Published in final edited form as:
Ecol Indic. 2020 August 1; 115: . doi:10.1016/j.ecolind.2020.106382.

Choice of field and laboratory methods affects the detection of anthropogenic disturbances using stream macroinvertebrate assemblages

Raphael Ligeiro[1,*], **Robert M. Hughes**[2], **Philip R. Kaufmann**[3], **Jani Heino**[4], **Adriano S. Melo**[5], **Marcos Callisto**[6]

[1]Universidade Federal do Pará, Instituto de Ciências Biológicas, Laboratório de Ecologia e Conservação, Av. Augusto Correia 01, CEP 66075-110, Belém, Pará, Brazil. [2]Amnis Opes Institute and Department of Fisheries & Wildlife, Oregon State University, 97331, Corvallis, Oregon, USA. [3]U.S. Environmental Protection Agency, Office of Research & Development, Center for Public Health & Environmental Assessment, Pacific Ecological Systems Division, 200 SW 35 Street, 97333, Corvallis, Oregon, USA, and Department of Fisheries & Wildlife, Oregon State University, 97331, Corvallis, Oregon, USA. [4]Finnish Environment Institute, Freshwater Centre, Paavo Havaksen Tie 3, 90570, Oulu, Finland. [5]Universidade Federal do Rio Grande do Sul, Instituto de Biociências, Departamento de Ecologia, Av. Bento Gonçalves 9500, CEP 91501-970, Porto Alegre, Rio Grande do Sul, Brazil. [6]Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas, Departamento de Genética, Ecologia e Evolução, Laboratório de Ecologia de Bentos, Av. Antônio Carlos 6627, CEP 30161-970, Belo Horizonte, Minas Gerais, Brazil.

Abstract

Accurate and precise detection of anthropogenic impacts on stream ecosystems using macroinvertebrates as biological indicators depends on the use of appropriate field and laboratory methods. We assessed the responsiveness to anthropogenic disturbances of assemblage metrics and composition by comparing commonly employed alternative combinations of field sampling and individuals counting methods. Four datasets were derived by, in the field 1) conducting multihabitat sampling (MH) or 2) targeting samples in leaf packs (single-habitat sampling – SH) and, in the laboratory A) counting all individuals of the samples, or B) simulating subsampling of 300 individuals per sample. We collected our data from 39 headwater stream sites in a drainage basin located in the Brazilian Cerrado. We used a previously published quantitative integrated disturbance index (IDI), based on both local and catchment disturbance measurements, to characterize the intensity of anthropogenic alterations at each site. Family richness and % Ephemeroptera, Plecoptera and Trichoptera (% EPT) individuals obtained from each dataset were tested against the IDI through simple linear regressions, and the differences in assemblage composition between least- and most-disturbed sites was tested using Permutational Multivariate Analysis of Variance (PERMANOVA). When counting all individuals, differences in taxonomic richness and assemblage composition of macroinvertebrate assemblages between least- and most-disturbed sites were more pronounced in the MH than in the SH sampling method. Leaf packs

[*]Corresponding author: ligeirobio@gmail.com.

seemed to concentrate high abundance and diversity of macroinvertebrates in highly disturbed sites, acting as ‘biodiversity hotbeds’ in these situations, which likely reduced the response of the assemblages to the disturbance gradient when this substrate was targeted. However, MH sampling produced weaker results than SH when subsampling was performed. The % EPT individuals responded better to the disturbance gradient when SH was employed, and its efficiency was not affected by the subsampling procedure. We conclude that no single method was the best in all situations, and the efficiency of a sampling protocol depends on the combination of field and laboratory methods being used. Although the total count of individuals with multihabitat sampling obtained the best results for most of the evaluated variables, the decision of which procedures to use depends on the amount of time and resources available, on the variables of interest, on the availability of habitat types in the sites sampled, and on the other methods being employed in the sampling protocol.

Keywords

Biomonitoring programs; multihabitat sampling; single-habitat sampling; leaf packs; subsampling effect; method comparisons

Introduction

Biomonitoring has long been established as a key tool for assessing and managing water resources worldwide (Ruaro and Gubiani, 2013; Buss et al., 2015; Ruaro et al., 2020), and macroinvertebrates are among the most common indicator groups used in the evaluation of stream ecological condition (Karr and Chu, 1999; Bonada et al., 2006). However, the sampling protocols used by various agencies and research groups vary widely, including procedures employed in the field (e.g., sampling apparatus used, sampled area, number and type of sampled habitats) and in the laboratory (e.g., number of subsamples or individuals counted, level of taxonomic resolution) (Carter and Resh, 2001; Cao and Hawkins, 2011; Buss et al., 2015). Accordingly, the effectiveness of macroinvertebrates in detecting anthropogenic pressures and stressors depends on the methods adopted (Gerth and Herlihy, 2006; Rehn et al., 2007; Stoddard et al., 2008; Chen et al., 2015). In this context, it is important to know which methods result in best responses, considering time and resource constraints (Doberstein et al., 2000; Hughes and Peck, 2008).

Concerning field methods, the choice of the number and type(s) of habitat(s) to be sampled is one of the issues that has generated most debate (Parsons and Norris, 1996; Buss et al., 2004; Gerth and Herlihy, 2006; Chessman et al., 2007; Rehn et al., 2007; Blocksom et al., 2008). There are two basic approaches for sampling benthic macroinvertebrates in stream biomonitoring: multihabitat (MH) and single-habitat (SH) sampling, the latter also known as targeted sampling (Blocksom et al., 2008; Hughes and Peck, 2008). In MH sampling, a combination of the common habitat types (substrate or hydraulic types) present at each stream site is sampled, usually yielding a composite sample to represent the entire site. The different habitats can be sampled systematically along the site (which the US EPA calls “reachwide sampling”, Stoddard et al., 2005; Hughes and Peck, 2008), or in proportion to the researcher’s visual estimate of their coverage (as used in Europe in the AQEM and

STAR projects, Hering et al., 2006). Both ways to conduct multihabitat sampling will yield similar faunal collections if actual and perceived habitat distributions are similar at a site.

In SH sampling, one habitat type (e.g., riffles, snags, channel edge, leaf packs) present in all sites is defined before the field sampling (e.g., Reynoldson et al., 1999; Peck et al., 2006). The main advantage of SH sampling is the intrinsic standardization obtained by not comparing sites where different kinds of habitats were collected. SH sampling is supposed to reduce the data 'noise' (i.e., data variability caused by other factors than anthropogenic disturbances) in relation to the disturbance signal, in this way contributing to a more accurate assessment (Parsons and Norris, 1996; Gerth and Herlihy, 2006). One drawback of SH sampling is the insensitivity of this method to changes in the proportion of the chosen habitat type caused by anthropogenic alterations, that is, the same amount of the same habitat type is always sampled in all sites, independent of changes in the availability of habitat types caused by human activities. Another, more practical, difficulty is finding the same pre-defined habitat type in all stream sites to be compared across large spatial extents (i.e., entire river basins, regions or countries) (Turak et al., 1999; Wells et al., 2002; Blocksom et al., 2008). Third, the fauna found in a particular habitat type is usually only a subset of the entire assemblage of a site, and there is always the possibility that other unsampled habitats may respond better to ongoing human caused changes (Kerans et al., 1992; Roy et al., 2003).

Regarding laboratory procedures, a fundamental decision is whether to process all the individuals of the samples or subsample them (Carter and Resh, 2001). Processing the whole sample is considered by many the most sensitive method (Courtemanch, 1996; Doberstein et al., 2000), and is desirable if the main concern is to enumerate rare taxa. Counting all the individuals is also preferable for obtaining less biased estimates of taxonomic density (Ligeiro et al., 2013a), defined as the number of taxa found in a given sampled area (Hurlbert, 1971). However, this method necessitates high counts of individuals and, consequently, it can be quite expensive and time-consuming (Buss et al., 2015). In addition, samples may be biased by fewer individuals (and taxa) occurring in naturally oligotrophic or homogeneous sites versus large numbers of individuals (and more taxa) in enriched or naturally heterogeneous sites (Gotelli and Cowell, 2001).

Subsampling procedures are implemented to reduce costs and make extensive biomonitoring programs more feasible (Vinson and Hawkins, 1996; Hughes and Peck, 2008). Arguably, the approach most commonly used is fixed-count subsampling (Carter and Resh, 2001), which consists of sorting and identifying a fixed number of individuals from each sample to generate standardized measurements of taxonomic richness (called numerical taxonomic richness, Hurlbert, 1971) and other related metrics. This is important because the number of taxa detected depends intrinsically not only on the area sampled, but also on the number of individuals sampled (Gotelli and Colwell, 2001). In biomonitoring protocols, the number of individuals counted varies from as few as 100 individuals for rapid assessments (Plafkin et al., 1989) to 500 individuals in national monitoring programs (Paulsen et al., 2008).

Although many studies have compared the efficiency of different methods in macroinvertebrate-based biomonitoring, very few have addressed the comparison of

different methods simultaneously (but see King and Richardson, 2002 and Chessman et al., 2007). In fact, this would be the most realistic way to approach the problem, because no step of the sampling protocol operates in isolation, and the overall efficiency of a sampling protocol is likely to result from the sum of all decisions that it comprises. A straightforward criterion to assess the efficiency of competing methods is measuring their capability to detect known disturbance gradients (Ostermiller and Hawkins, 2004). Therefore, in this study, we formed four different macroinvertebrate datasets through a combination of methods commonly used in the field (MH and SH sampling) and in the laboratory (counting all individuals and subsampling). We used assemblage metrics and taxonomic composition derived from the four datasets to assess their responsiveness to an index of anthropogenic pressures calculated for the stream sites. We hypothesized that 1) SH sampling performs best because standardizing microhabitat conditions among the sites introduces less environmental variability to the assessment; and 2) processing the entire sample provides the clearest distinction of the disturbance gradient because it maximizes differences in taxonomic diversity.

Materials and Methods

Study area and site selection

We sampled streams in the Upper Araguari River Basin ($46^0$30'W - $48^0$0'W; $19^0$0'S - $20^0$0'S), southeastern Brazil, located in the Cerrado biome of Minas Gerais State. The Cerrado is the second largest biome of Brazil, originally covering 2,045,064 km^2 (20% of Brazil). It is marked by predominantly savanna-like vegetation and two well-defined seasons: a wet season from October to March and a dry season from April to September, with 1200-1800 mm of rainfall per year (Brasil 1992). The Cerrado is considered a terrestrial biodiversity hotspot (Myers et al., 2000) because of its high floral and faunal diversity and endemism (Oliveira and Marquis, 2002), and high rates of habitat loss over the past 50 years (Wantzen et al., 2006; Françoso et al., 2015).

The Araguari Basin has an extensive and well-developed system of irrigated/mechanized agriculture, mainly of soy, coffee, corn, and sugar cane. Pasture and small patches of relatively undisturbed vegetation are also present. Most people dwell in small towns, although a few small cities with up to 80,000 inhabitants are present. Thirty-nine stream sites from 1^{st} to 3^{rd} order (*sensu* Strahler, 1957, map scale 1:100,000) were sampled in a hydrologic unit of 7,376 km^2. They were randomly selected through a computerized probability-based design (Olsen and Peck, 2008) that assures a spatially balanced distribution of sites (Stevens and Olsen, 2003).

Field sampling and laboratory procedures

Field sampling was conducted in September of 2009, at the end of the dry season. Following Peck et al. (2006), each site consisted of a length equal to 40 × the mean wetted width, with a minimum site length of 150 m. Then, 11 equidistant cross-sectional transects were marked from downstream to upstream, defining 10 longitudinal sections of the same length within each site.

For the MH sampling, we employed the reachwide method as described in Peck et al. (2006). One macroinvertebrate sample unit was taken per transect, following a systematic zig-zag pattern along transects (right-middle-left). Each of these 11 sample units was taken through use of a D-net (30 cm mouth width, 500 μm mesh size) in 0.09 m^2 per sample unit summing to 0.99 m^2 of stream bottom area sampled per site. This method assures that many types of habitats, including different substrates and surface water profiles, are sampled at each site. It is expected that the habitats will be sampled in proportion to their occurrence within each stream site, although rare habitats with areal cover <10% of the stream channel may be missed.

For the SH sampling, eight leaf packs were sampled per site, preferably located in different site sections. The same D-net was used, summing to 0.72 m^2 of leaf pack area sampled per site. Leaf packs are microhabitats formed by a mixture of leaves from many plant species that accumulate in the streambed, coming mostly from the stream riparian vegetation (Moretti et al., 2007). In contrast to the pulsed input of leaf litter of temperate streams, leaf detritus inputs continue throughout the year in Cerrado streams, producing leaf packs in stream channels in all seasons (Gonçalves et al., 2006; França et al., 2009). Therefore, given their ecological importance, and high abundance and availability in tropical streams, we targeted our sampling on leaf packs when performing SH sampling. However, other SH sampling options are possible (e.g., riffles, pools, boulders, snags, macrophytes), as used in other protocols (Chessman et al., 2007; Rehn et al., 2007).

The individual sample units from each method were placed in separate plastic buckets, generating one composite sample for MH sampling and one composite sample for SH sampling per stream site. Both composite samples were preserved with 10% formalin in the field.

In the laboratory, all samples were fully processed (all individuals counted). Insects and gastropods were identified to family level through use of taxonomic keys (Pérez, 1988; Fernández and Domínguez, 2001; Costa et al., 2006; Mugnai et al., 2010). Only seven taxa, together representing < 4% of all individuals collected, were not identified to family (Collembola, Hydracarina, Tricladida, Nematoda, Hirudinea, Oligochaeta and Bivalvia). Hence, for simplicity we will refer to all identified taxa as families. Family-level identification of macroinvertebrates has proven to be efficient for biomonitoring purposes, with results comparable to those obtained with genus and species level (Melo, 2005; Marshall et al., 2006; Chessman et al., 2007; Whittier and Van Sickle, 2010), and it is a good option in many tropical regions that show high diversity of organisms and scarce taxonomic knowledge (Godoy et al., 2019).

Datasets compared and subsampling procedures

We compared four datasets with respect to their responsiveness to a known disturbance gradient, each dataset being a combination of the different field (MH and SH sampling) and laboratory (total counts of individuals and subsampling) methods considered in this study. We simulated subsampling of individuals via computer routines made in R software (R Development Core Team, 2018). Starting from the total counts in datasets of each of the two field sampling methods, we used the R function *rrarefy*, available in the *vegan* package

(Oksanen et al., 2018), to simulate the random subsampling (without replacement) of 300 individuals per site. Samples that originally yielded < 300 individuals were kept unaltered. We performed 200 subsampling simulations for each field sampling method, totaling 400 simulations.

There are protocols that subsample 400 or more individuals (Carter and Resh, 2001; Hughes and Peck, 2008), which can be desirable in terms of differentiation strength of the biological metrics (Cao et al., 2002; Chen et al., 2015). However, subsampling 300 individuals has also demonstrated adequate performance in metric estimations and site classifications, and has been recommended for biomonitoring purposes (Larsen and Herlihy, 1998; Sovell and Vondracek, 1999; Klemm et al., 2003; Boonsoong et al., 2009; Silva et al., 2017). In addition, many of our sites did not yield more than 300 individuals. Low densities of macroinvertebrates are common in the streams of many tropical regions (Heino et al., 2018).

Macroinvertebrate variables

To evaluate the responsiveness of the different methods to the disturbance gradient, we calculated from each dataset the total family richness and the percentage of Ephemeroptera, Plecoptera and Trichoptera (% EPT) individuals of the sites. Those two indicators represent important general aspects of taxonomic diversity and sensitivity of assemblages, which are among the most commonly used variables included in macroinvertebrate multimetric indices (Klemm et al., 2003; Stoddard et al., 2008; Waite et al., 2012; Macedo et al., 2016; Silva et al., 2017; Fierro et al., 2018; Ruaro et al., 2020).

Besides these two univariate metrics, we also evaluated the taxonomic composition of the datasets. This is because many macroinvertebrate metrics used in multimetric indices are derived from the assemblage composition of the samples (e.g., presence-absence or relative abundance of many groups) and predictive models are primarily based on the difference between taxonomic composition observed and expected at the sites (Wright, 1995; Reynoldson et al., 1997; Hawkins et al., 2000; Clarke et al., 2003).

Anthropogenic disturbance gradient

To quantitatively characterize the exposure of the stream sites to anthropogenic pressures, we used the Integrated Disturbance Index (IDI), described in detail in Ligeiro et al. (2013b). This index is based on the disturbances observed at the local scale (in-channel and riparian vegetation) and also at the catchment scale (land uses), because the ecological condition of any stream reach depends on both local and upstream catchment conditions (Allan, 2004; Whittier et al., 2007; Herlihy et al., 2020). We estimated local disturbance through use of the habitat metric *W1_hall* (Kaufmann et al., 1999), which is the mean number of specified types of anthropogenic disturbances observed at each transect (i.e., presence of buildings, channel revetment, pavement, roads, pipes, trash and landfill, parks and lawns, row crop agriculture, pasture, logging and mining), distance-weighted relative to their proximity to the stream channel. Catchment disturbance was calculated by summing the proportional areas of human land uses (i.e., pasture, agriculture and urban) in each site's catchment. The different land uses were weighted according to their potential to impair the aquatic environment (Rawer-Jost et al., 2004; Maloney et al., 2011). The IDI was then calculated as the Euclidean

distance between each stream site and the origin of the Cartesian plane formed by the local and the catchment indices (Ligeiro et al., 2013b). The greater the IDI score of a site, the greater the intensity of anthropogenic disturbance expected at that site, a zero value representing a site lacking the anthropogenic disturbances measured. For the pool of sites analyzed in the present study, the IDI values ranged from 0.05 (for a site located inside a forested protected area) to 0.93 (for a site inside a heavily urbanized area). Therefore, the sampled sites covered a wide range of anthropogenic disturbance.

The effects of human activities on the ecological condition of stream sites operate at many spatial levels and involve many intricate pathways (Macedo et al., 2014; Leal et al., 2016; Leitão et al., 2017). Accordingly, they are too complex to be summarized perfectly by any single index. Nonetheless, the IDI has proven to be a useful practical tool to rank sites according to their overall intensity of exposure to anthropogenic alterations (Terra et al., 2013; Macedo et al., 2016; Carvalho et al., 2017; Chen et al., 2017; Silva et al., 2017; Castro et al., 2018; Fierro et al., 2018; Sanches et al., 2019; Martins et al., 2020).

Data analyses

Comparisons between field sampling methods—We first compared the data obtained by MH and SH sampling methods, considering total counts of individuals of the samples of the 39 stream sites. We conducted paired t-tests on the number and density (individuals/m^2) of organisms (both ln(y) transformed), family richness, and % EPT individuals (logit transformed, ln(y/[1 / y]), as suggested by Warton and Hui, 2011). To test for the congruency of the assemblage composition between the two methods we performed PROCRUSTES analysis (Peres-Neto and Jackson, 2001), which uses a rotation algorithm that minimizes the sum of squared residuals between two dissimilarity matrices under comparison. We used as dissimilarity measures the Jaccard index for presence/absence data and the modified Gower distance for relative abundances. Following the advice of Anderson et al. (2006), data were transformed by $\log_2(y) + 1$, but with zeros not being transformed, instead remaining as zeros. The modified Gower distance gives a clearer and more effective representation of differences on relative abundances than other more popular dissimilarity measures (e.g., Bray-Curtis index) (Anderson et al., 2006). A correlation-like coefficient (r) was calculated between the dissimilarity matrices of both sampling methods, considering each dissimilarity measure (following Mardia et al., 1979). A randomization test (10,000 iterations) was made to estimate the statistical significance of the congruency observed. These analyses were performed with R software (R Development Core Team, 2018) using the *vegan* package (Oksanen et al., 2018) for Procrustes (function *protest*).

Assemblage metrics versus disturbance gradient—To test the performance of the four datasets in detecting the intensity of anthropogenic disturbances, family richness and % EPT individuals (logit transformed, as above) were regressed through simple linear regressions (SLR) against the IDI values of the sites. We generated one regression model for each total-count dataset, and 200 regression models for each subsampled dataset (one model per subsampling simulation). The regression models were conducted with STATISTICA 7.0 software (StatSoft Inc., 2004).

The strength of response of each dataset was measured by the *F* values of their regression models. In SLR, the *F* value indicates how many times the mean square model is greater than the mean square error (Zar, 2010). Therefore, in our study, higher *F* values indicate greater responsiveness of the datasets to the disturbance gradient.

To determine the degree that the two methods of counting individuals differed in their strength of response, we compared the single *F* values obtained from the total-count datasets with the distribution of the 200 individual *F* values obtained from the subsampling simulations of 300 individuals per site through a standardized measurement of differentiation (*Z* values):

$$Z = (F_{observed} - \text{Mean } F_{simulations}) \; / \; \text{Standard deviation of } F_{simulations}$$

The higher the modular value of *Z*, the greater the difference between the observed F value and the distribution of simulated *F* values. Usually, *Z* values > 1.96 (or < −1.96) indicate that the *F* value observed in the total counts is highly distinct from the mean *F* value of the simulations (Zar, 2010).

Assemblage compositional dissimilarities—We compared the assemblage composition between groups of sites having low and high intensities of anthropogenic disturbance. We included in the least-disturbed category all sites with IDI values < 0.3 and in the most-disturbed category all sites with IDI values > 0.6. Those thresholds clearly distinguish two groups of stream sites in terms of their intensity of exposure to anthropogenic pressures, as suggested by Ligeiro et al. (2013b). The least- and the most-disturbed categories were represented by six and seven sites, respectively.

To test which dataset best discriminated the assemblage composition between least- and most-disturbed sites, we performed Permutational Multivariate Analysis of Variance (PERMANOVA, Anderson, 2001), employing as dissimilarity measures the Jaccard index and the altered Gower distance.

We used the *adonis* function in the *vegan* package (Oksanen et al., 2018) of R software (R Development Core Team, 2018), and employed 10,000 randomizations in each comparison to test model significance. Again, we generated a single PERMANOVA model for each total-count dataset, whereas we generated 200 PERMANOVA models for each subsampled dataset (one model per subsampling simulation). Once more, the *F* values of PERMANOVA models were used to measure the discrimination strength of each dataset, and *Z* values were calculated to determine the degree to which the two laboratory processing methods (total counts of individuals and subsampling) differed in their discrimination strength. Our general analytical framework is summarized in Figure 1.

Results

Comparisons between field sampling methods

We collected a total of 22,345 and 21,508 individuals in the MH and SH field methods, respectively. The number of families found was also similar; 69 in MH and 66 in SH,

totaling 77 families. Most families (58) were sampled using both methods, with 11 families found exclusively in MH and 8 found only in SH sampling (Supplementary Material 1).

The number of individuals and % EPT individuals per site did not differ significantly between the two methods (Table 1). However, MH sampling produced significantly more families, whereas SH sampling produced a higher density of macroinvertebrates per site (Table 1).

In general, the relative abundances of major macroinvertebrate groups relative to the total number of individuals collected differed little between the two sampling methods (Figure 2). In both cases, insects comprised 96% of the individuals collected, with Diptera being the dominant insect order and Chironomidae the most abundant family. MH sampling produced more EPT individuals (28%, versus 20.8% in SH sampling), particularly Ephemeroptera (17.6%, versus 10.6% in SH sampling), whereas SH sampling resulted in a higher percentage of Chironomidae (46.2%, versus 40.4% in MH sampling) (Figure 2). PROCRUSTES analysis showed a significant correspondence between the assemblage composition yielded by both MH and SH methods, with r = 0.73 for Jaccard index and r = 0.79 for Gower distance ($p < 0.001$ in both cases).

Assemblage metrics versus disturbance gradient

Considering the response of macroinvertebrate family richness to the quantitative disturbance gradient, MH sampling showed better results than SH, independent of the method of counting individuals employed (Table 2). The regression model of SH family richness was not even significant when total counts of individuals were considered (p = 0.18, Table 02, Supplementary Material 02). However, subsampling increased considerably the responsiveness of SH family richness to disturbance (Z = −2.41) (Table 2, Supplementary Material 03). Subsampling did not affect considerably MH performance regarding the family richness metric (Z = 0.67).

For % EPT individuals, the pattern was inversed, SH presenting a better response than MH independent of the method for counting individuals used (Table 2, Supplementary 03). For MH sampling, the regression model of % EPT individuals was not significant considering total counts of individuals (p = 0.064, Table 2, Supplementary Material 02) and most subsampling simulations (91.5%) were non-significant as well. Subsampling did not affect the performance of the models, as showed by the low Z values (0.18 in MH, and −0.01 in SH), indicating that this metric was highly stable across the counting methods.

Assemblage compositional dissimilarities

Regarding the comparisons of assemblage composition between least- and most-disturbed sites, the altered Gower distance, accounting for the relative abundance of the taxa, always presented better results (higher PERMANOVA F values) than the Jaccard index (presence/absence data), for all field and counting of individuals methods (Table 3, Supplementary Material 04). For both dissimilarity measures, MH presented better results than SH when considering total counts of individuals, whereas SH performed better than MH when subsampling was employed (Table 3, Supplementary Material 05). In three out of four cases, subsampling reduced responsiveness of the datasets to disturbances, but not for SH using

altered Gower, for which models obtained from subsampling simulations usually had PERMANOVA F values higher than in total counts (Z = −2.95). This decrease in performance was most pronounced in the MH field sampling method, for which the PERMANOVA *F* values for both dissimilarity measures were almost halved in the subsampled datasets (Table 3).

Differences between the field sampling methods according to the disturbance gradient

Given the better performance of MH sampling for almost all assemblage variables when considering total counts of individuals, we further explored how the differences between these two field sampling methods varied along the disturbance gradient. To do so, we conducted Pearson correlations of the IDI values of the sites versus differences calculated between MH and SH in terms of number and density of individuals, family richness and % EPT individuals. We also obtained Jaccard index and altered Gower distance between the two methods for each site and correlated them with the IDI values. That is, we wanted to know whether the magnitude of the differences between MH and SH samples were influenced or not by the disturbance status of the stream sites.

We found that differences of % EPT individuals and assemblage composition (assessed through both dissimilarity measures) between the two field sampling methods were not influenced by the disturbance gradient (Figure 3), i.e., the same patterns of difference were observed from least- to most-disturbed sites. However, we found significant ($p < 0.05$) negative correlations between the IDI and differences in the number and density of individuals (r = −0.42 and −0.41, respectively) and differences in family richness (r = −0.38) (Figure 3). In other words, in least-disturbed sites (low IDI values), MH samples presented higher quantity and density of individuals and higher family richness than SH samples, whereas in most-disturbed sites (high IDI values), SH samples presented higher quantity and density of individuals and higher family richness than MH samples.

Discussion

The effects of methodological decisions on the observation of biodiversity patterns is a long-lasting discussion among stream ecologists, and it is one of central importance because the understanding of assemblage patterns and dynamics rely on the sampled data (Melo, 2005; Godoy et al., 2019; Sgarbi et al., 2020). These effects are even more dramatic in the interpretation of ecological indicators of anthropogenic alterations, considering that the outputs of biomonitoring studies are supposed to guide conservation and management decisions (Bonada et al., 2006; Hughes and Peck, 2008). These decisions often involve social actors (e.g., stakeholders, decision makers, community representatives) who are typically unaware of the intricacies and pitfalls of biological studies.

In line with other studies that also aimed to test the effects of different methods on the detectability of disturbance gradients (e.g., Buss et al., 2004; Gerth and Herlihy, 2006; Rehn et al., 2007; Waite et al., 2012), we were not interested here to test the ‘endpoints’ of the ecological indicators (e.g., biotic indices, MMI’s, predictive models). These are, by definition, very particular for each regional context (Cao and Hawkins, 2011). Our purpose was to assess the behavior of the ‘building blocks’ of these indicators (i.e., univariate

metrics, assemblage composition), in this way providing more general conclusions about the effects of methodological decisions.

Influence of field sampling methods on the detection of anthropogenic disturbances

Our first hypothesis was that targeting our sampling in a single habitat (leaf packs), in this way standardizing to some degree the biophysical environment that sustains macroinvertebrates, we would obtain better responses to the disturbance gradient, as suggested by Gerth and Herlihy (2006). By reducing natural habitat variability among streams, we expected that the anthropogenic disturbance signal would be stronger (Parsons and Norris, 1996). However, in general, this hypothesis was not supported by our data. When considering total counts of individuals, the low, statistically non-significant, F value of the regression model indicated that the number of families did not differ along the disturbance gradient. Actually, SH samples in most-disturbed sites presented much higher abundance and diversity of macroinvertebrates than MH samples, inverting the general pattern observed in other sites. Accordingly, SH samples discriminated the assemblage composition regarding the disturbance gradient to a lesser extent than MH sampling, mainly when total counts of individuals were considered.

Sampling in leaf packs seemed to impair the detection of anthropogenic disturbances when total counts of individuals were employed and taxonomic richness and assemblage composition were considered as indicators. Leaf packs are very attractive microhabitats to macroinvertebrates, serving as food resources, habitat and shelter (Gjerlov and Richardson, 2004; Kobayashi and Kagaya, 2005; Ligeiro et al., 2010). Hence, this microhabitat may have served as refuges for organisms in our highly disturbed streams. In other words, leaf packs acted as ‘biodiversity hotbeds’ in most-disturbed sites, to some degree buffering the assemblages from the effects of anthropogenic alterations. Haapala et al. (2003) also found that leaf detritus aggregations in the streambed concentrated macroinvertebrates to a much higher degree in most-disturbed streams than in streams with good ecological condition.

In line with our findings, Chessman et al. (2006) also found that MH sampling better responded to anthropogenic disturbance gradients than SH sampling, even though they did not consider leaf packs as targeted habitats. These findings suggest that MH sampling, being capable of tracking the impairment of habitat heterogeneity caused by anthropogenic alterations, is more likely to provide an overall more powerful tool for biomonitoring. On the other hand, the % EPT individuals responded strongly to the disturbance gradient only when SH sampling was considered. This demonstrates that, although harboring a high number of individuals and macroinvertebrate families along the entire disturbance gradient, the number of sensitive individuals decreased greatly in leaf packs of most-disturbed sites, which can result from physical and chemical alterations in water quality (Feio et al., 2005).

We aimed to describe assemblage compositional differences between least- and most-disturbed sites considering pure compositional variation (Jaccard Index) and compositional plus relative abundance variations (altered Gower distance) (Anderson at al., 2006). In this study, all datasets yielded higher dissimilarities between least- and most-disturbed sites when the altered Gower distance was employed. Anthropogenic disturbances on streams can change both the macroinvertebrate relative abundances, decreasing the number of

individuals of some taxa and increasing the number of others, and also the taxonomic composition, via local extinctions of some taxa and invasion of others (Karr, 1999; Norris and Thoms, 1999; Davies and Jackson, 2006). Therefore, we suggest that dissimilarity measures that account for both features of assemblage compositional dissimilarity are likely to respond better to anthropogenic alterations.

Effects of subsampling procedures on the assessment of disturbance effects

We found that subsampling of individuals had differing effects on the ability of the datasets to detect the anthropogenic disturbance gradient, depending on the field sampling method employed. In this way, our second hypothesis was not completely corroborated either, since we expected that counting all the individuals would always generate better results than subsampling.

Barbour and Gerritsen (1996), Vinson and Hawkins (1996), and Walsh (1997) recommended subsampling when comparing sites of markedly different ecological conditions, a practice widely performed in recent large-extent bioassessments (Hughes and Peck, 2008; Cao and Hawkins, 2011). Subsampling can standardize the sampling effort in terms of number of individuals, which is highly recommended when comparing sites with very different macroinvertebrate densities (Gotelli and Cowell, 2001).

On the other hand, Courtemanch (1996) argued that counting all individuals (i.e., considering taxonomic density) better describes the relative abundance among taxa and enhances the importance of rare taxa, which often encompasses the majority of macroinvertebrate diversity. Doberstein et al. (2000) achieved weaker models when employing subsampling to analyze taxonomic richness and other assemblage metrics, and they also advocated counting all individuals in samples for a more comprehensive understanding of anthropogenic alterations. Besides taxonomic diversity, total abundance of macroinvertebrates may also be responsive to anthropogenic pressures, and this information is completely lost during subsampling.

Neither taxonomic density (generated through the count of all individuals of the samples) nor numerical taxonomic richness (generated through subsampling of individuals) is necessarily the ‘correct’ way to measure taxonomic diversity, each method emphasizing different aspects of diversity patterns (Larsen and Herlihy, 1998; Gotelli and Cowell, 2001). Indeed, the primary reason to subsample is to help make regional and national biomonitoring programs that involve hundreds or thousands of sites more timely and cost-effective (Vinson and Hawkins, 1996; Hughes and Peck, 2008), and provide a standard method that would facilitate national monitoring data syntheses (Cao and Hawkins, 2011). Thus, a key point is to define which aspect of taxonomic diversity better responds to the effects of anthropogenic disturbances, and to which degree subsampling impairs or increase the responsiveness of other assemblage variables.

In our study, the subsampling procedure mostly impaired the responses given by the MH method to the disturbance gradient, mainly considering assemblage composition. However, for the SH sampling, the responses of the univariate metrics and the assemblage composition were mostly increased by subsampling 300 individuals. Whether this positive effect of

subsampling on SH can be attributed to targeting leaf packs specifically or any other standardized habitat/microhabitat is a matter for future studies. Nonetheless, our results suggest that the advantages of sampling across various habitat types are only fully realized when all individuals are counted. Since SH standardizes variation in macroinvertebrate assemblages among microhabitat conditions, it seems that it performs better with standardizing the number of individuals.

It is noteworthy that the number of individuals collected in this study (average of about 550 individuals / site) was relatively small compared to the abundances found in temperate or boreal streams (see Heino et al. 2018). Even showing smaller abundances, the impact of the subsampling procedure on the responses of the datasets was substantial in our case, indicating that this impact may be even greater in streams with greater numbers of individuals.

In contrast with other variables assessed, the % EPT individuals gave fairly uniform results for total counts and subsampled datasets. Therefore, this metric can be considered more stable than family richness and assemblage composition with respect to the method of counting individuals. This result is in agreement with Courtemanch (1996) who argued that, once individuals are collected randomly during subsampling, metrics that deal with proportions of individuals would be more stable than taxonomic richness and composition, and other related metrics as well. Despite that, % EPT individuals presented poor performance in MH sampling, which indicates that microhabitat heterogeneity may buffer the impacts of anthropogenic disturbance on EPT populations.

Effects of spatial extent on macroinvertebrate assessments

Many authors have agreed that MH and SH samplings rarely lead to markedly different responses of macroinvertebrate assemblages to anthropogenic disturbances, even when methods differ on the assemblages obtained (Plafkin et al., 1989; Ostermiller and Hawkins, 2004; Gerth and Herlihy, 2006; Rehn et al., 2007). For instance, Hewlett (2000) found that, despite assemblage differences observed between the methods, both MH and SH datasets generated the same site classifications. However, this emerging conclusion was generated mostly from studies dealing with very large spatial extents (> 200,000 km^2), frequently related to regional or national biomonitoring programs. The larger the study spatial extent, the greater the environmental heterogeneity present within that region (Jackson et al., 2001; Heino et al., 2015). Therefore, as spatial extent increases, the main determinants of changes in the structure and composition of the assemblages moves from local factors (e.g., substrate, water velocity) to regional factors (e.g., precipitation, geomorphology, land use) (Wu and Loucks, 1995; Wiens, 2002; Bonada et al., 2008; Heino, 2009), reducing the proportional variability generated by the sampled habitat (Gerth and Herlihy, 2006).

Perhaps because our study was conducted in a much smaller area (approximately 7,400 km^2), we found that the choice of the sampled habitat had a clear effect on the responses of assemblage metrics and composition to disturbance gradients. Chessman et al. (2007), studying a relatively small spatial extent in western Australia, also found differences in the performances of metrics derived from different habitats in detecting anthropogenic alterations in streams. These results suggest that biological assessments conducted across

smaller spatial extents, including those made in smaller river basins and BACI designs, tend to be more sensitive to the choice of the field sampling method than studies conducted across larger geographical areas.

Conclusions

In this work, we found that the choice of sampling and processing method had a significant impact on the detection of disturbance gradients by macroinvertebrate assemblages in a relatively small tropical basin. Moreover, the responsiveness of the datasets was affected by the combination of the field and laboratory sampling methods employed, which makes this matter even more problematic. If time and personnel are sufficient and available, employing MH with total counts of individuals provided the best results for almost all assemblage variables analyzed. On the other hand, the responsiveness of MH data to the disturbance gradient was mostly diminished after employing subsampling of 300 individuals. Future studies should investigate if subsampling more individuals can overcome this hindrance. For instance, some authors propose larger individual counts (as large as 500 individuals) to improve the accuracy and precision of comparisons (Cao et al., 2002; Chen et al., 2015). However, in areas with low densities of macroinvertebrates (as in many low-latitude streams, Heino et al., 2018), including our sites, it is not possible to subsample more than 300 organisms per sample. The same can happen anywhere if resources constrain the number of subsamples taken at a site or the number of individuals that can be identified in the laboratory.

Leaf packs, acting as ‘biodiversity hotbeds’ in most-disturbed streams, seemed to mask the human impacts when all individuals where counted. Despite the limitations we discussed, SH performed very well with subsampling, and can be an option to be considered if this procedure is necessary. In conclusion, no single field or laboratory method performed best in all cases, and the decision of which procedure to use depends largely on the amount of time and resources available, on the biological variables of interest, and on the other methods being adopted in the sampling protocol.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors are grateful to Programa Peixe-Vivo of Companhia Energética de Minas Gerais (CEMIG) and P&D Aneel-Cemig GT-599, for scholarships and financial support, to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES; finance code 001), to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq 307587/2017-7 to ASM and CNPq 303380/2015-2 to MC), to Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), and to Fulbright-Brasil for financial support. Carlos B. M. Alves provided logistical support. Colleagues from Universidade Federal de Minas Gerais (UFMG), Universidade Federal de Lavras (UFLA), Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG) and Pontifícia Universidade Católica de Minas Gerais (PUC-MINAS) helped with field sampling. Several colleagues from the Laboratório de Ecologia de Bentos ICB/UFMG helped with sample processing. Critical comments by Alan Herlihy, Joseline Molozzi, Paulo Pompeu, José Fernandes Bezerra Neto, Gilmar Bastos Santos and three anonymous referees improved the manuscript. This manuscript was reviewed by U.S. EPA’s Center for Public Health and Environmental Assessment’s Pacific Ecological Division and approved for publication. Approval does not signify that the contents reflect the views of the EPA, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

References

Allan JD 2004. Landscape and riverscapes: the influence of land use on river ecosystems. Annual Review of Ecology, Evolution, and Systematics 35:257–284. 10.1146/annurev.ecolsys.35.120202.110122.

Anderson MJ 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecology 26:32–46. 10.1111/j.1442-9993.2001.01070.pp.x.

Anderson MJ, Ellingsen KE, and McArdle BH. 2006. Multivariate dispersion as a measure of beta diversity. Ecology Letters 9:683–693. 10.1111/j.1461-0248.2006.00926.x. [PubMed: 16706913]

Barbour MT, and Gerritsen J. 1996. Subsampling of benthic samples: a defense of the fixed-count method. Journal of the North American Benthological Society 15:386–391. 10.2307/1467285.

Blocksom KA, Autrey BC, Passmore M, and Reynolds L. 2008. A comparison of single and multiple habitat protocols for collecting macroinvertebrates in wadeable streams. Journal of the American Water Resources Association 44:1–17. 10.1111/j.1752-1688.2008.00183.x.

Bonada N, Rieradevall M, Dallas H, Davis J, Day J, Figueroa R, Resh VH, and Prat N. 2008. Multi-scale assessment of macroinvertebrate richness and composition in Mediterranean-climate rivers. Freshwater Biology 53:772–788. 10.1111/j.1365-2427.2007.01940.x.

Bonada N, Prat N, Resh VH, and Statzner B. 2006. Developments in aquatic insect biomonitoring: a com-parative analysis of recent approaches. Annual Review of Entomology 51:495–523. 10.1146/annurev.ento.51.110104.151124.

Boonsoong B, Sangpradub N, and Barbour MT. 2009. Development of rapid bioassessment approaches using benthic macroinvertebrates for Thai streams. Environmental Monitoring and Assessment 155:129–147. 10.1007/s10661-008-0423-2. [PubMed: 18633723]

Brasil. 1992. Normais Climatológicas (1960-1990). Brasilia: Ministério da Agricultura e Reforma Agrária, Secretaria Nacional de Irrigação, Departamento Nacional de Meteorologia. 84p.

Buss DF, Baptista DF, Nessimian JL, and Egler M. 2004. Substrate specificity, environmental degradation and disturbance structuring macroinvertebrate assemblages in Neotropical streams. Hydrobiologia 518:179–188. 10.1023/B:HYDR.0000025067.66126.1c.

Buss DF, Carlisle DM, Chon T, Culp J, Harding JS, Keizer-Vlek HE, Robinson WA, Strachan S, Thirion C, and Hughes RM. 2015. Stream biomonitoring using macroinvertebrates around the globe: a comparison of large-scale programs. Environmental Monitoring and Assessment 187:4132. 10.1007/s10661-014-4132-8. [PubMed: 25487459]

Cao Y, and Hawkins CP. 2011. The comparability of bioassessments: a review of conceptual and methodological issues. Journal of the North American Benthological Society 30:680–701. 10.1899/10-067.1.

Cao Y, Larsen DP, Hughes RM, Angermeier PL, and Patton TM. 2002. Sampling effort affects multivariate comparisons of stream communities. Journal of the North American Benthological Society 21:701–714. 10.2307/1468440.

Carter JL, and Resh VH. 2001. After site selection and before data analysis: sampling, sorting, and laboratory procedures used in stream benthic macroinvertebrate monitoring programs by USA state agencies. Journal of the North American Benthological Society 20:658–682. 10.2307/1468095.

Carvalho DR, Leal CG, Junqueira NT, Castro MA, Fagundes DC, Alves CBM, Hughes RM, and Pompeu PS. 2017. A fish-based multimetric index for Brazilian savanna streams. Ecological Indicators 77:386–396. 10.1016/j.ecolind.2017.02.032.

Castro DMP, Dolédec S, and Callisto M. 2018. Land cover disturbance homogenizes aquatic insect functional structure in neotropical savanna streams. Ecological Indicators 84:573–582. 10.1016/j.ecolind.2017.09.030.

Chen K, Hughes RM, and Wang B. 2015. Effects of fixed-count size on macroinvertebrate richness, site separation, and bioassessment of Chinese monsoonal streams. Ecological Indicators 53:162–170. 10.1016/j.ecolind.2015.01.011.

Chen K, Hughes RM, Brito JG, Leal CG, Leitão RP, Oliveira-Júnior JMB, Oliveira VC, Dias-Silva K, Ferraz SFB, Ferreira J, Hamada N, Juen L, Nessimian J, Pompeu PS, and Zuanon J. 2017. A

multi-assemblage, multi-metric biological condition index for eastern Amazonia streams. Ecological Indicators 78:48–61. 10.1016/j.ecolind.2017.03.003.

Chessman BC, Thurtell LA, and Royal MJ. 2006. Bioassessment in a harsh environment: a comparison of macroinvertebrate assemblages at reference and assessment sites in an Australian inland river system. Environmental Monitoring and Assessment 119:303–330. 10.1007/s10661-005-9027-2. [PubMed: 16741821]

Chessman BC, Williams S, and Besley C. 2007. Bioassessment of streams with macroinvertebrates: effect of sampled habitat and taxonomic resolution. Journal of the North American Benthological Society 26:546–565. 10.1899/06-074.1.

Clarke RT, Wright JF, and Furse MT. 2003. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. Ecological Modelling 160:219–233. 10.1016/S0304-3800(02)00255-7.

Costa C, Ide S, and Simonka CE. 2006. Insetos Imaturos - Metamorfose e Identificação. Holos Editora, Ribeirão Preto, SP.

Courtemanch DL 1996. Commentary on the subsampling procedures used for rapid bioassessments. Journal of the North American Benthological Society 15:381–385. 10.2307/1467284.

Davies SP, and Jackson SK. 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. Ecological Applications 16:1251–1266. 10.1890/1051-0761(2006)016[1251:TBCGAD]2.0.CO;2. [PubMed: 16937795]

Doberstein CP, Karr JR, and Conquest LL. 2000. The effect of fixed-count subsampling on macroinvertebrate biomonitoring in small streams. Freshwater Biology 44:355–371. 10.1046/j.1365-2427.2000.00575.x.

Feio MJ, Vieira-Lanero R, Ferreira V, and Graça MAS. 2005. The role of the environment in the distribution and composition of Trichoptera assemblages in streams. Archiv für Hydrobiologie 164:493–512. 10.1127/0003-9136/2005/0164-0493.

Fernández HR, and Domínguez E. 2001. Guia para la determinación de los artrópodos bentônicos sudamericanos. Universidad Nacional de Tucumán, Tucumán.

Fierro P, Arismendi I, Hughes RM, Valdovinos C, and Jara-Flores A. 2018. A benthic macroinvertebrate multimetric index for Chilean Mediterranean streams. Ecological Indicators 91:13–23. 10.1016/j.ecolind.2018.03.074.

França JS, Gregório RS, De Paula JD, Gonçalves JF Jr., Ferreira FA, and Callisto M. 2009. Composition and dynamics of allochthonous organic matter imputs and benthic stock in a Brazilian stream. Marine and Freshwater Research 60:990–998. 10.1071/MF08247.

Françoso RD, Brandão R, Nogueira CC, Salmona YB, Machado RB, and Colli GR. 2015. Habitat loss and the effectiveness of protected areas in the Cerrado Biodiversity Hotspot. Natureza & Conservação 13:35–40. 10.1016/j.ncon.2015.04.001.

Gerth WJ, and Herlihy AT. 2006. The effect of sampling different habitat types in regional macroinvertebrate bioassessment surveys. Journal of the North American Benthological Society 25:501–512. 10.1899/0887-3593(2006)25[501:EOSDHT]2.0.CO;2.

Gjerløv C, and Richardson JS. 2004. Patchy resources in a heterogeneous environment: effects of leaf litter and forest cover on colonisation patterns of invertebrates in a British Columbian stream. Archiv für Hydrobiologie 161:307–327. 10.1127/0003-9136/2004/0161-0307.

Godoy BS, Faria APJ, Juen L, Sara L, and Oliveira LG. 2019. Taxonomic sufficiency and effects of environmental and spatial drivers on aquatic insect community. Ecological Indicators 107:105624. 10.1016/j.ecolind.2019.105624.

Gonçalves JF Jr., França JS, and Callisto M. 2006. Dynamics of Allochthonous Organic Matter in a Tropical Brazilian Headstream. Brazilian Archives of Biology and Technology 49:967–973. 10.1590/S1516-89132006000700014.

Gotelli NJ, and Cowell RK. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecology Letters 4:379–391. 10.1046/j.1461-0248.2001.00230.x.

Haapala A, Muotka T, and Laasonen P. 2003. Distribution of benthic macroinvertebrates and leaf litter in relation to streambed retentivity: Implications for headwater stream restoration. Boreal Environmental Research 8:19–30.

Hawkins CP, Norris RH, Hogue JN, and Feminella JW. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecological Applications 10:1456–1477. 10.1890/1051-0761(2000)010[1456:DAEOPM]2.0.CO;2.

Heino J 2009. Biodiversity of Aquatic Insects: Spatial Gradients and Environmental Correlates of Assemblage-Level Measures at Large Scales. Freshwater Reviews 2:1–29. 10.1608/FRJ-2.1.1.

Heino J, Melo AS, and Bini LM. 2015. Reconceptualising the beta diversity-environmental heterogeneity relationship in running water systems. Freshwater Biology 60:223–235. 10.1111/fwb.12502.

Heino J, Melo AS, Jyrkänkallio-Mikkola J, Petsch DK, Saito VS, Tolonen KT, Bini LM, Landeiro VL, Silva TSF, Pajunen V, Soininen J, and Siqueira T. 2018. Subtropical streams harbour higher genus richness and lower abundance of insects compared to boreal streams, but scale matters. Journal of Biogeography 45:1983–1993. 10.1111/jbi.13400.

Hering D, Feld CK, Moog O, and Ofenböck T. 2006. Cook book for the development of a multimetric index for biological condition of aquatic ecosystems: experiences from the European AQEM and STAR projects and related initiatives. Hydrobiologia 566:311–342. 10.1007/s10750-006-0087-2.

Herlihy AT, Sifneos JC, Hughes RM, Peck DV, and Mitchell RM. 2020. Relation of lotic fish and benthic macroinvertebrate condition indices to environmental factors across the conterminous USA. Ecological Indicators 112 (5):105958. 10.1016/j.ecolind.2019.105958.

Hewlett R 2000. Implications of taxonomic resolution and sample habitat for stream classification at a broad geographic scale. Journal of the North American Benthological Society 19:352–361. 10.2307/1468077.

Hughes RM, and Peck DV. 2008. Acquiring data for large aquatic resource surveys: the art of compromise among science, logistics, and reality. Journal of the North American Benthological Society 27:837–859. 10.1899/08-028.1.

Hurlbert SH 1971. The nonconcept of species diversity: a critique and alternative parameters. Ecology 52:577–586. 10.2307/1934145. [PubMed: 28973811]

Jackson DA, Peres-Neto PR, and Olden JD. 2001. What controls who is where in freshwater fish communities – the roles of biotic, abiotic and spatial factors. Canadian Journal of Fisheries and Aquatic Sciences 58:157–170. 10.1139/f00-239.

Kaufmann PR, Levine P, Robison EG, Seeliger C, and Peck DV. 1999. Quantifying Physical Habitat in Wadeable Streams. EPA/620/R-99/003. U.S. Environmental Protection Agency, Washington, DC.

Karr JR 1999. Defining and measuring river health. Freshwater Biology 41:221–234. 10.1046/j.1365-2427.1999.00427.x.

Karr JR, and Chu EW. 1999. Restoring life in running waters: better biological monitoring. Island Press, Washington, DC.

Kerans BL, Karr JR, and Ahlstedt SA. 1992. Aquatic invertebrate assemblages: spatial and temporal differences among sampling protocols. Journal of the North American Benthological Society 11:377–390. 10.2307/1467559.

King RS, and Richardson CJ. 2002. Evaluating subsampling approaches and macroinvertebrate taxonomic resolution for wetland bioassessment. Journal of the North American Benthological Society 21:150–171. 10.2307/1468306.

Klemm DJ, Blocksom KA, Fulk FA, Herlihy AT, Hughes RM, Kaufmann PR, Peck DV, Stoddard JL, Thoeny WT, Griffith MB, and Davis WS. 2003. Development and evaluation of a macroinvertebrate biotic integrity index (MBII) for regionally assessing Mid- Atlantic Highlands streams. Environmental Management 31:656–669. 10.1007/s00267-002-2945-7. [PubMed: 12719895]

Kobayashi S, and Kagaya T. 2005. Hot spots of leaf breakdown within a headwater stream reach: comparing breakdown rates among litter patch types with different macroinvertebrate assemblages. Freshwater Biology 50:921–929. 10.1111/j.1365-2427.2005.01371.x.

Larsen DP, and Herlihy AT. 1998. The Dilemma of Sampling Streams for Macroinvertebrate Richness. Journal of the North American Benthological Society 17:359–366. 10.2307/1468338.

Leal CG, Pompeu PS, Gardner TA, Leitão RP, Hughes RM, Kaufmann PR, Zuanon J, de Paula FR, Ferraz SFB, Thomson JR, Mac Nally R, Ferreira JN, and Barlow J. 2016. Multi-scale assessment

of human-induced changes to Amazonian instream habitats. Landscape Ecology 31:1725–1745. 10.1007/s10980-016-0358-x.

Leitão RP, Zuanon J, Mouillot D, Leal CG, Hughes RM, Kaufmann PR, Villéger S, Pompeu PS, Kasper D, de Paula FR, Ferraz SFB, and Gardner TA. 2017. Disentangling the pathways of land use impacts on the functional structure of fish assemblages in Amazon streams. Ecography 40:01–13. 10.1111/ecog.02845.

Ligeiro R, Moretti MS, Gonçalves JF Jr., and Callisto M. 2010. What is more important for invertebrate colonization in a stream with low-quality litter inputs: exposure time or leaf species? Hydrobiologia 654:125–136. 10.1007/s10750-010-0375-8.

Ligeiro R, Ferreira W, Hughes RM, and Callisto M. 2013a. The problem of using fixed-area subsampling methods to estimate macroinvertebrate richness: a case study with Neotropical stream data. Environmental Monitoring and Assessment 185:4077–4085. 10.1007/s10661-012-2850-3. [PubMed: 22930188]

Ligeiro R, Hughes RM, Kaufmann PR, Macedo DR, Firmiano KR, Ferreira WR, Oliveira D, Melo AS, and Callisto M. 2013b. Defining quantitative stream disturbance gradients and the additive role of habitat variation to explain macroinvertebrate taxa richness. Ecological Indicators 25:45–57. 10.1016/j.ecolind.2012.09.004.

Macedo DR, Hughes RM, Ligeiro R, Ferreira WR, Castro MA, Junqueira NT, Oliveira DR, Firmiano KR, Kaufmann PR, Pompeu PS, and Callisto M. 2014. The relative influence of catchment and site variables on fish and macroinvertebrate richness in cerrado biome streams. Landscape Ecology 29:1001–1016. 10.1007/s10980-014-0036-9.

Macedo DR, Hughes RM, Ferreira WR, Firmiano KR, Silva DRO, Ligeiro R, Kaufmann PR, and Callisto M. 2016. Development of a benthic macroinvertebrate multimetric index (MMI) for Neotropical Savanna headwater streams. Ecological Indicators 64:132–141. 10.1016/j.ecolind.2015.12.019.

Maloney KO, Munguia P, and Mitchell RM. 2011. Anthropogenic disturbance and landscape patterns affect diversity patterns of aquatic benthic macroinvertebrates. Journal of the North American Benthological Society 30:284–295. 10.1899/09-112.1.

Mardia KV, Kent JT, and Bibby JM. 1979. Multivariate analysis. First Edition, Academic Press. 518p.

Marshall JC, Steward AL, and Harch BD. 2006. Taxonomic resolution and quantification of freshwater macroinvertebrate samples from an Australian dryland river: the benefits and costs of using species. Hydrobiologia 572:171–194. 10.1007/s10750-005-9007-0.

Martins I, Macedo DR, Hughes RM, and Callisto M. 2020. Are multiple multimetric indices effective for assessing ecological condition in tropical basins? Ecological Indicators 110(3):105953. 10.1016/j.ecolind.2019.105953.

Melo AS 2005. Effects of taxonomic and numeric resolution on the ability to detect ecological patterns at a local scale using stream macroinvertebrates. Archiv für Hydrobiologie 164:309–323. 10.1127/0003-9136/2005/0164-0309.

Moretti MS, Gonçalves JF Jr., and Callisto M. 2007. Leaf breakdown in two tropical streams: differences between single and mixed species packs. Limnologica 37:250–58. 10.1016/j.limno.2007.01.003.

Mugnai R, Nessimian JL, and Baptista DF. 2010. Manual de identificação de macroinvertebrados aquáticos do Estado do Rio de Janeiro. Technical Books, Rio de Janeiro.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, and Kent J. 2000. Biodiversity hotspots for conservation priorities. Nature 403:853–858. 10.1038/35002501. [PubMed: 10706275]

Norris RH, and Thoms MC. 1999. What is river health? Freshwater Biology 41:197–211. 10.1046/j.1365-2427.1999.00425.x.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Henry M, Stevens H, Szoecs E, and Wagner H. 2018. Vegan: Community Ecology Package (R package version 2.5-3). Available at: http://CRAN.Rproject. org/package=vegan.

Oliveira AFM, and Marquis RJ. 2002. The Cerrados of Brazil. Columbia University Press, New York.

Olsen AR, and Peck DV. 2008. Survey design and extent estimates for the Wadeable Streams Assessment. Journal of the North American Benthological Society 27:822–836. DOI: 10.1899/08-050.1. 10.1899/08-050.1.

Ostermiller JD, and Hawkins CP. 2004. Effects of sampling error on bioassessments of stream ecosystems: application to RIVPACS-type models. Journal of the North American Benthological Society 23:363–382. 10.1899/0887-3593(2004)023<0363:EOSEOB>2.0.CO;2.

Paulsen SG, Mayio A, Peck DV, Stoddard JL, Tarquinio E, Holdsworth SM, Van Sickle J, Yuan LL, Hawkins CP, Herlihy A, Kaufmann PR, Barbour MT, Larsen DP, and Olsen AR. 2008. Condition of stream ecosystems in the US: an overview of the first national assessment. Journal of the North American Benthological Society 27:812–821. 10.1899/08-098.1.

Parsons M, and Norris RH. 1996. The effect of habitat specific sampling on biological assessment of water quality using a predictive model. Freshwater Biology 36:419–434. 10.1046/j.1365-2427.1996.00088.x.

Peck DV, Herlihy AT, Hill BH, Hughes RM, Kaufmann PR, Klemm DJ, Lazorchak JM, McCormick FH, Peterson SA, Ringold PL, Magee T, and Cappaert MR. 2006. Environmental Monitoring and Assessment Program-Surface Waters: Western Pilot Study field operations manual for wadeable streams. EPA/620/R-06/003. USEPA. Washington, DC.

Peres-Neto PR, and Jackson DA. 2001. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. Oecologia 129:169–178. 10.1007/s004420100720. [PubMed: 28547594]

Pérez GR 1988. Guía para el estudio de los macroinvertebrados acuáticos del Departamento de Antioquia. Fondo Fen. Colombia/Colciencias, Universidad de Antioquia, Colombia.

Plafkin JL, Barbour MT, Porter KD, Gross SK, and Hughes RM. 1989. Rapid bioassessment protocols for use in streams and rivers: benthic macroinvertebrates and fish. EPA/444/4–89–001. Office of Water, US Environmental Protection Agency, Washington, DC.

R Development Core Team. 2018. R: A Language and Environment for Statistical Computing (3.5.1). R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org.

Rawer-Jost C, Zenker A, Bohmer J. 2004. Reference conditions of German stream types analysed and revised with macroinvertebrate fauna. Limnologica 34:390–397. 10.1016/S0075-9511(04)80008-2.

Rehn AC, Ode PR, and Hawkins CP. 2007. Comparisons of targeted-riffle and reach-wide benthic macroinvertebrate samples: implications for data sharing in stream condition assessments. Journal of the North American Benthological Society 26:332–348. 10.1899/0887-3593(2007)26[332:COTARB]2.0.CO;2.

Reynoldson TB, Bombardier M, Donald DB, O'Neill H, Rosenberg DM, Shear H, Tuominen TM and Vaughan HH. 1999. Strategy for a Canadian Aquatic Biomonitoring Network. National Water Research Institute, Environment Canada, Burlington, ON. NWRI Contribution No. 99-248. 24 p.

Reynoldson TB, Norris RH, Resh VH, Day KE, and Rosenberg DM. 1997. The reference condition - a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. Journal of the North American Benthological Society 16:833–852. 10.2307/1468175.

Roy AH, Rosemond AD, Leigh DS, Paul MJ, and Wallace JB. 2003. Habitat-specific responses of stream insects to land cover disturbance: biological consequences and monitoring implications. Journal of the North American Benthological Society 22:292–307. 10.2307/1467999.

Ruaro R, and Gubiani EA. 2013. A scientometric assessment of 30 years of the Index of Biotic Integrity in aquatic ecosystems: Applications and main flaws. Ecological Indicators 29:105–110. 10.1016/j.ecolind.2012.12.016.

Ruaro R, Gubiani EA, Hughes RM, and Mormul RP. 2020. Global trends and challenges in multimetric indices of ecological condition. Ecological Indicators 110(3):105862. 10.1016/j.ecolind.2019.105862.

Sanches BO, Becker B, Hughes RM, Petesse ML, Ribeiro JR, and Santos GB. 2019. Fish-based multimetric index for evaluating land use effects on large neotropical reservoirs. Journal of Applied Ichthyology 35:1129–1140. 10.1111/jai.13954.

Sgarbi LF, Bini LM, Heino J, Jyrkänkallio-Mikkola J, Landeiro VL, Santos EP, Schneck F, Siqueira T, Soininen J, Tolonen KT, and Melo AS. 2020. Sampling effort and information quality provided by

rare and common species in estimating assemblage structure. Ecological Indicators 110: 105937. 10.1016/j.ecolind.2019.105937.

Silva DRO, Herlihy AT, Hughes RM, and Callisto M. 2017. An improved macroinvertebrate multimetric index for the assessment of wadeable streams in the neotropical savanna. Ecological Indicators 81:514–525. 10.1016/j.ecolind.2017.06.017.

Sovell LA, and Vondracek B. 1999. Evaluation of the Fixed-Count Method for Rapid Bioassessment Protocol III with Benthic Macroinvertebrate Metrics. Journal of the North American Benthological Society 18:420–426. 10.2307/1468455.

StatSoft, Inc. 2004. STATISTICA (data analysis software system), version 7. Available at:www.statsoft.com.

Stevens DL, and Olsen AR. 2003. Variance estimation for spatially balanced samples of environmental resources. Environmetrics 14:593–610. 10.1002/env.606.

Stoddard JL, Herlihy AT, Peck DV, Hughes RM, Whittier TR, and Tarquinio E. 2008. A process for creating multi-metric indices for large-scale aquatic surveys. Journal of the North American Benthological Society 27:878–891. 10.1899/08-053.1.

Stoddard JL, Peck DV, Olsen AR, Paulsen SG, Van Sickle J, Herlihy AT, Kaufmann PR, Hughes RM, Whittier TR, Lomnicky G, Larsen DP, Peterson SA, Ringold PL. 2005. An Ecological Assessment of Western Streams and Rivers. U.S. Environmental Protection Agency and Dynamic Corporation, Corvallis, OR.

Strahler AN 1957. Quantitative analysis of watershed geomorphology. Transactions American Geophysical Union 38:913–920. 10.1029/TR038i006p00913.

Terra BF, Hughes RM, Francelino MR, and Araújo FG. 2013. Assessment of biotic condition of Atlantic Rain Forest streams: A fish-based multimetric approach. Ecological Indicators 34:136–148. 10.1016/j.ecolind.2013.05.001.

Turak E, Flack LK, Norris RH, Simpson J, and Waddell N. 1999. Assessment of river condition at a large spatial scale using predictive models. Freshwater Biology 41:283–298. 10.1046/j.1365-2427.1999.00431.x.

Vinson MR, and Hawkins CP. 1996. Effects of sampling area and subsampling procedure on comparisons of taxa richness among streams. Journal of the North American Benthological Society 15:392–399. 10.2307/1467286.

Waite IR, Kennen JG, May JT, Brown LR, Cuffney TF, Jones KA, and Orlando JL. 2012. Comparison of stream invertebrate response models for bioassessment metrics. Journal of the American Water Resources Association 48:570–583. 10.1111/j.1752-1688.2011.00632.x.

Walsh CJ 1997. A multivariate method for determining optimal subsample size in the analysis of macroinvertebrate samples. Marine and Freshwater Research 48:241–248. 10.1071/MF96096.

Wantzen KM, Siqueira A, Nunes Da Cunha C, and Sa MFP. 2006. Stream–valley systems of the Brazilian cerrado: impact assessment and conservation scheme. Aquatic Conservation: Marine and Freshwater Ecosystems 16:713–732. 10.1002/aqc.807.

Warton DI, and Hui FKC. 2011. The arcsine is asinine: the analysis of proportions in ecology. Ecology 92:3–10. 10.1890/10-0340.1. [PubMed: 21560670]

Wells F, Metzeling L, and Newall P. 2002. Macroinvertebrate regionalisation for use in the management of aquatic ecosystems in Victoria, Australia. Environmental Monitoring and Assessment 74:271–294. 10.1023/A:1014235211968. [PubMed: 11944800]

Whittier TR, and Van Sickle J. 2010. Macroinvertebrate tolerance values and an assemblage tolerance index (ATI) for western USA streams and rivers. Journal of the North American Benthological Society 29:852–866. 10.1899/09-160.1.

Whittier TR, Stoddard JL, Larsen DP, and Herlihy AT. 2007. Selecting reference sites for stream biological assessments: best professional judgment or objective criteria. Journal of the North American Benthological Society 26:349–360. 10.1899/0887-3593(2007)26[349:SRSFSB]2.0.CO;2.

Wiens JA 2002. Riverine landscapes: taking landscape ecology into the water. Freshwater Biology 47:501–515. 10.1046/j.1365-2427.2002.00887.x.

Wright JF 1995. Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. Australian Journal of Ecology 20:181–197. 10.1111/j.1442-9993.1995.tb00531.x.

Wu J, and Loucks OL. 1995. From balance of nature to hierarchical patch dynamics: a paradigm shift in ecology. The Quarterly Review of Biology 70:439–466. 10.1086/419172.

Zar JH 2010. Biostatistical Analysis. Fifth Edition, Prentice-Hall/Pearson, 944p.

Highlights

- We examined how combinations of methods detected a known anthropogenic disturbance gradient.
- Multihabitat sampling performed best when total counts of individuals were employed.
- The responsiveness of targeted sampling on leaf packs was increased after subsampling.
- Leaf packs behaved as ‘biodiversity hotbeds’ in highly disturbed sites.
- Macroinvertebrate responsiveness to anthropogenic disturbances depends on assessment methodologies and on their combinations.

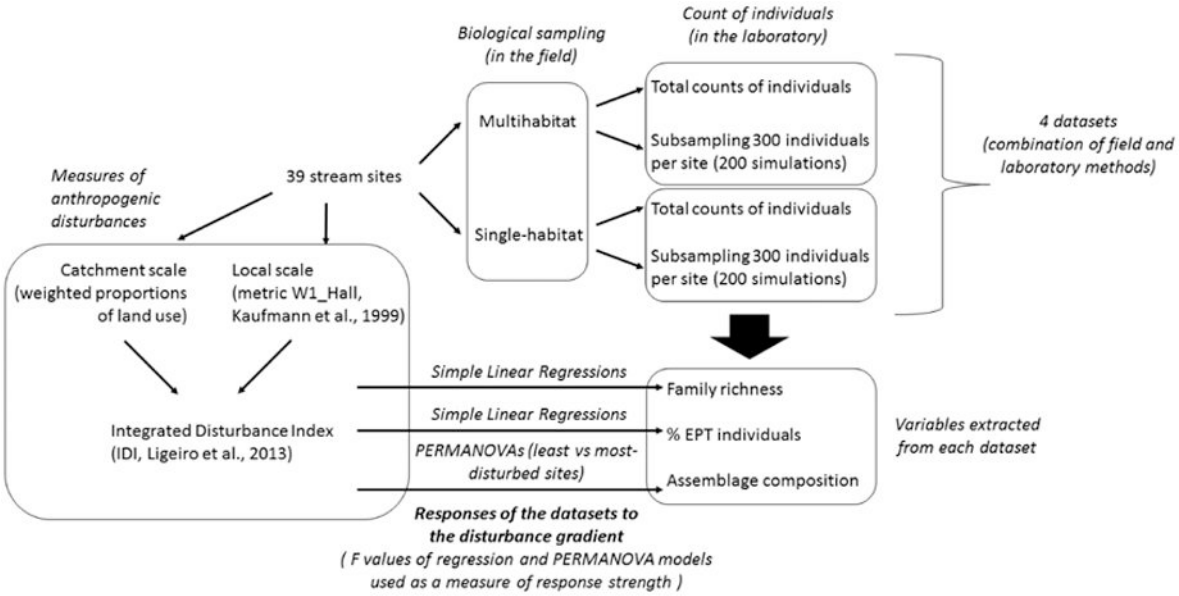

Figure 1.

Analytical design to assess the response strength of macroinvertebrate assemblages to a known anthropogenic disturbance gradient, comparing four combinations of field (multihabitat vs single-habitat) and laboratory (total counts of individuals vs subsampling) methods.

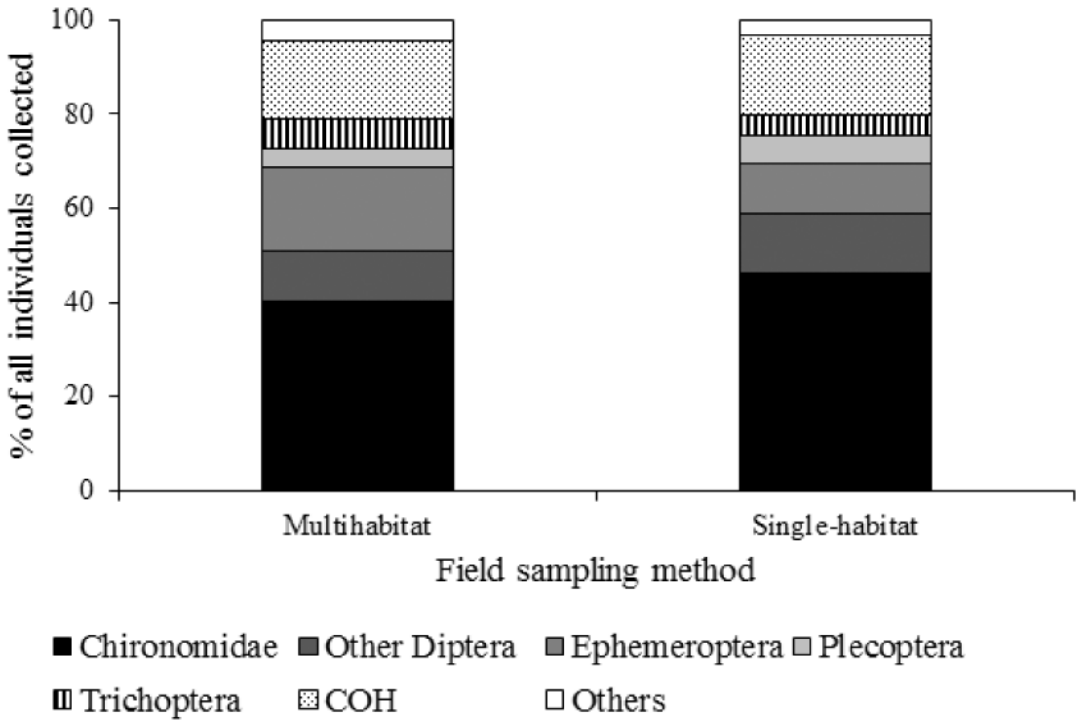

Figure 2.

Relative abundance of major groups of macroinvertebrates observed in multihabitat (MH) and single-habitat (SH) field sampling methods (sum of all individuals found in all stream sites). COH = Coleoptera+Odonata+Heteroptera.

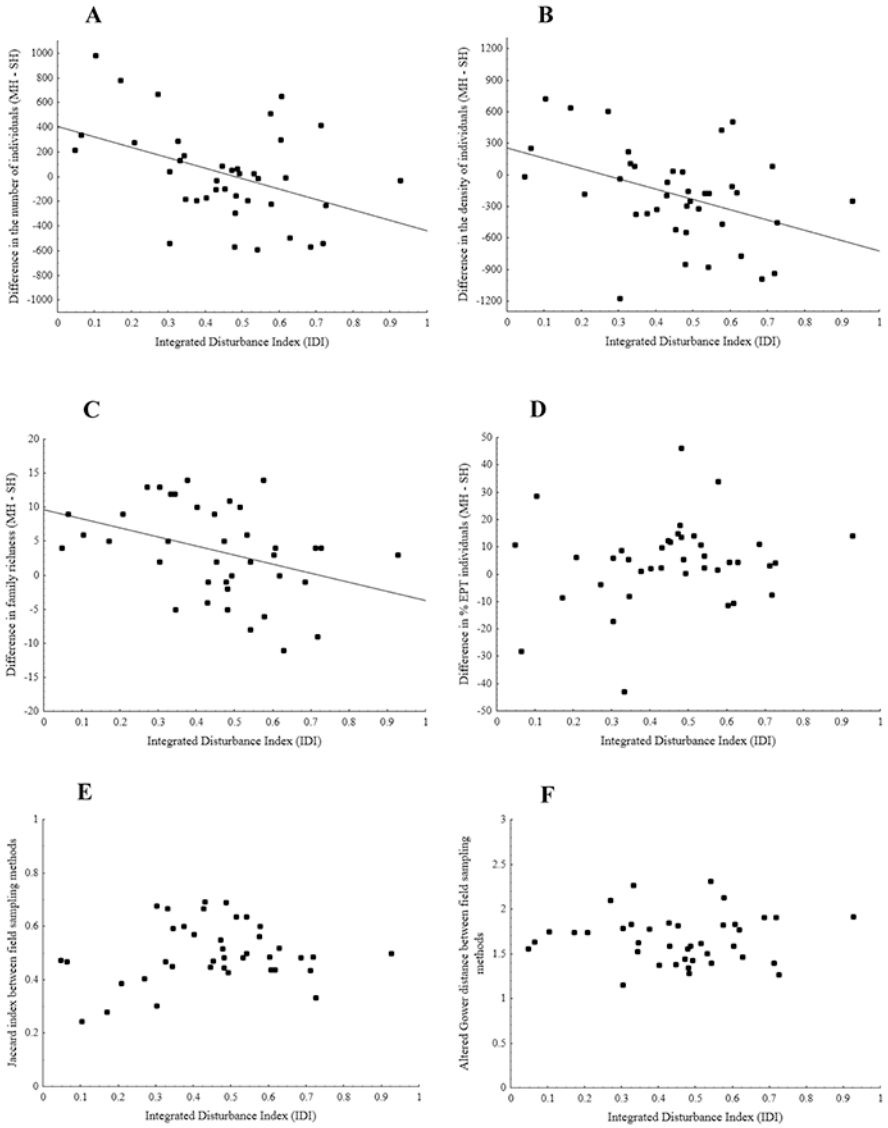

Figure 3.

Relationship of the integrated disturbance index (IDI) of the stream sites and differences in the number (A) and density (B) of individuals, family richness (C), % EPT individuals (D), Jaccard index (E) and altered Gower distance (F) between samples obtained from multihabitat and single-habitat field sampling methods (difference = MH - SH), considering total counts of individuals.

Table 1.

Comparisons between macroinvertebrate assemblage variables obtained with multihabitat (MH) and single-habitat (SH) field sampling methods, showing averages (± standard errors) and paired t-test results (38 degrees of freedom). *p* values < 0.05 are followed by an asterisk.

Variable	Sampling method		Statistics (paired t-tests)	
	MH	SH	*t* value	*p* value
Number of individuals	573 (± 68)	551 (± 58)	0.03	0.974
Density (ind./m^2)	579 (± 69)	766 (± 80)	− 2.63	0.012 *
Number of families	24 (± 1)	21 (± 1)	3.29	0.002 *
% EPT individuals	26 (± 3)	21 (± 2)	1.90	0.065

Table 2.

Results of simple linear regression models of macroinvertebrate family richness and % EPT individuals (logit transformed) against the Integrated Disturbance Index (IDI) of the stream sites. The four datasets compared were generated from two different field sampling methods; 1) multihabitat (MH), 2) single-habitat (SH), and two laboratory procedures; A) total count of individuals of the samples, and B) subsampling 300 individuals from the samples. For the subsampled datasets (200 simulations per field sampling method), we show the mean F values and the proportion of significant regression models (which represented $p < 0.05$). We compared the single F values obtained from the total counts with the respective 200 F values obtained from the simulations through a standardized measurement of differentiation (Z values). Degrees of freedom for F statistics were 1, 37. p values < 0.05 are followed by an asterisk.

		Total counts		Subsampling (300 individuals)		
Metric	**Sampling method**	***F* value**	***p* value**	**Mean *F* values**	**% significant models**	***Z* value**
Family richness	MH	14.76	< 0.001 *	12.93	100.0	0.67
	SH	1.87	0.180	6.50	92.5	−2.41
% EPT individuals	MH	3.65	0.064	3.58	8.5	0.18
	SH	14.99	< 0.001 *	15.00	100.0	−0.01

Table 3.

Results of PERMANOVA models performed between least- and most-disturbed stream sites using the Jaccard index and the altered Gower distance. The four datasets compared were generated from two different field sampling methods; 1) multihabitat (MH), 2) single-habitat (SH), and two laboratory procedures; A) total count of individuals of the sites, and B) subsampling 300 individuals per site. For the subsampled datasets (200 simulations per field sampling method), we show the mean *F* values and the proportion of significant PERMANOVA models (which presented $p < 0.05$). We employed 10,000 randomizations in each PERMANOVA to test model significance. We compared the single *F* values obtained from the total counts with the respective 200 *F* values obtained from the simulations through use of a standardized measurement of differentiation (Z values). Degrees of freedom of PERMANOVAs were 1, 37. *p* values < 0.05 are followed by an asterisk.

		Total counts		Subsampling (300 individuals)		
Dissimilarity measure	**Sampling method**	***F* value**	***p* value**	**Mean *F* values**	**% significant models**	***Z* value**
Jaccard index	MH	2.73	0.002 *	1.61	48.0	4.39
	SH	1.95	0.010 *	1.76	69.5	0.62
Altered Gower distance	MH	3.60	0.003 *	1.74	87.0	12.52
	SH	2.05	0.005 *	2.63	100.0	−2.95