Ecological Indicators 10 (2010) 980–989

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

# Defining ecological thresholds to determine class boundaries in a bioassessment tool: The case of the Eastern Canadian Diatom Index (IDEC)

Martine Grenier[a,*], Isabelle Lavoie[a], Alain N. Rousseau[a], Stéphane Campeau[b]

[a] Institut national de la recherche scientifique, Centre Eau, Terre et Environnement, 490 de la Couronne, Québec, Québec, Canada, G1K 9A9
[b] Université du Québec à Trois-Rivières, 3351 boul. des Forges, Trois-Rivières, Québec, Canada, G9A 5H7

ARTICLE INFO

*Article history:*
Received 12 June 2009
Received in revised form 26 February 2010
Accepted 1 March 2010

*Keywords:*
Diatoms
Biological assessment
Biological integrity
Biotic index
Biotype
Ecological classes

ABSTRACT

Most bioassessment tools are based on the Reference Condition Approach (RCA), where the biological integrity of a site is defined by the "distance" between current conditions and its reference condition status. Among them, the Eastern Canadian Diatom Index (IDEC) was developed to evaluate the ecological integrity of streams along an alteration gradient, as a function of the dissimilarity of their diatom community from their suitable reference communities. In the first version of the IDEC, the alteration gradient was arbitrarily divided, like most traditional approaches, into five classes of equal size representing the qualitative statuses of the site. In this article, we propose a remodeling of those classes by introducing ecologically meaningful thresholds, reducing the subjectivity in the determination of the number and the range (widths) of classes. We developed a new approach which uses biotypes issued from a classification technique (Self-Organizing Maps) to determine thresholds between integrity classes of the IDEC. These biotypes represent relatively homogenous ecological entities composed of taxa adapted to a particular biological integrity status. Based on these biotypes, four classes were established. The new limits between classes are now based on the maximum ecological distance between diatom groups, and the transition from one class to another may be related to major ecological thresholds that induce important shifts in community structure. The proposed biotype-based approach allows for the identification of ecologically meaningful differences among biological communities and provides a more relevant interpretation of the community changes along the alteration gradient.



## 1. Introduction

Over the last decades, numerous bioassessment tools have been developed to evaluate the biological integrity of aquatic ecosystems and to determine the effect of a wide range of stressors on this integrity. Most of these tools are based on the Reference Condition Approach (RCA), where the biological integrity of a site is defined by the "distance" between current conditions and reference condition status. The indices are usually expressed as a value along an alteration gradient (e.g., 0–100) or as an overall statistical difference from reference conditions. To facilitate the interpretation of the biological "distance" of a site from its reference status, the alteration gradient is often divided into classes reflecting qualitative levels of biological integrity. These approaches allow for a rapid and easy overall picture of the ecosystem status and are particularly interesting for water-quality managers interacting with the general public. Although the creation of biological integrity classes is useful, the approaches used to define the number of classes and the limits between these classes are arbitrary and lack meaningful biological considerations. Indeed, the number of classes is usually arbitrarily determined with widths of either equal proportion (e.g., Lavoie et al., 2006a; Ponader et al., 2007) or unequal proportion (e.g., Karr, 1981; Coste et al., 2004), often with boundaries subjectively set to certain percentiles or probabilities (e.g., Wright et al., 1993; Reynoldson et al., 1995; Parsons and Norris, 1996; Bowman et al., 2006).

RCA-based tools can be divided into two categories according to the alteration gradient covered, "probability-tools" and "gradient-tools". "Probability-tools", such as RIVPACS (Wright et al., 1993), AusRivAS (Parsons and Norris, 1996), BEAST (Reynoldson et al., 1995) and TSA (Bowman et al., 2003) use potential reference samples for biological status evaluation. The level of alteration is based on the probability that a test site falls within the normal range of variation of the references sites. A test site exceeding the normal range of variation is considered as an altered site. In RIVPACS and AusRivAS, the assessment of a test site is based on the overall ratio of observed-to-expected taxa (i.e., O/E score). Sites that fall within the central 80% of reference O/E values about the mean (i.e., 10th to 90th percentiles) are considered equivalent to the reference. Five classes of different widths were defined depending on the

* Corresponding author.
*E-mail address:* martine.grenier@uqtr.ca (M. Grenier).

1470-160X/$ – see front matter 
doi:10.1016/j.ecolind.2010.03.003

variability of the reference O/E values. BEAST uses an ordination to evaluate the distance between a community and the reference communities (Rosenberg et al., 1999). Four classes of different widths were determined according to the normal range of variation, enclosing 90%, 99% and 99.9% of the reference sites (Reynoldson and Wright, 2000).

"Gradient-tools", such as the IBI (Karr, 1981), IPS (Coste, 1982), IBD (Lenoir and Coste, 1996) and IDEC (Lavoie et al., 2006a), were developed using samples covering the entire alteration gradient of their respective study regions. Samples composed of the most different biological communities, compared to reference communities, determine the maximum alteration level. Difference from reference communities and similarity with altered communities are used to assess the degree of alteration of a test site. The original version of the IBI (Karr, 1981) consisted of 12 metrics that reflected different fish attributes (e.g., species richness, abundance of indicator species, trophic organization and function). The sum of the 12 metrics yielded an IBI score which ranged from 60 (best integrity) to 12 (worst integrity), and five unequal classes were determined. Various versions have been developed for different regions, ecosystems, and organisms (e.g., invertebrates, diatoms); each version differing from the original version in terms of number, identity, and scoring of metrics. New versions of the IPS and the IBD now account for the natural variability in diatom reference communities and the water-quality classes have been modified accordingly (Coste et al., 2004; Tison et al., 2007). Five unequal classes were determined. The median index value for the reference sites is used as the value representing the reference conditions, while the value of the 75% percentile of reference sites is used as the limit between the very good/good biological statuses. The limits for the intermediate classes are calculated by dividing the range in index values between the very good/good ecological statuses and the lowest value by four.

Lavoie et al. (2006a,b, 2008a,b,c) developed a diatom-based index in Eastern Canada that is highly sensitive to eutrophication, mineral pollution, and organic enrichment. The IDEC indicates the distance of an altered diatom community from its specific reference biotype. Reference biotypes are used to separate the effect of natural variations from human-related changes. Grenier et al. (2006) showed that pH was the main discriminating factor for diatom reference communities and two main biotypes were identified. Considering these results, the IDEC was divided into two sub-indices: one for sites corresponding to a reference biotype typical of naturally alkaline waters (IDEC-alkaline), and one for sites corresponding to a reference biotype typical of naturally acidic/neutral waters (IDEC-neutral). Modeling techniques (random forests and artificial neural networks) were used to assign each test site to the appropriate sub-index (Grenier et al., submitted for publication). Each sub-index was developed based on a correspondence analysis (CA) (Lavoie et al., 2006a). The position of each site along its respective CA alteration gradient ($x$-coordinate) was rescaled to range between 0 and 100. A high index value characterizes reference or less-impacted sites, while a low index value represents more heavily impacted sites. In this version of the IDEC, the alteration gradient (CAs first axis) was arbitrarily divided into five classes of equal size representing the qualitative status of the site.

The main purpose of the present study was to develop an approach that considers ecologically meaningful thresholds between classes, thus eliminating the subjectivity in the determination of the number and the range (widths) of classes. This new approach uses biotypes issued from a classification technique (Self-Organizing Maps) to determine thresholds between the biological integrity classes of the IDEC. These biotypes represent relatively homogenous ecological entities composed of taxa adapted to a particular biological integrity status, and the differences between biotypes represent the dissimilarity thresholds.

## 2. Materials and methods

### 2.1. Diatom sampling and water analyses

For the development of the IDEC, 204 diatom samples were collected in 2002 and 2003 at 154 sampling locations distributed throughout 34 rivers and streams in the St. Lawrence River Basin (Fig. 1). The sampling sites were distributed within three ecoregions of southern Québec (Canada): the Canadian Shield, the St. Lawrence Lowlands and the Appalachians; enabling the inclusion of a wide range of watershed characteristics, pollution types and alteration levels. Grenier et al. (2006) and Lavoie et al. (2006a) presented a detailed description of the study area. 65 samples collected in 2005 and 2006 were added to this upgraded version of the IDEC. Benthic diatom communities were sampled and analysed as described in Lavoie et al. (2006a). Following the recommendation of Lavoie et al. (2008b), a taxon was included in the analyses if it was present with a relative abundance of $\geq$2% in at least one sample. The 269 samples available for this study were divided into the following two groups: 74 samples for the neutral sub-index and 195 samples for the alkaline sub-index (Grenier et al., submitted for publication).

Water analyses for most of the sites were performed by the Québec Ministry of Sustainable Development, Parks and Environment (MDDEP) as part of their water-quality monitoring program. Water samples were collected every four weeks. Additional sites not included in the MDDEP's monitoring program were selected for this study and water analyses were performed by the 'Institut national de la recherche scientifique' (INRS). The parameters considered in this study included: total phosphorus (TP), total dissolved nitrogen (TN), ammonia-nitrogen ($NH_3$), pH, conductivity (CON), temperature (TEMP), turbidity (TUR), and dissolved organic carbon (DOC).

### 2.2. Ordinations

Two CAs were used to develop each sub-index. Diatom samples were represented on CAs to indirectly infer environmental gradients from diatom community data. Canonical correspondence analyses (CCAs) were performed a posteriori to evaluate which environmental variables were the most important in structuring the diatom communities. For each CCA, variables with a variance inflation factor (VIF) exceeding 10 were excluded because of their correlation with other variables. Monte-Carlo permutation tests were used to select variables explaining a significant portion of the variance ($p \leq 0.01$). Further information on IDEC's development can be found in Lavoie et al. (2006a). All ordination analyses were performed using CANOCO (version 4.5, Center for Biometry, Wageningen; ter Braak and Smilauer, 2002).

### 2.3. Biotype classification

For each sub-index, biotypes were determined using Self-Organizing Maps (SOMs; Kohonen, 1982). A Kohonen SOM was derived using neural networks with unsupervised learning rules. A SOM performed a topology-preserving projection of the data space onto a regular two-dimensional space. The data were projected onto a rectangular grid map containing multiple hexagonal cells. Sites with similar characteristics were mapped in the same vicinity depending on the distance measure metric used. Classification of the SOM cells was performed using the Bray–Curtis distance measure (McCune and Mefford, 1999). A formula was used to calculate the number of cells needed to map the sites ($5 \times (\sqrt{}$ number of