Incorporating natural variability in the bioassessment of stream condition in the Atlantic Forest biome, Brazil

Priscilla S. Pereira a, Natália F. Souza a, Darcilio F. Baptista a, Jaime L.M. Oliveira b, Daniel F. Buss a,∗

a Fundação Oswaldo Cruz, Instituto Oswaldo Cruz, Laboratório de Avaliação e Promoção da Saúde Ambiental (LAPSA/IOC/FIOCRUZ), Av. Brasil 4365, Manguinhos, Pavilhão Laura Travassos, Rio de Janeiro, RJ CEP 21040-360, Brazil
b Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública Sérgio Arouca (DSSA/ENS/FIOCRUZ), Rua Leopoldo Bulhões 1480, Manguinhos, Rio de Janeiro, RJ CEP 21041-210, Brazil

A R T I C L E   I N F O

Article history:
Received 24 August 2015
Received in revised form 13 May 2016
Accepted 19 May 2016
Available online 28 May 2016

Keywords:
Multimetric index
Biological monitoring
Bioassessment protocol
Aquatic insects
Neotropical streams

A B S T R A C T

Most bioassessment programs in Brazil face difficulties when scaling up from small spatial scales because larger scales usually encompass great environmental variability. Covariance of anthropogenic pressures with natural environmental gradients can be a confounding factor in the evaluation of biologic responses to anthropogenic pressures. The objective of this study was to develop a multimetric index (MMI) with macroinvertebrates for two stream types and two ecoregions in the Atlantic Forest biome in Rio de Janeiro state, Brazil. We hypothesized that by using two approaches – (1) testing and adjusting metrics to landscape parameters, and (2) selecting metrics using a cluster analysis to avoid metrics redundancy – the final MMI would perform better than the traditional approach (unadjusted metrics, one metric representing each category). Four MMIs were thus developed: MMI-1 – adjusted MMI with metrics selected after cluster analysis; MMI-2 – adjusted MMI with one metric from each category; MMI-3 – unadjusted MMI with metrics selected after cluster analysis; MMI-4 – unadjusted MMI with one metric from each category. We used three decision criteria to assess MMI’s performance: precision, responsiveness and sensitivity. In addition, we tested the MMIs by using an independent set of sites to validate the results. Although all MMIs performed well in the three criteria, adjusting metrics to natural variation increased MMI response and sensitivity to impairment. In addition, the selected MMI-2 was able to classify sites of two stream types and two ecoregions. The use of cluster analysis, however, did not avoid high redundancy between metrics of different branches. The MMI-4 had the poorest performance among all tested MMIs and it was not able to distinguish adequately reference and impaired sites from both ecoregions. We present some considerations on the use of metrics and on the development of MMIs in Brazil and elsewhere.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Multimetric indices (MMIs) with macroinvertebrates are the most widely used approach for biological assessment and monitoring of aquatic ecosystems. This approach has been widely accepted because: (a) it is based on ecological concepts and processes, (b) it has the potential to assess ecological functions, and (c) it can discriminate human impacts (Helson and Williams, 2013; Mereta et al., 2013). However, one difficulty many bioassessment programs in Brazil face is scaling up from small spatial scales – from where most indices are developed (e.g. a river basin or hydrographic region) – to larger scales. Buss et al. (2015), analyzing 13 large-scale macroinvertebrate protocols from around the world discussed that limitations for scaling up may be associated with lack of logistics and funding, a reluctance to change established techniques or gear, or the fact that locally developed methods sometimes yield more accurate results than regionally applicable ones. Despite these difficulties, many countries succeeded in building widely applicable monitoring programs by using the same and/or compatible sampling protocols and by selecting metrics that are sensitive on large-scales (e.g. Stoddard et al., 2008; Moya et al., 2011; Jun et al., 2012).

Large-scale studies, however, may encompass great environmental variability. One approach to describe and account for variability in ecological studies is to classify areas in ecoregions. Ecoregions are usually defined as relatively homogeneous areas that have similar environmental conditions (Omernik, 1995). Ecoregions can be defined at different spatial scales, and aim to
serve as a territory for investigation, assessment, management and monitoring of ecosystems, including the development of biological criteria and water quality standards (Kong et al., 2013). Some authors have found relationships between the macroinvertebrate fauna and environmental conditions at the ecoregional level (Barbour et al., 1996; Reynolds et al., 1997). Other authors, however, have found macroinvertebrate fauna to be more strongly associated with micro/local scales, such as substrate and micro-habitat (Gerth and Herlihy, 2006; Costa and Melo, 2008; Ligeiro et al., 2013). Acknowledging that, some countries incorporate local characteristics (stream types) in different biomonitoring protocols (AQEM, 2002; Buss et al., 2015). For example, New Zealand has different protocols for hard-bottomed and soft-bottomed streams (Stark et al., 2001). The authors argue that this separation is necessary because of significant differences in the structure and composition of aquatic communities among stream types, and thus different methods are required for sample collection and processing to be cost-effective. The definition of stream types allows the correct establishment of reference conditions that are comparable to the ecological status classifications within each group of rivers with similar characteristics (Stark et al., 2001).

Covariance of anthropogenenic pressures with natural environmental gradients can be a confounding factor in the evaluation of biologic responses to anthropogenic pressures (Stoddard et al., 2008; Hawkins et al., 2010; Moya et al., 2011). One simple technique for normalizing metrics for natural gradients is to remove the stressor gradient from the data by focusing on reference-site data and to quantify the remaining correspondence between the metric value and the natural gradient (Stoddard et al., 2008). Recent studies aimed to test this and other alternative approaches for the development of MMIs (e.g. Chen et al., 2014; Macedo et al., 2016). The objective of this study was to develop a MMI for the Atlantic Forest biome in Rio de Janeiro state, Brazil. We hypothesized that by using two of these approaches – (1) testing and adjusting metrics to landscape parameters, and (2) selecting metrics using a cluster analysis to avoid metrics redundancy – the final MMI would perform better than the traditional approach (unadjusted metrics, one metric representing each category).

2. Materials and methods

2.1. Study area

The geomorphology of Rio de Janeiro state is composed of a group of coastal plains separated by hills and two mountain chains that run parallel to the ocean (Serra do Mar, ranging from altitudes 0–2000 m a.s.l and Serra da Mantiqueira, ranging from 800 to 2500 m a.s.l.). In between the two mountain chains, lies the valley of the state’s main river, Paraiba do Sul (at an altitude of around 800 m a.s.l.). According to Alvares et al. (2013) 44% of Rio de Janeiro state’s mid-lower portions is classified as tropical with a summer rainy season, with the most mountainous regions and plateaus classified as humid tropical with hot summer, without dry season or with a dry winter. Temperatures oscillate between 15 °C and 28 °C and annual rainfall is around 1000–1500 mm. The Atlantic Forest biome, which originally covered virtually the entire region, now represents less than 12% of its original extent, and is mostly spread in the higher parts of the mountains and in remnants interspersed with agriculture and pasture (Ribeiro et al., 2011).

2.2. Sampled sites

We sampled 73 sites (once, during the dry season, in streams ranging from 1st to 4th order, according to Strahler classification using 1:50,000 scale maps) representing two stream types, two ecoregions and three classes of impairment. Ecoregions were based on the classification of ‘dominions’ of RadamBrasil (Brasil, 1983). Stream types were classified as “transitional/sedimentary areas” (stream type 1) and “rocky substrates” (stream type 2; see below for details). We chose sites based on ad hoc indication and/or by previous knowledge of the area to represent sites classified as reference, intermediate or impaired. Sites classified as “reference” had “Optimal” or “Good” environmental condition according to the Habitat Assessment Protocol (HAP; see below for logic and measurement), absence of channelization and <25% upstream urban or industrial areas. Impaired stream reaches were classified as “Poor” condition according to HAP and >40% of upstream area affected by urban areas or intense farming or livestock grazing. Intermediate sites had characteristics between these two classes.

We sampled forty-nine sites in the sedimentary deposits ecoregion (SD). The SD ecoregion is located at the piedmont of Serra do Mar mountain range, with altitudes about 200 m a.s.l. and is a depositional zone formed by marine, lacustrine and fluvi-sedimentation processes (Brasil, 1983). Being a transitional zone between erosive/depositional zones, sampled streams were divided into two predominant substrate types: reaches with >80% sand and clay (“transitional/sedimentary areas”; stream type 1) and reaches with >70% particles of gravel size or greater (“rocky substrates”; stream type 2). In this region, land use is dominated by patches of small-scale agriculture and livestock grazing, and minimally impacted areas are scarce. Reference areas in this ecoregion were classified as “least disturbed areas”, according to the reference condition approach (RCA; Stoddard et al., 2006). Twenty-two sites of stream type 1 were sampled; of which six were reference, six intermediates and ten impaired. Twenty-seven sites of stream type 2 were sampled, of which fifteen were reference, three intermediate and nine impaired (Fig. 1).

We sampled twenty-four sites in the mountainous scarpas ecoregion (MS). This ecoregion is located at higher altitude (from >200 m a.s.l. to around 1,800 m a.s.l.) in a mountainous region with high slope and steep scarpas. Streams in MS have > 80% predominance of rocky substrates (stream type 2) – bedrocks in some reaches – with few patches of sand and formation of pools intertwined with riffles or runs. All sites were sampled within or near protected areas (conservation units), which were classified as “minimally disturbed areas” or “best attainable” (RCA; Stoddard et al., 2006). The latter occurred in areas outside conservation units, but had low to moderate impact by rural activities, and full or partial riparian vegetation and forest fragments.

2.3. Environmental and biological data

We sampled macroinvertebrates using a kick-net sampler with mesh size of 500 μm. Twenty samples (around 20 m²) were taken proportional to the substrates available in each site, following the multi-habitat method (Barbour et al., 1998). The percentage of available habitats was estimated by visual inspection. Substrates with less than 5% of the site area were not sampled. Samples were obtained from a site length of approximately 20 times the channel width. Samples were composited and conserved in the field in 80% ethanol and taken to the laboratory for further inspection. In the laboratory, samples were washed to remove coarse organic matter, such as leaves and twigs. The remaining material was placed in a sub-sampler (64 × 36 cm), divided into 24 quadrats, each measuring 10.5 × 8.5 cm (Patent application number PCT/BR2011/000144). Sub-sampling is a procedure widely used in formulating multitemetric indices, to assure randomness of the procedure, making it less subject to inherent variations from changes in team members (Oliveira et al., 2011). Eight quadrats were chosen at random, following the procedures described in Oliveira...
et al. (2011). Two sites from stream type 1 (one reference and one impaired) were excluded from the analyses because fewer than 150 individuals were found in the entire sample, following Carter and Resh (2001). Specimens were identified to the lowest taxonomic level possible (mostly genus) – except for Hemiptera and Diptera, which were identified to family level, and Annelida to class level.

In each sampling occasion, physicochemical and environmental parameters were recorded in the field. Dissolved oxygen (DO) was determined using a YSI 550A analyzer, pH with a MPA 210p (LabConte) and conductivity and total dissolved solids by using MCA 150p (LabConte). Water samples were preserved in sterile plastic bags (whirl-pak) according to APHA (2000). In the laboratory, ammonia was determined by using a HACH (DR 2500), chloride and total alkalinity by the titrimetric method following APHA (2000). Sampling sites were classified using the visual-based habitat assessment protocol (HAP; Barbour et al., 1999) for low-gradient (ecoregion SD) and high-gradient (MS ecoregion). The HAP analyzes ten environmental parameters, such as substrate availability for colonization by benthic fauna, water velocity and embeddedness (pool variability for low-gradient streams), channel condition (sinuosity for low-gradient streams), sediment deposition, margin stability and riparian vegetation. For each parameter, a score between 0 and 20 is assigned. Sites are classified according to the mean score obtained in the HAP, as follows: 0–5 “Poor”, 5.1–9.9 “Marginal”, 10–14.9 “Suboptimal” and 15–20 an “Optimal” environmental condition (Barbour et al., 1999).

2.4. MMI development and validation

2.4.1. Metric selection

We examined 42 benthic macroinvertebrate assemblage metrics commonly used in biomonitoring protocols (e.g. Barbour et al., 1999; Baptista et al., 2007; Henson and Williams, 2013). These metrics comprise five categories, based on their logic (according to Barbour et al. (1999) and Stoddard et al. (2008)):

1) Richness – express the number of taxa of the whole macroinvertebrate assemblage or of its subgroups, in family or genus level.

2) Diversity/evenness – diversity indices include information on the taxonomic richness and the relative abundances of taxonomic groups. This category includes diversity indices such as Shannon-Wiener, Margalef and Dominance (1-Simpson index), as well as equitability/evenness indices.

3) Composition – measures of the relative abundance of selected groups, in taxonomic level of order or lower.

4) Tolerance – metrics that consider the sensitivity and tolerance of organisms to impairment. This category includes biotic indices and information on other sensitive or tolerant indicator groups. We included indices developed for Brazil such as the Biological Monitoring Working Party (BMWP-CETEC; Junqueira and Campos, 1998) and the Indice Biotico Estendido (IBE-IOC; Mugnai et al., 2008).
5) Trophic – encompass functional feeding groups and provide information on the balance of feeding strategies (food acquisition and morphology) in the benthic assemblage and their trophic roles (shredders, collectors, scrapers, filterers or predators).

The first step for metric selection was to evaluate if metrics met all these four criteria: (a) present similar results for reference areas from typologies 1 and 2 using Mann-Whitney U tests. Our intention was to have a single index for both typologies because often the two typologies are mixed in a site and it is difficult to determine which is predominant. (b) Discriminate impairment conditions (reference, intermediate and impaired) according to Kruskal-Wallis test. (c) Have a linear response to impairment – i.e. intermediate sites should present intermediate values between those found for reference and impaired classes. (d) Do not present a narrow range of values (e.g., richness metrics based on only a few taxa) or that have similar values at most sites (e.g., most sites have values = 0), or many outliers (following Stoddard et al., 2008; Nelson and Williams, 2013), analyzed after inspection of Box-and-Whisker plots.

With the approved metrics, we developed and tested four MMIs in order to evaluate the potentially confounding effects of natural variation into biological response to anthropogenic impacts, and to reduce the effects of redundancy between metrics. For the first target, we used four GIS-extracted environmental variables (based on Chen et al. (2014) and Macedo et al. (2016)). Two parameters were related to climate, available from WorldClim database (http://www.worldclim.org): seasonality of temperature (i.e. standard deviation × 100), coded Bio10 and annual precipitation, coded Bio12. The two others were related to topography, available from AMBDATA (http://www.dpi.inpe.br/ambdata): altitude and slope.

Both geographic information systems had spatial resolution of 30 arc-seconds (~1 km). After verifying and correcting for normal distribution according to Shapiro-Wilk test, metrics were regressed against environmental variables by stepwise (forward) multiple linear regression procedures and the Akaike information criteria to build the simplest possible model that adequately explained the candidate metric (Moya et al., 2007; Chen et al., 2014). If natural environmental gradients were significantly associated (p < 0.05) with the variation in metric values, the residuals of each model (residual = observed value – predicted value) were used as metric values. MMI-1 and MMI-2 used the residuals of metrics (hereafter referred as “adjusted” metrics and MMIs). MMI-3 and MMI-4 used raw metric values (hereafter “unadjusted” metrics and MMIs). For the second target, we measured metric redundancies by means of a cluster analysis using Pearson’s r correlation coefficients as the similarity measure and Ward linkage as the clustering method (Cao et al., 2007; Chen et al., 2014). Metrics in a cluster branch are potentially redundant (Stoddard et al., 2008). In order to decide which metric from each cluster should be retained for the MMI we calculated discrimination efficiency (DE) and t-tests between reference and impaired sites (Barbour et al., 1999; Stoddard et al., 2008). For metrics expected to decrease at impaired sites, DE was calculated as the percentage of impaired sites with a metric value <25th percentile of the reference site values. For metrics expected to increase at impaired sites, DE was calculated as the percentage of impaired sites with a metric value >75th percentile of the reference site values. The metric with the highest DE and significant t-test scores in each cluster was retained for the MMI. This approach was used for MMI-1 and MMI-3. For MMI-2 and MMI-4, we calculated Pearson correlations between metrics of the same category (richness, diversity/evenness, composition, tolerance and trophic). Metrics were considered redundant if r > 0.70 (Stoddard et al., 2008). We then retained the metrics with the greater power to discriminate reference and impaired conditions according to t-tests, and if similar, the one with lower cost and/or easier to calculate. As a summary:

- **MMI-1** – adjusted MMI with metrics selected after cluster analysis;
- **MMI-2** – adjusted MMI with one metric from each category;
- **MMI-3** – unadjusted MMI with metrics selected after cluster analysis;
- **MMI-4** – unadjusted MMI with one metric from each category.

### 2.4.2. MMI standardization

The selected metrics were included in the four MMIs, using a continuous system, following the recommendation of Blockson (2003). For positive raw metrics (values that increase as the environment improves), Eq. (1) was used, and for negative raw metrics (values that decrease as the environment improves), Eq. (2) was used, as follows:

\[
\text{Metric} = \frac{25\% \text{ impaired} - 75\% \text{ reference}}{25\% \text{ reference} - 25\% \text{ impaired}} \times 10
\]

\[
\text{Metric} = \frac{75\% \text{ impaired} - 25\% \text{ impaired}}{25\% \text{ reference} - 75\% \text{ impaired}} \times 10
\]

where 25% impaired is the 25th percentile of impaired sites, 75% reference is the 75th percentile of reference sites, 75% impaired is the 75th percentile of impaired sites, and 25% reference is the 25th percentile of reference sites.

Negative values were assigned a value of zero and those above 10 were assigned a value of 10. The sum of all metrics scores was then adjusted to fit a scale from 0 to 100, which was used as the final MMI score. The MMI score was used to define the class of biological integrity, according to the arbitrary scale: severely impaired (0–19), impaired (20–39), regular (40–59), good (60–79), excellent (80–100).

### 2.4.3. MMI evaluation

Three criteria were used to evaluate the performance of the MMIs: precision, responsiveness and sensitivity to anthropogenic pressures (Chen et al., 2014; Macedo et al., 2016). For the precision, we calculated the standard deviation of each MMI for the reference sites. Lower standard deviation meant higher precision. The responsiveness was tested with t-tests to verify if each MMI discriminated the pre-determined impairment conditions. To evaluate the sensitivity, we ran stepwise-forward regression between each MMI and anthropogenic-related variables (pH, dissolved oxygen, conductivity, total dissolved solids, chloride and ammonia).

### 2.4.4. MMI validation

Validation tests consisted on evaluating MMI’s ability to correctly classify independent sites (not used to develop the index) and separate reference and impaired conditions. For that, we used 24 reference sites from the MS ecoregion (stream type 2), and nine intermittently impaired sites from SD ecoregion (six of stream type 1 and three of stream type 2). In addition, two Principal Component Analysis (PCA) were performed – one for each set of sites (the 38 sites used to develop and the 33 sites used to validate the index) – with environmental variables related to impairment (total dissolved solids, conductivity, alkalinity, ammonia and HAP). Prior to analysis, data were standardized by subtracting each value from its mean and dividing it by its standard deviation to reduce the effects of different scales used in the variables. The values for PCA axis 1 were correlated (Pearson) with each corresponding MMI value.

### 2.4.5. Data analysis

All statistical analyses were conducted in R (version 3.2.3; R Development Core Team, 2015, http://www.r-project.org/).
3. Results

3.1. Environmental and biological data

Significant differences between reference and impaired sites used to develop the MMIs were found for total dissolved solids, conductivity, alkalinity, ammonia and the HAP index (Table 1).

Regarding the biological samples, 44,892 individuals (78 families and 141 genera) were identified. From this total, 9083 individuals (65 families and 83 genera) were collected in transitional/sedimentary areas (stream type 1) and 19,088 (62 families and 97 genera) in rocky substrate (stream type 2), both in the SD ecoregion. A total of 16,721 (60 families and 101 genera) were collected in stream type 2 in the MS ecoregion.

3.2. MMI development and validation

All metrics, except 'Number of Ephemeroptera families' (Mann-Whitney U test, p < 0.05), were approved for not discriminating reference areas of stream types 1 and 2 (Table 2). Twelve metrics failed in discriminating the two classes of impairment (Table 2; Kruskal-Wallis tests, p < 0.05) and fourteen other metrics were excluded because they presented narrow ranges of values and/or had non-linear responses to the stress gradient (Table 2). Thus, fifteen metrics were approved for further screening, of which ten required identification to family level, two to genus level and three that required identification to order level (Table 2).

The approved metrics were regressed against environmental variables (altitude, Bio04, Bio12, slope). All metrics followed a normal distribution and no transformation was necessary. Only six of the fifteen metrics were significantly associated with environmental variables (four with temperature seasonality, Bio04, and two with annual precipitation, Bio12; Table 3). No metric was significantly associated with altitude or slope. The residuals of those six metrics (hereafter added a “re” after the name of the metric) and the unadjusted values for the other nine metrics were used in further analysis.

For MMI-1, the selected metrics were based on a cluster analyses (Fig. 2a) and the ones with the highest DE and t-test scores in each cluster were retained (Shanf, %EPTsBH, IBE-IOC_re, BMWP_re and %Ple). Only two of the five metrics had to be adjusted.

For MMI-2, the selected metrics for each category were the ones with the highest t-test scores to discriminate impairment classes. Two of the four metrics were adjusted: MarEp, EPTsBHf_re, %Ple and IBE-IOC_re.

For MMI-3, selected metrics were based on a cluster analyses (Fig. 2b) and we selected the ones with the highest DE and t-test scores in each cluster (Shanf, %Ple, BMWP, IBE-IOC and Domf). However, using only the reference sites, we found a high correlation between two metrics that were in different clusters (Domf × Shanf r = −0.91). The original MMI-3 with five metrics had precision of 12.11, responsiveness t-score of 11.04 (p < 0.0001) and sensitivity of r2 = 0.86 (p < 0.0001). First, we decided to substitute the second-best performance metric in the cluster, provided it had no high correlation (r > 0.70) with the ones already selected. However, when substituting Shanf by %EPT or Equif, or when substituting Domf by %Coele, it reduced MMI precision, responsiveness and sensitivity. We then tested excluding one of these metrics. Both MMIs had better performance than the original index, so we retained Shanf because it yielded slightly better responsiveness (higher t-score, 13.12 versus 13.03 with Domf), although with slightly lower precision (standard deviation, 10.85 versus 9.55), and sensitivity (r2 0.91 versus 0.92). The final MMI-3 with four metrics (%Ple, BMWP, IBE-IOC and Shanf) was used in further analysis (Table 4).

For the MMI-4, when metrics were correlated (r > 0.70), we selected the ones with the lowest p-value according to Kruskal-Wallis tests to separate the three impairment classes. In the richness category, EPTsBHf was selected over Trif. In the diversity category, Margalef at family and genus levels were selected over Shannon-Wiener index at family-level and Equitability index at family-level. In the composition category, %EPTsBH was selected over %EPT/Chi, %Ple, and %EPT. In the tolerance category, BMWP-CETEC (family level) and IBE-IOC (genus level) were selected over Dominance index in families and genus level, respectively. For the latter, we decided to retain both metrics because they have different logic and evaluate different aspects of macroinvertebrate assemblage. In addition, they were not highly correlated.

The four MIMIs performed similarly (Table 4). MMI-3 had better precision (lowest standard deviation in reference site scores) and the lowest mean correlation between metrics in reference sites. The MMI-2 had a slightly better response (highest t-score to discriminate impairment classes) and sensitivity (higher association with anthropogenic variables), but all MIMIs performed well (p < 0.0001) for both criteria. The MMI-1, which adopted all the hypothesis approaches, did not outperform the other MIMIs in any criteria, nor had the lowest mean correlation between metrics considering only the reference sites (Table 4).

Considering the responsiveness criteria, MMI-2 (Table 5) had the best performance, for both the development database and the validation databases (Table 4; Fig. 3). MMI-4 performed poorly when considering the validation database, with a very narrow range of values separating reference and impaired sites (Fig. 3). For MMI-2, metric ranges were similar and showed no significant differences between reference sites of the two stream types, and in both cases it was able to discriminate reference and impaired sites (Fig. 4). MMI-2 had a 76% correct classification of validation sites (25 of 610 valid sites).
Table 2
Candidate metrics to integrate the MMIs, predicted responses to impairment, evaluation if metric range was similar for both stream types (p > 0.05, Mann-Whitney test), if metric was able to discriminate between classes of impairment (p < 0.05, Kruskal-Wallis test), if metric had a linear response to impairment and final evaluation. EPTs-BH = Ephemeroptera, Plecoptera and Trichoptera excluding Baetidae and Hydropsychidae; MOLD = Mollusca + Diptera; IBE-IOC = Índice Biótico Estendido Instituto Oswaldo Cruz; BMWP-CETEC = adaptation of Biological Monitoring Working Party score system, Dominance = 1-Simpson index.

<table>
<thead>
<tr>
<th>Category</th>
<th>Metrics (codes for valid metrics)</th>
<th>Predicted response to impairment</th>
<th>Metric range is similar for both stream types?</th>
<th>Metric discriminate between classes of impairment?</th>
<th>Metric present a linear response to impairment?</th>
<th>Final evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diversity</td>
<td>Number of genera</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Number of families</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Number of Plecoptera genera</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Number of Ephemeroptera genera</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Number of Ephemeroptera families</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Number of Trichoptera genera</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Number of Trichoptera families (Trif)</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td>Number of EPT families (EPTF)</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td>Number of EPTsBH families (EPTsBHf)</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td>Shannon-Wiener index families (Shanf)</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td>Simpson index genera</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Simpson index families</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Evenness index genera</td>
<td>Decrease</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Evenness index families</td>
<td>Decrease</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Margalef index genera (Margg)</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td>Margalef index families (Margf)</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td>Equitability index genera</td>
<td>Decrease</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Equitability index families (Equif)</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td>Dominance index families (Domf)</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Valid</td>
</tr>
<tr>
<td>Composition</td>
<td>% Coleoptera (%Coleo)</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td>% Odonata</td>
<td>Variable</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>% Ephemeroptera</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>% Plecoptera (%Ple)</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td>% Trichoptera</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>% EPT</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td>% EPTsBH</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td>% EPT/Chironomidae (%EPT/%Ch)</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td>% Chironomidae</td>
<td>Increase</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>% Diptera</td>
<td>Increase</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>MOLD</td>
<td>Increase</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>% MOLD</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>Tolerance</td>
<td>IBE-IOC</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td>BMWP-CETEC</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td>Hydropsychidae/Trichoptera</td>
<td>Decrease</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Chironomidae</td>
<td>Increase</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Chironomidae/Diptera</td>
<td>Increase</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Baetidae/Ephemeroptera</td>
<td>Decrease</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Trophic</td>
<td>% Scrapers</td>
<td>Decrease</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>% Shredders</td>
<td>Decrease</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>% Filterers</td>
<td>Decrease</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>% Predators</td>
<td>Variable</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>% Collector-gatherer</td>
<td>Variable</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 3
Stepwise forward regression between each metric and environmental variables. Bio04 = temperature seasonality (standard deviation × 100), Bio12 = annual precipitation. Residual indicate if residuals of regressions were used instead of raw metric values.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>$r^2$</th>
<th>P</th>
<th>Intercept</th>
<th>Slope</th>
<th>Environmental Variable</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Margg</td>
<td>0.286</td>
<td>&lt;0.05</td>
<td>–49.109</td>
<td>0.078</td>
<td>Bio04</td>
<td>Residual</td>
</tr>
<tr>
<td>BMWP_re</td>
<td>0.4058</td>
<td>&lt;0.01</td>
<td>–9.349</td>
<td>0.009</td>
<td>Bio04</td>
<td>Residual</td>
</tr>
<tr>
<td>EPTsBH_re</td>
<td>0.4888</td>
<td>&lt;0.001</td>
<td>–17.280</td>
<td>0.013</td>
<td>Bio04</td>
<td>Residual</td>
</tr>
<tr>
<td>EPT</td>
<td>0.2389</td>
<td>&lt;0.05</td>
<td>–4.707</td>
<td>0.006</td>
<td>Bio12</td>
<td>Residual</td>
</tr>
<tr>
<td>Equif</td>
<td>0.3092</td>
<td>&lt;0.05</td>
<td>–28.058</td>
<td>0.041</td>
<td>Bio12</td>
<td>Residual</td>
</tr>
<tr>
<td>Shanf</td>
<td>0.3154</td>
<td>&lt;0.01</td>
<td>–3.209</td>
<td>0.006</td>
<td>Bio04</td>
<td>Residual</td>
</tr>
<tr>
<td>IBE-IOC_re</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>%Ple</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Domf</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Margg – Margalef index in family-level; BMWP_re – Residuals for natural variation of metric adaptation of the Biological Monitoring Working Party; EPTsBH_re – Residuals for natural variation of metric number of Ephemeroptera, Plecoptera and Trichoptera excluding Baetidae and Hydropsychidae families; SEPTsBH – Percentage of Ephemeroptera, Plecoptera and Trichoptera excluding Baetidae and Hydropsychidae families; EPT – Number of Ephemeroptera, Plecoptera and Trichoptera families; Equif – Equitability index; SEPT/CCh_re – Residuals for natural variation of metric ratio of Percentage of Ephemeroptera, Plecoptera and Trichoptera and %Chironomidae; %Coleo – Percentage of Coleoptera; Shanf – Shannon-Wiener index families; IBE-IOC_re – Residuals for natural variation of metric Índice Biótico Estendido Instituto Oswaldo Cruz; %Ple – Percentage of Plecoptera and Domf – Dominance index (1-Simpson).
33 sites. Only two out of 23 reference sites were not classified correctly (one as “impaired”, one as “regular”). Six intermediate sites were classified as “impaired”. Regarding the PCAs, only axis 1 was significant according to the broken-stick model (56.7% of the variance explained for the development dataset and 49% for the validation dataset). In both cases, axis 1 displayed sites according to the impairment gradient. Significant correlations were found between MMI-2 scores and their corresponding PCA axis 1 values for both datasets ($r = -0.61$, $p < 0.001$ for the development dataset and $r = -0.59$, $p < 0.001$ for the validation dataset).

4. Discussion

Accounting for naturally occurring variation is an ideal and logical approach for MMIs (Stoddard et al., 2008; Hawkins et al., 2000). However, this is a major difficulty in metric selection and MMI development (Cao et al., 2007; Buss et al., 2015). Other studies showed that calibrating metrics for natural gradients improved MMI performance and decreased type I and type II errors of inference (Hawkins et al., 2010; Chen et al., 2014). This was only partially true in our study. MMI-3 – built only with unadjusted metrics – had the best precision. In addition, in our “adjusted” MMI-2 only half the metrics had to be adjusted because of their response to natural variability. Still, as shown in Table 4, MMI-2 increased its responsiveness and sensitivity to detect impairment, which is the ultimate goal of bioassessment tools (Stoddard et al., 2008).

Other studies also reported an apparent low response of macroinvertebrate to natural variation. According to Stoddard et al. (2008) calibrating metrics appears to be more important for aquatic vertebrate MMIs than for macroinvertebrate MMIs. Macedo et al. (2016) found that only one in four metrics included in their MMI needed to be adjusted. They also found – as well as Chen et al. (2014)
and our results – that all indices, adjusted or unadjusted, performed significantly well in the three criteria (precision, responsiveness and sensitivity).

Some authors argue that selecting one metric from each cluster (or each axis of a principal component analysis) is an effective way of identifying groups of statistically covarying metrics, which enable to select non-redundant metrics for MMIs (Cao et al., 2007). Following this procedure improved MMI performance (Chen et al., 2014). Stoddard et al. (2008) argue that redundancy between metrics should not be avoided per se, but that it could occur because each metric could be responding to a single environmental parameter that may covariate with each other in the database used for the study. One proposed solution to circumvent that is eliminating the stressor gradient from this step. They state that this way it is possible to avoid eliminating metrics that respond to similar stressor gradients, but that reflect different taxonomic information. In our study, using only reference sites to reduce the effect of stressors, MMI-1 and MMI-3 – which used cluster analysis – did not have lower correlations between metrics than the ones where metrics were selected from each category (Table 4). Also, although MMI-1 had no pair of metrics with correlation $r > 0.70$, MMI-3 had a pair with very high correlation ($Domf \times Shanf r = -0.91$). Chen et al. (2014) have also reported high correlations ($r > 0.8$) between metrics when using the cluster approach. They have decided to exclude one of the correlated metrics, choosing to miss one potential unique response (since this metric was in a different cluster branch), in a way to avoid redundancy. An alternative would be selecting the second-best metric in that cluster, provided it was not...
redundant with the other selected metrics, but that could decrease MMI discrimination power – as we showed in our study.

We had two hypotheses in our study: (1) testing and adjusting metrics to landscape parameters and (2) selecting metrics using a cluster analysis to avoid metrics redundancy, would increase MMI performance. Our results do not support it. In fact, MMI-1 did not outperform other MMIs in any criteria. However, the two best performance MMIs used one of this approaches each: MMI-2, with two out of four adjusted metrics, one selected from each category, had best responsiveness and sensitivity; MMI-3, only with unadjusted metrics using a cluster approach for metric selection, had lower correlation among metrics and better precision. We believe that MMI’s development should benefit from incorporating responses to natural variation, and also by using a system less subjective than choosing one from each category. For example, many metrics (e.g. EPT richness) could be classified either as a “richness” or as a “tolerance” metric. We support that further studies are necessary to validate those methods. In addition, we notice that some recent publications failed to report the full set of information to allow MMI calculation. We believe these tools should be made readily available for managers and we urge studies to provide the following: intercept and slope for the residuals of each environmental variable significantly associated with the selected metrics; the transformation used for each metric to reach normalization (if any); and the quartiles 25% and 75% of reference and impaired sites.

4.1. Considerations on selected metrics

EPT-derived metrics had been used in several MMIs around the world (e.g. Nelson and Williams, 2013; Lakew and Moog, 2015), and also in Brazil (e.g. Ferreira et al., 2011; Couceiro et al., 2012). In our study, two metrics were selected representing this group: %Ple and the number of EPT families, excluding Baetidae and Hydropsychidae (EPTsBH). Plecoptera is often classified as sensitive taxa, and have been used in many MMIs around the world (Rosenberg and Resh, 1993; Törnblom et al., 2011). Recent studies indicate this group is also a good surrogate for the assessment of climate change effects (Tierno de Figueroa et al., 2010). The metric EPTs-BHF performed better than EPT richness because Baetidae and Hydropsychidae had higher relative abundance in intermediate and impaired sites (46% and 69%, respectively), thus decreasing their discriminating power. Among Hydropsychidae, the genus Smicridea was responsible for 98% abundance of the family. Other MMIs around the world report that (Barbour et al., 1999; Bellucci et al., 2013). This is also supported by other studies in Brazil. For example, Buss and Salles (2007) described the habitat and habits of Baetidae species, and classified the genus Americabaetis as “tolerant” and the genera Baetodes, Camelobaetidius and Cryptonympha as “somewhat sensitive”. We recommend other studies to consider this metric for screening while developing MMIs in Brazil and elsewhere.

The other selected metrics for MMI-2 were the Margalef index in family level and the IBE-IOC index. Other recent multimetric indices selected the Margalef index (Nguyen et al., 2014), including Nelson and Williams (2013) that developed a multimetric index in lowland streams in Panama, an ecological condition similar to the one described in the present study. The better performance of the IBE-IOC is likely to have occurred because the index was developed using information from a similar study area (Rio de Janeiro state; Mungai et al., 2008).

4.2. Further considerations for MMIs development

No trophic metrics were retained for our final MMI. Trophic metrics are often used in MMIs (e.g. Stoddard et al., 2008; Couceiro et al., 2012) because they can serve as a proxy for ecosystem functioning (Petchey and Gaston, 2006). Two metrics were discarded because they did not discriminate the three impairment conditions (Scrapers and Filterers) and the three other metrics (Shredders, Predators and Gatheries-collector) were excluded because they did not respond linearly to the impairment gradient. Moya et al. (2011), studying Bolivian streams, also reported that trophic measures were not approved while screening for a MMI. In the neotropical region only a few studies aimed to determine macroinvertebrate functional feeding groups (FFGs) either by analyzing gut contents (Tomanova et al., 2006; Miserendino, 2007; Chará-Serna et al., 2012) or ultra-structure of mouthparts (Baptista et al., 2007). As a result, most MMIs developed in the region simply apply the FFG attributed to a taxon based on temperate species, which may lead to wrong classifications. Some studies in neotropical region suggest most genera have omnivorous feeding habits (Tomanova et al., 2006), adding difficulties to correctly applying this approach. Furthermore, in many cases MMIs are calculated in family-level, which make FFG assignment increasingly prone to errors considering there is considerable within-family variation in the feeding habits of macroinvertebrate species in neotropical streams (Moya et al., 2007). Another limitation for using these metrics is that since they express the energy flow through the compartments of the food chain and the ecosystem, it is necessary to evaluate the biomass of each FFG, and not their relative abundance (a common approach for calculating other metrics). Using data on relative abundance as a proxy clearly change the relative importance of FFGs. For example, Masese et al. (2014) showed that shredder relative abundance in closed-canopy forested streams was ~20%, but the same groups yielded corresponded to ~80% of total macroinvertebrate wet mass in Kenya. Similarly, Chá et al. (2007) found that shredders represented 13% of the abundance and 68% of the biomass of invertebrates colonizing leaf litter bags in Andean Colombian streams. The easiest method for calculating the biomass is using length-biomass regressions based on total length, head width or interocular distance (Benke et al., 1999). That requires a previous effort to build the regressions and then measure organisms individually to convert length to biomass. The use of FFG biomass may be used as a tool for assessing ecosystem functioning. However, there are few length-biomass regression curves available for neotropical macroinvertebrates, not covering most groups (Cressa, 1999; in Venezuela; Miserendino, 2001; in Argentina; Chará-Serna et al., 2012, for leaf litter fauna in Colombia; and Becker et al., 2009; only for the shredder genus Phylloicus – family Calamoceratidae, order Trichoptera – in Brazil). More studies are necessary in this region, which will allow not only the application of FFG-based metrics in biomonitoring programs, but also on studies of ecosystem functioning and to provide information on the supposed scarcity of shredders in the tropics (Masese et al., 2014; Boyero et al., 2011).

Substrate availability is one of the main drivers for the colonization and the structure of macroinvertebrate assemblages in streams. Several papers have demonstrated the importance of the substrate for macroinvertebrate assemblages (Jiang et al., 2010; Demars et al., 2012). In Brazil, Buss et al. (2004) reported that taxa occurrence was highly dependent on substrate type (82 out of 86 taxa had a >40% preference for one substrate), and although degradation influence significantly the macroinvertebrate fauna on each substrate, they were responding to different physicochemical parameters. Many biomonitoring programs recognize the importance of substrate types for macroinvertebrates, and recommend a multi-habitat sampling approach, usually in proportion to their occurrence (e.g. Barbour et al., 1999; AQEM, 2002). In New Zealand, Stark et al. (2001) showed significant differences in the structure and composition of the community of macroinvertebrates between hard- and soft-bottom stream types. They reported the latter as having less productive habitats and lower diversity and number of invertebrates. In our study, however, we did not find signifi-
cant differences for any metric values between reference sites of the two stream types ("transitional/seasonal areas" and "rocky substrates"), and reference sites were significantly different than impaired sites, for both stream types.

5. Conclusions

Although all MMIs performed well in all three decision criteria, adjusting metrics to natural variation increased MMI response and sensitivity to impairment. In addition, the selected MMI-2 (adjusted metrics, one representing each category) was able to classify sites of two stream types and two ecoregions. The use of cluster analysis, however, did not avoid high redundancy between metrics of different branches. The MMI developed using the traditional approach had the poorest performance among all tested MMIs and it was not able to distinguish adequately reference and impaired sites from ecoregion 2.

Acknowledgments

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq/PROEP 400107/2011-2). We are grateful to LPASA team members for field assistance.

References


