria or biological criteria for Aquatic Life Uses (ALUs). In the USA, the terms 'Use', 'Designated Uses', and 'Aquatic Life Use' have specific meanings for water-quality management in the context of the CWA. A state defines the uses for its waters and develops physical, chemical, and biological criteria to protect those uses. Designated Uses are the water-quality goals for a specific water body and identify the functions and activities that are supported by a state-defined level of water quality. Water-quality standards are reviewed periodically based on new information that may indicate change in appropriateness of use and changes in what might have been considered irreversible.

Designated Uses also include potential quality or condition that may not be attained currently, but could be attained with appropriate controls or restoration. Thus, ALUs can be set according to the biological potential of water bodies, rather than their current condition. For example, infrastructure is not always irreversible, but it can be modified to reduce stresses on water bodies. The BCG may be more robust than current indexes because it allows for nonlinear responses, and has requirements for combinations of metric values in the condition levels.

The BCG models have been used to refine Minnesota's designated uses known as Tiered Aquatic Life Uses (TALUs; Bouchard et al. 2016, MPCA 2014a). TALUs are refined ALUs that articulate the goal for a water body better than a single one-size-fits-all ALU (e.g., Yoder and Rankin 1995, Bouchard et al. 2016). In Minnesota, the BCG was used to develop biological criteria for TALUs and to address differences in the current condition of streams across the state (Bouchard et al. 2016). For example, the prairie regions in Minnesota have been highly altered, resulting in few if any sites that meet the requirements for minimally disturbed reference sites. This situation poses challenges when the typical reference condition approach is used because minimally disturbed streams are needed to establish benchmarks (i.e., biological criteria) for ALUs. The BCG was used as a universal yardstick to set consistent and protective biological criteria across a diverse landscape (Bouchard et al. 2016). It also aligned biological criteria with the narrative language established by the CWA with the proposed TALU narratives. Levels of the BCG are not a priori equivalent to TALUs or water-quality criteria, although a given criterion could be set to a level of the BCG as a policy decision. The BCG is a measurement yardstick, and it does not express policy decisions and breakpoints for designated uses.

The BCG provides a powerful approach for an operational monitoring and assessment program, for communicating resource condition to the public, and for management decisions to protect or remediate water resources. If formalized properly, any person with data can follow the rules to obtain the same level assignments as the group of experts. This property makes the actual decision criteria transparent to stakeholders. Description of BCG Levels 1 and 2 in the BCG process establishes a fixed natural reference (which may no longer exist) to prevent shifting baselines. Understanding of the natural baseline may be modified with new and better information on historic conditions, but both original and modified baselines are documented and not simply a present-day sample. The levels of the BCG are biologically recognizable stages in condition of stream water bodies. They can inform a biological basis for biological criteria and regulation of water bodies. Development of quantitative BCG models yield the technical tools for protecting the highest quality waters through TALU and for developing realistic restoration goals for waters affected by legacy activities (e.g., ditching, impoundments). The BCG allows practical and operational implementation of ALUs in a state's water-quality criteria and standards.

## ACKNOWLEDGEMENTS

Author contributions: JG developed the fuzzy-set modeling, facilitated workshops, and prepared drafts. RWB prepared sections of drafts, assisted in facilitation, and participated as an expert in the workshops. LZ developed stress-response models. EWL organized the database and developed workbooks for the workshops. COY revised sections of drafts, assisted in facilitation, and participated as an expert in the workshops.

This work was funded by contracts between MPCA and the Midwest Biodiversity Institute. This effort depended on the work of the biologists who participated in the BCG development workshops, the subsequent teleconferences, review of the model outputs, and review of the final report. The principal participants of the macroinvertebrate work group workshops for warm-water streams included: Joel Chirhart (MPCA), John Genet (MPCA), Dan Helwig (MPCA), Gary Montz (Minnesota Department of Natural Resources), and Kevin Stroom (MPCA). The fish work group for warm-water streams included: Brenda DeZiel (MPCA), Mike Feist (MPCA), Brett Nagle (Minnesota Department of Natural Resources), Scott Niemela (MPCA), John Sandberg (MPCA), and Konrad Schmidt (Minnesota Department of Natural Resources). We are grateful to all. Susan Jackson (US EPA, Office of Science and Technology) and Ed Hammer (US EPA, Region 5) have been instrumental in promoting the BCG approach. They reviewed the original reports and provided critical insights. Luke Cole, Scott Niemela, an anonymous referee, and Associate Editor Daren Carlisle provided insights and comments that strengthened the manuscript. We thank Clair Meehan for technical review and editing.

## LITERATURE CITED

Barbour, M. T., and J. Gerritsen. 2006. Key features of bioassessment development in the United States of America. Pages 351–366 *in* G. Ziglio, M. Siligardi, G. Flaim (editors). Biological monitoring of rivers. Applications and perspectives. John Wiley and Sons, Chichester, UK.

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

Bay, S., W. Berry, P. M. Chapman, R. Fairey, T. Gries, E. Long, D. MacDonald, and S. B. Weisberg. 2007. Evaluating consistency of best professional judgment in the application of a multiple lines of evidence sediment quality triad. Integrated Environmental Assessment and Management 3:491–497.

Bay, S. M., and S. B. Weisberg. 2010. Framework for interpreting sediment quality triad data. Integrated Environmental Assessment and Management 8:589–596.

Bouchard, R. W., S. Niemela, J. A. Genet, C. O. Yoder, J. Sandberg, J. W. Chirhart, M. Feist, B. Lundeen, and D. Helwig. 2016. A novel approach for the development of tiered use biological criteria for rivers and streams in an ecologically diverse landscape. Environmental Monitoring and Assessment 188:1–26.

Cao, Y., C. Hagedorn, O. C. Shanks, D. Wang, J. Ervin, J. F. Griffith, B. A. Layton, C. D. McGee, T. E. Reidel, and S. B. Weisberg. 2013. Towards establishing a human fecal contamination index in microbial source tracking. International Journal of Environmental Science and Engineering Research 4:46–58.

Carlisle, D. M., M. R. Meador , S. R. Moulton, and P. M. Ruhl. 2007. Estimation and application of indicator values for common macroinvertebrate genera and families of the United States. Ecological Indicators 7:22–33.

Castella, E., and M. Speight. 1996. Knowledge representation using fuzzy coded variables: an example based on the use of Syrphidae (Insecta, Diptera) in the assessment of riverine wetlands. Ecological Modelling 85:13–25.

Cheung, W. W., T. J. Pitcher, and D. Pauly. 2005. A fuzzy logic expert system to estimate intrinsic extinction vulnerabilities of marine fishes to fishing. Biological Conservation 124:97–111.

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

Dayton, P. K., M. J. Tegner, P. B. Edwards, and K. L. Riser. 1998. Sliding baselines, ghosts, and reduced expectations in kelp forest communities. Ecological Applications 8:309–322.

Demicco, R. V., and G. J. Klir. 2004. Introduction. Pages 1–9 *in* R. V. Demicco and G. J. Klir (editors). Fuzzy logic in geology. Elsevier Academic Press, San Diego, California.

Droesen, W. J. 1996. Formalisation of ecohydrological expert knowledge applying fuzzy techniques. Ecological Modelling 85:75–81.

EU Commission. 2015. Report from the commission to the European Parliament and the Council on the implementation of the Water Framework Directive (2000/60/EC) river basin management plans. (Available from: http://ec.europa.eu/environment/water/water-framework/pdf/3rd_report/CWD-2012-379_EN-Vol1.pdf)

Frey, D. G. 1977. Biological integrity of water—an historical approach. Pages 127–140 *in* R. K. Ballentine and L. J. Guarraia (coordinators). The Integrity of Water. Proceedings of a Symposium, March 10–12, 1975. US Environmental Protection Agency, Washington, DC.

Gerritsen, J., and E. W. Leppo. 2005. Biological Condition Gradient for Tiered Aquatic Life use in New Jersey. Prepared by TetraTech, Owings Mills, Maryland, for the US Environmental Protection Agency Office of Water, Environmental Protection Agency Region 2, and New Jersey Department of Environmental Protection. (Available at: http://www.nj.gov/dep/wms/bears/docs/FINAL%20TALU%20NJ%20RPT_2.pdf)

Gerritsen, J., and J. Stamp. 2013. Calibration of the Biological Condition Gradient (BCG) in cold and cool waters of the Upper Midwest BCG for fish and benthic macroinvertebrate assemblages. Prepared by TetraTech, Owings Mills, Maryland, for the US Environmental Protection Agency, Washington, DC. (Available from: https://www.uwsp.edu/cnr-ap/biomonitoring/Documents/pdf/USEPA-BCG-Report-Final-2012.pdf)

Gerritsen, J., L. Zheng, E. Leppo, and C. O. Yoder. 2013. Calibration of the biological condition gradient for streams of Minnesota. Prepared for the Minnesota Pollution Control Agency, St. Paul, Minnesota. (Available from: https://www.pca.state.mn.us/sites/default/files/wq-s6-32.pdf)

Hausmann, S., D. F. Charles, J. Gerritsen, and T. J. Belton. 2016. A diatom-based biological condition gradient (BCG) approach for assessing impairment and developing nutrient criteria for streams. Science of the Total Environment 562:914–927.

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

Hering, D., A. Borja, J. Carstensen, L. Carvalho, M. Elliott, C. K. Feld, A. Heiskanen, R. K. Johnson, J. Moe, D. Pont, A. L. Solheim, and W. van de Bund. 2010. The European Water Framework Directive at the age of 10: a critical review of the achievements with recommendations for the future. Science of the Total Environment 408:4007–4019.

Hose, G., E. Turak, and N. Waddell. 2004. Reproducibility of AUSRIVAS rapid bioassessments using macroinvertebrates. Journal of the North American Benthological Society 23:126–139.

Huttunen, K.-L., H. Mykrä, and T. Muotka. 2012. Temporal variability in taxonomic completeness of stream macroinvertebrate assemblages. Freshwater Science 31:423–441.

Ibelings, B. W., M. Vonk, H. F. Los, D. T. van der Molen, and W. M. Mooij. 2003. Fuzzy modeling of cyanobacterial surface waterblooms: validation with NOAA-AVHRR satellite images. Ecological Applications 13:1456–1472.

Joss, B. N., R. J. Hall, D. M. Sidders, and T. J. Keddy. 2008. Fuzzy-logic modeling of land suitability for hybrid poplar across the Prairie Provinces of Canada. Environmental Monitoring and Assessment 141:79–96.

Karr, J. R., K. D. Fausch, P. L. Angermeier, P. R. Yant, and I. J. Schlosser. 1986. Assessing biological integrity in running waters: a method and its rationale. Illinois Natural History Survey Special Publication 5:23.

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

Klir, G. J. 2004. Fuzzy logic: a specialized tutorial. Pages 11–61 *in* R. V. Demicco and G. J. Klir (editors). Fuzzy logic in geology. Elsevier Academic Press, San Diego, California.

Kopf, R. K., C. M. Finlayson, P. Humphries, N. C. Sims, and S. Hladyz. 2015. Anthropocene baselines: assessing change and managing biodiversity in human-dominated aquatic ecosystems. BioScience 65:798–811.

Lyons, J. 1992. Using the index of biological integrity (IBI) to measure environmental quality in warmwater streams of Wisconsin. General Technical Report NC-149. North Central Experiment Station, US Department of Agriculture Forest Service, St Paul, Minnesota.

MBI (Midwest Biodiversity Institute). 2015. Refining state water quality monitoring programs and aquatic life uses: evaluation of the Minnesota PCA Bioassessment Program. MBI technical memorandum. Midwest Biodiversity Institute, Columbus, Ohio. (Available from: https://www.pca.state.mn.us/)

Merritt, R. W., K. W. Cummins, and M. B. Berg (editors). 2008. An introduction to the aquatic insects of North America. 4th edition. Kendall/Hunt, Dubuque, Iowa.

MPCA (Minnesota Pollution Control Agency). 2014a. Development of biological criteria for tiered aquatic life uses: fish and macroinvertebrate thresholds for attainment of aquatic life use goals in Minnesota streams and rivers. Minnesota Pollution Control Agency, Environmental Analysis and Outcomes Division, St Paul, Minnesota. (Available from: http://www.pca.state.mn.us/)

MPCA (Minnesota Pollution Control Agency). 2014b. Development of fish indices of biological integrity (FIBI) for Minnesota rivers and streams. Minnesota Pollution Control Agency, St. Paul, Minnesota. (Available from: http://www.pca.state.mn.us/)

MPCA (Minnesota Pollution Control Agency). 2014c. Development of macroinvertebrate indices of biological integrity (MIBI) for Minnesota streams. Minnesota Pollution Control Agency, St Paul, Minnesota. (Available from: http://www.24,24.pca.state.mn.us/)

MPCA (Minnesota Pollution Control Agency). 2014d. Fish community sampling protocol for stream monitoring sites. Minnesota Pollution Control Agency, St Paul, Minnesota. (Available from: http://www.pca.state.mn.us/)

MPCA (Minnesota Pollution Control Agency). 2014e. Invertebrate community sampling protocol for stream monitoring sites. Minnesota Pollution Control Agency, St Paul, Minnesota. (Available at: http://www.pca.state.mn.us/)

MPCA (Minnesota Pollution Control Agency). 2016. Guidance for calculating the MPCA's human disturbance score. Minnesota Pollution Control Agency, St Paul, Minnesota. (Available from: http://www.pca.state.mn.us/)

Nair, R., R. Aggarwal, and D. Khanna. 2011. Methods of formal consensus in classification/diagnostic criteria and guideline development. Seminars in Arthritis and Rheumatism 41:95–105.

Niemela, S., E. Pearson, T. P. Simon, R. M. Goldstein, and P. A. Bailey. 1999. Development of an index of biotic integrity for the species-depauperate Lake Agassiz Plain ecoregion, North Dakota and Minnesota. Pages 339–365 *in* T. P. Simon (editor). Assessing the sustainability and biological integrity of water resources using fish communities. CRC Press, Boca Raton, Florida.

Odum, E. P., J. T. Finn, and E. H. Franz. 1979. Perturbation theory and the subsidy–stress gradient. BioScience 29:349–352.

Pauly, D. 1995. Anecdotes and the shifting baseline syndrome of fisheries. Trends in Ecology and Evolution 10:430.

Pont, D., R. M. Hughes, T. R. Whittier, and S. Schmutz. 2009. A predictive index of biotic integrity model for aquatic-vertebrate assemblages of Western U.S. streams. Transactions of the American Fisheries Society 138:292–305.

Santavy, D. L., P. Bradley, J. Gerritsen, and L. Oliver. The Biological Condition Gradient, a tool used for describing the condition of US coral reef ecosystems. Coral Reefs (*in press*).

Silvert, W. 2000. Fuzzy indices of environmental conditions. Ecological Modelling 130:111–119.

Simpson, J. C., and R. H. Norris. 2000. Biological assessment of river quality: development of AUSRIVAS models and outputs. Pages 125–142 *in* J. F. Wright, D. W. Sutcliffe, and M. T. Furse (editors). Assessing the biological quality of fresh waters. Freshwater Biological Association, Ambleside, UK.

Stamp, J., and J. Gerritsen. 2012. A biological condition gradient (BCG) assessment model for stream fish communities of Connecticut. Prepared by TetraTech, Owings Mills, Maryland for US Environmental Protection Agency Office of Science and Technology and Connecticut Department of Energy and Environmental Protection. (Available at: http://www.ct.gov/deep/lib/deep/water/water_quality_management/monitoring/ct_fishreport.pdf)

Stoddard, J. L., A. T. Herlihy, D. V. Peck, R. M. Hughes, T. R. Whittier, and E. Tarquinio. 2008. A process for creating multimetric indices for large-scale aquatic surveys. Journal of the North American Benthological Society 27:878–891.

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.

Suter, G. W. 1993. A critique of ecosystem health concepts and indices. Environmental Toxicology and Chemistry 12:1533–1539.

Teixeira, H., Á. Borja, S. B. Weisberg, J. A. Ranasinghe, D. B. Cadien, D. M. Dauer, J. Dauvin, S. Degraer, R. J. Diaz, A.

Grémare, I. Karakassis, R. J. Llansó, L. L. Lovell, J. C. Marques, D. E. Montagne, A. Occhipinti-Ambrogi, R. Rosenberg, R. Sardá, L. C. Schaffner, and R. G. Velarde. 2010. Assessing coastal benthic macrofauna community condition using best professional judgement—developing consensus across North America and Europe. Marine Pollution Bulletin 60:589–600.

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

USEPA (US Environmental Protection Agency). 2013. Biological assessment program review: assessing level of technical rigor to support water quality management. EPA 820-R-13-001. Office of Science and Technology, US Environmental Protection Agency, Washington, DC.

USEPA (US Environmental Protection Agency). 2016. A practitioner's guide to the Biological Condition Gradient: a framework to describe incremental change in aquatic ecosystems. EPA 820-R-13-001. US Environmental Protection Agency, Washington, DC.

Van Sickle, J. 1998. Documentation for MEANSIM, version 6.0. Western Ecology Division US Environmental Protection Agency, Corvallis, Oregon.

Van Sickle, J., and R. M. Hughes. 2000. Classification strengths of ecoregions, catchments, and geographic clusters for aquatic vertebrates in Oregon. Journal of the North American Benthological Society 19:370–384.

Weisberg, S. B., B. Thompson, J. A. Ranasinghe, D. E. Montagne, D. B. Cadien, D. M. Dauer, D. Diener, J. Oliver, D. J. Reish, R. G. Velarde, and J. Q. Word. 2008. The level of agreement among experts applying best professional judgment to assess the condition of benthic infaunal communities. Ecological Indicators 8:389–394.

Whittier, T., R. Hughes, J. Stoddard, G. Lomnicky, D. Peck, and A. Herlihy. 2007. A structured approach for developing indices of biotic integrity: three examples from streams and rivers in the western USA. Transactions of the American Fisheries Society 136:718–735.

Wright, J. F. 2000. An introduction to RIVPACS. Pages 1–24 *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.

Yoder, C. O., and E. T. Rankin. 1995. Biological criteria program development and implementation in Ohio. Pages 109–144 *in* W. S. Davis and T. P. Simon (editors). Biological assessment and criteria: tools for water resource planning and decision making. Lewis Publishers, Boca Raton, Florida.

Zadeh, L. A. 1965. Fuzzy sets. Information and Control 8:338–353.

Zadeh, L. A. 2008. Is there a need for fuzzy logic? Information Sciences 178:2751–2779.