



Refining benthic macroinvertebrate kick sampling protocol for wadeable rivers and streams in Ethiopia

Melaku Getachew · Worku Legesse Mulat ·
Seid Tiku Mereta · Geremew Sahilu Gebrie ·
Mary Kelly-Quinn

Received: 17 June 2021 / Accepted: 5 November 2021 / Published online: 17 February 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract Streams and rivers cover a larger proportion of the Earth's surface but are highly affected by human pressures. Conversely, bioassessment methods are in their infancy in developing countries such as Ethiopia. In this study, we compared 2- and 3-min macroinvertebrate kick samples at multiple locations for both riffle habitat (RH) and multihabitat (MH) approaches. The performance of each method was evaluated statistically using benthic macroinvertebrate metrics and diversity indices. Results of the Kruskal–Wallis analysis in this study showed no significant differences among methods tested in minimally impacted streams in Ethiopia and generally

performed equally irrespective of the methods employed except for total abundances and Ephemeroptera abundances. Furthermore, multivariate analysis of the relative abundances of macroinvertebrate communities using analysis of similarity (ANOSIM), RELATE, non-metric multidimensional scaling (MDS), and classification strength-sampling method comparability (CS-SMC) indicated a high similarity in the macroinvertebrate communities recorded among all methods employed in this study area. However, the index of multivariate dispersion (IMD) test showed variations in relative abundances of macroinvertebrate communities among the methods. In summary, if the focus is not on rare taxa and the required information is not dependent on additional evidence provided by the use of lower taxonomic levels of identification (genus and species), the results of the present study support the use of the shorter 2-min RH kick sampling method for the bioassessment of wadeable rivers and streams in Ethiopia.

M. Getachew (✉)
Ethiopian Institute of Water Resources, Addis Ababa
University, Addis Ababa, Ethiopia
e-mail: melakugetachew@gmail.com

M. Getachew · W. L. Mulat
Department of Environmental Health, Wollo University,
Dessie, Ethiopia

S. T. Mereta
Department of Environmental Health Sciences
and Technology, Jimma University, Jimma, Ethiopia

G. S. Gebrie
School of Civil and Environmental Engineering, Addis
Ababa Institute of Technology, Addis Ababa University,
Addis Ababa, Ethiopia

M. Kelly-Quinn
School of Biology and Environmental Science, University
College Dublin, UCD Earth Institute, Dublin, Ireland

Keywords Bioassessment of streams and rivers ·
Fixed sampling time · Single and multihabitat
sampling · Macroinvertebrates · Ethiopia

Introduction

Highland streams in Ethiopia have been reported to show deteriorated water quality from severe

anthropogenic impacts (Alemneh et al., 2017) including poor waste management practices, urbanization, population growth, industrialization, unregulated agriculture, floriculture farms that use large quantities of fertilizers and pesticides as well as deforestation (Getachew et al., 2020). A sampling strategy to monitor water quality must be capable of distinguishing anthropogenic impacts from natural variation (Callanan et al., 2008). Natural variation in biological community structure occurs in the river continuum especially across ecoregions in response to changes in altitude (Wang et al., 2012) and seasonal change (Alemneh et al., 2019; Gratwicke, 1998). At the same time, the delivery of pollutants to rivers and streams may be from a point or diffuse sources as chronic or acute inputs. According to Buss et al. (2015) several factors also dictate the limitations to adopting standardized biomonitoring protocols: (1) Climatic differences among countries create differences in season, month, or flow condition for sampling in all regions; (2) lack of taxonomic expertise to identify specimens to genus or species level; and (3) the fact that most biological indicators and indices have been developed for specific geographic regions, states, or countries, as well as for specific pressures.

Inferring water quality and in particular ecological water quality from biological communities is well established (Geist, 2011; Resh & Rosenberg, 1993). Biological sampling protocols have been established for most developed regions such as America, Europe, and Australia (Barbour et al., 1999; Koester & Gergs, 2017; Stark et al., 2001) based on their ecological contexts but often not consistently applied at local, national, or continental scales—even in their home countries (Buss et al., 2015). The protocols follow several levels of sampling effort that can be categorized into three groups (Buss et al., 2015) namely: fixed sample number, fixed sampling length/area, and fixed sampling time.

Benthic macroinvertebrates are among the most widely used globally for the assessment of water quality as they reflect prevailing conditions, respond rapidly to environmental stresses, and are often the first ecological indicators to react to changes in the environment (Barbour et al., 1999). For example, Carlisle et al. (2008) showed that the macroinvertebrate model was the most precise compared to the diatom and fish models. However, the use of macroinvertebrates has been criticized due to long-time requirements and

expensive costs of sampling, sorting, counting, and identifying them (De Bikuna et al., 2015) and this is likely to affect the effectiveness of biomonitoring protocols using macroinvertebrate communities (Pinna et al., 2014).

Among the most commonly applied methods for bioassessment using macroinvertebrates is kick sampling. However, the approach to kick sampling is variable especially in terms of timing; examples include 2-min (Mandaville, 2002), 3-min (Larsen et al., 2012), 4-min (Haase, et al., 2004a, b), and 5-min (Cheshmedjiev et al., 2011; Luo et al., 2018) to collect samples of macroinvertebrate communities at each location. Conversely, there is often little attention given to the duration the “kicking” should be conducted to take a sample (Bradley & Ormerod, 2002; Clarke et al., 2002). Furthermore, there is variability in terms of the mesohabitats (e.g., riffle, run/glide, pool) targeted for sampling. Both single habitat (Blocksom et al., 2008) and multihabitat (Barbour et al., 2006) approaches have been adopted. However, most developing countries including Ethiopia have no agreed macroinvertebrate kick sampling protocols. So far, several bioassessment studies conducted in Ethiopia have used various kick sampling methods, for example, 3-min (Getachew et al., 2012; Lakew & Moog, 2015) (most frequently used method), 5-min (Ambelu et al., 2014), and even 10-min (Mereta et al., 2012, 2013) in different parts of the country without any subsampling techniques for benthic invertebrates. There is no consensus on whether a standardized approach should be adopted for monitoring wadeable rivers and streams in the country in terms of the kick sampling method. This was the impetus for the current research that undertook to test several approaches to kick sampling.

In identifying the approach that might be most appropriate we considered both processing time and costs. As the kick time increases, the cost of sample processing also increases unless subsampling is used. For example, Haase et al. (2004a, b) explained that sorting is often the most time-consuming aspect of a bioassessment study that led to either economic or deliverability consequences. Any decrease in time spent on sample processing while not decreasing sampling effectiveness will increase productivity and reduce costs. For example, Feeley et al. (2012) compared two multihabitat (MH) kick sampling methods in small streams and found that a kick sampling

method as short as 20-s was as effective as 60-s to produce a reasonable representation of the taxa present. Mykra et al. (2006) also found that there was no indication of differences in community structure among different sample sets collected for 1-, 2-, 3-, and 4-min, and suggested that even 1-min kick samples can adequately represent macroinvertebrate community structure in streams for that particular ecoregion.

Decreasing effort and cost in developing a protocol is not the only aim of sampling methods. It should also maintain accuracy to answer research or monitoring questions (Barbour & Gerritsen, 1996). This needs effective yet rapid sampling and processing methods. However, there is a scarcity of studies in Ethiopia that have evaluated different kick sampling methods in rivers and streams. As previously explained, studies in the country used different kick sampling methods varying from 3- to 10-min, all of which have been developed elsewhere in different ecoregions such as non-tropical countries (Barbour et al., 1999; Gabriels et al., 2010; Ostermiller & Hawkins, 2004) without determining the habitat type to sample. The 2-min kick sampling method has not been used in previous bioassessment studies in Ethiopia; however, it has been effectively used in different ecoregions (Bradley & Ormerod, 2002; Heino et al., 2002, 2003). Therefore, the objective of this study was to determine whether there are differences in four sampling approaches; MH kick samples collected for 2- and 3-min and riffle habitat (RH) kick samples collected for 2- and 3-min at minimally impacted sites (Table 2). The study tested the hypothesis that “no differences occur in terms of a range of macroinvertebrate metrics as well as community structure between samples collected using the four methods.”

Methods and materials

Study areas

Ethiopia has a diverse climate due to its equatorial positioning and varied topography (Block, 2008) ranging from a semi-arid desert to a humid and warm (temperate) type in southwest Ethiopia (Beyene & Meissner, 2010) where this study was conducted. Traditionally climatic conditions in Ethiopia are classified into five major agroecological zones based on the altitude and temperature variations (Berhanu et al., 2014; MOA, 1998) (Table 1). The study sites for this study are located in the “Kola” and “Weynadega” agroecological zones having nearly 80% share of the country’s surface area (MOA, 1998).

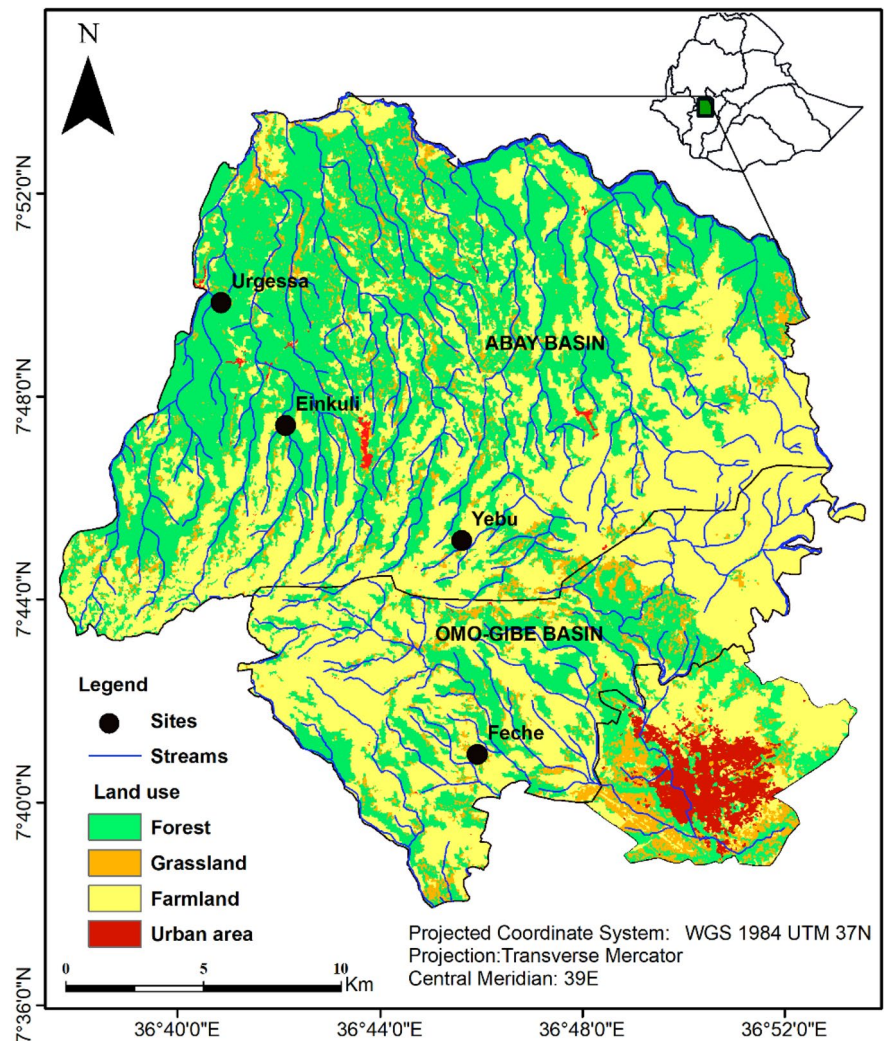
Ethiopia has twelve major river basins (Awulachew et al., 2007). The study sites were located in the Abay and Omo-Gibe river basins (Fig. 1) based on their physical characteristics (e.g., slope, soil type, stream order, and land use types) and known ecological conditions (Alemneh et al., 2019; Lakew, 2017) to reflect typical Ethiopian wadeable rivers and streams.

Twenty reaches on four reference streams (Urgessa, Yebu, Enkulu, and Feche) with minimal anthropogenic pressures were selected using the standards representing minimally impacted waters (Table 2). The collection of macroinvertebrate samples within designated index periods is critical for repeated assessments (KDOW, 2015) and to minimize seasonal variation (Montana, 2012). We selected the dry season as an index period to collect the samples when water levels are sufficiently low to enable sampling. All samples were collected in the same year (2019) in one month (May). Previous bioassessment studies in Ethiopia have also been conducted in this season (mid-September to mid-June) (Ambelu et al., 2013; Getachew et al., 2012). Prior to field sampling

Table 1 Traditional climatic zones of Ethiopia and their physical characteristics. Source: (Berhanu et al., 2014; MOA, 1998)

Agroecological zones	Altitude (m)	Mean annual rainfall (mm)	Mean annual temperature (°C)	Area share (%)
Wurch (cold to moist)	> 3200	900–2,200	< 11.5	0.98
Dega (cool to humid)	2300–3200	900–1,200	11.5–17.5	9.94
Weynadega (sub humid)	1500–2300	800–1,200	17.5–20.0	26.75
Kola (warm semiarid)	500–1500	200–800	20.0–27.5	52.94
Bereha (hot arid)	< 500	Below 200	> 27.5	9.39

Fig. 1 Land use types and location of study sites in the Abay and Omo-Gibe River basins of Ethiopia (Source: ArcGIS 10.5)



background information on land use, geology, and soils where available were compiled (Table 2).

The coordinates and altitudes were recorded using a global positioning system (GERMIN-72H). Dissolved oxygen (mg/L), pH, water temperature ($^{\circ}\text{C}$), conductivity ($\mu\text{S}/\text{cm}$ @ 25°C), total dissolved solids (mg/L) were measured in-situ using a digital handheld portable multi-parameter (Hach HQD) probe. Habitat assessment (representation of flow mesohabitats and benthic substrates) was also conducted using the rapid bioassessment protocol set for wadeable streams and rivers (Barbour et al., 1999). Nutrients such as nitrate (mg/L $\text{NO}_3\text{-N}$) and phosphate (mg/L) were measured using a screen touch spectrophotometer in the Department of Environmental Health Sciences and Technology laboratory of Jimma University.

Sampling method evaluation

We first selected an accessible length of the stream at each site to locate five 50 m reach (250 m in total on each river) for sampling after Feeley et al. (2012) and Mabidi et al. (2017). From these stream reaches, 15 samples for each 2- and 3-min MH method were collected from all mesohabitats including riffles, run/glides, and pools in proportion to their representation in each of the 50-m stream reach. Similarly, 5 samples of 2- and 3-min duration were collected from riffle habitats only within similar reaches where the MH samples were collected. There was no overlap in the sampling area covered by each sampling method. For all methods, sampling was started from downstream and continued upstream to avoid habitat disturbance

Table 2 Details of streams selected for the collection of macroinvertebrate samples in southwestern Ethiopia, 2019

Variables	Streams				Standard	References
	Urgessa	Yebu	Einkuli	Feche		
Longitude	36.86	36.723	36.682	36.78	–	–
Latitude	7.7	7.77	7.80	7.68	–	–
Altitude (m.a.s.l)	1487	2231	1710	1723	–	–
Stream order	2	2	2	3	–	–
Soil type (all Nitrosols-Nitr)	Nitr	Nitr	Nitr	Nitr	–	–
Point source pollution	None	None	None	None	None	Sanchez-Montoya et al. (2012), p.47
% Urbanization	None	None	None	None	<0.4%	Lakew (2017), p.951
Sand/gravel extraction	None	None	None	None	None	Sanchez-Montoya et al. (2012), p.47
Intensive use of grazing	None	None	None	None	None	Sanchez-Montoya et al. (2012), p.47
Water diversion and dam	None	None	None	None	None	Alemneh et al. (2019), p.129
Dry land farming	10%	5%	10%	15%	<20%	Sanchez-Montoya et al. (2012), p.47
Tree removal	5%	None	5%	8%	<10%	Alemneh et al. (2017), p.454
Electrical conductivity (µS/cm)	65	64	59	68	45–265	Lakew (2017), p.951
Total dissolved solids (mg/L)	46	45	41	48	<500	Oluyemi et al. (2010), p.692
pH	7.49	7.65	6.55	6.85	7.3–8.2	Sanchez-Montoya et al. (2012), p.47
Dissolved oxygen (DO) (mg/L)	8.41	7.73	7.45	7.26	6.8–9.4	Lakew (2017), p.951
% DO saturation	98	98	99	95	95–120	Lakew (2017), p.951
Phosphate (mg/L)	0.013	0.021	0.017	0.024	0.01–0.03	Fadiran et al. (2008), p.201
Nitrate-nitrogen (NO ₃ -N) (mg/L)	0.35	0.67	0.42	0.78	2.0	Camargo et al. (2005), p.1264

prior to sampling. A total of 120 MH replicates (60 for each method) and 40 RH replicates (20 for each method) were collected using a D-frame pond net (0.5 mm mesh). Immediately after the collection of benthic macroinvertebrates, large leaves, sticks, and stones were individually rinsed and inspected for organisms and discarded at the site. The entire sample was then preserved in 96% ethanol alcohol after Barbour et al. (1999) and transported to the laboratory for sorting and identification to the family level and counting the various taxa. Wright et al. (2000) highlighted that family-level data are adequate for rapid bioassessment of water quality. Furthermore, Barbour et al. (1999) showed that family-level identifications provide a high degree of precision among samples, require less expertise to perform, and accelerate the production of results. However, the necessary level of taxonomic resolution is determined depending on the purpose(s) of a study (Bailey et al., 2001). Fundamentally, most biomonitoring programs acknowledged a tradeoff between efficiency and sensitivity for large-scale monitoring, thus many strategies for reducing processing time and cost have been implemented (Buss et al., 2015).

Data and statistical analysis

The non-parametric Kruskal–Wallis test was used to test for significant differences in a range of metrics among the four kick sampling methods (2- and 3-min RH and 2- and 3-min MH samples). The analyzed metrics include taxonomic richness at the family level, total abundance, Ephemeroptera, Plecoptera, and Trichoptera (EPT) richness, %EPT, Ephemeroptera richness, Ephemeroptera abundance, family biotic index (FBI), the Biological Monitoring Working Party (BMWP) score, Average Score Per Taxon (ASPT) (Hawkes, 1998), benthic macroinvertebrates based on the biotic score (ETHbios), and ASPT calculated from ETHbios (Aschalew & Moog, 2015). Also, the Shannon–Wiener’s index, evenness, Simpson’s index, %scrapers, %shredders, %collector-gatherers, %collectors-filterer, and %predators were used to evaluate the four kick sampling methods. The post hoc Mann–Whitney test was calculated to evaluate pairwise differences between the groups. The Bonferroni corrected *p*-values were calculated in PAST statistical software to reduce type I error, or false rejection of the null hypothesis (i.e.,

there is no difference between the kick sampling methods). The bar charts with standard error of the mean (Cumming et al., 2007) were also displayed for selected metrics (richness, total abundance, and Ephemeroptera abundance).

We used the relative abundances of macroinvertebrate communities to carry out the multivariate tests to control for the differences in kick net, sampling times, and habitat differences. Analysis of similarity (ANOSIM), a randomization test (Clarke & Green, 1988), was performed on the resemblance matrix based on the Bray–Curtis similarity coefficient (Armitage et al., 1987; Clarke & Gorley, 2006). This was carried out in PRIMER 6, based on $\log(x + 1)$ transformed macroinvertebrate community relative abundance data using 999 permutations. It allows a test of the null hypothesis that there are no macroinvertebrate assemblage differences among groups of replicated samples collected by different kick sampling methods (Clarke & Gorley, 2006). The differences among the groups were measured by the global R statistic, which is calculated as $R = \frac{4(\bar{r}_B - \bar{r}_W)}{n(n-1)}$, where \bar{r}_B and \bar{r}_W are the mean rank between-group similarity and within-group similarity, respectively, and n is the total number of samples (Cao et al., 2005). The ANOSIM calculates a test statistic R that ranges from -1 to 1 (Chapman & Underwood, 1999). An R -value close to 1 indicates good separation or differences exist among groups and a value close to 0 indicates weak separation or complete random grouping (Chapman & Underwood, 1999; Clarke & Gorley, 2006) and the negative values occur when the samples within a group are less similar to one another than to the samples of other groups, probably due to inappropriate sampling designs (Chapman & Underwood, 1999). ANOSIM provides associated p -values to R -statistics which highlights the significance level of the test (Clarke & Gorley, 2006).

The two-dimensional MDS ordination plot which is a visual representation of differences in macroinvertebrate structures among the methods was also constructed using a relative abundance matrix. An MDS is a rank-based approach where the relative abundance data are substituted with ranks (Buttigieg & Ramette, 2014). An ordination plot with stress values equal to or below 0.05 indicated a good fit and values > 0.2 are arbitrary or more of a random pattern indicating that there is a little

explainable pattern in the MDS plot (Clarke, 1993). To elucidate the differences in macroinvertebrate distribution in the MDS plot, an index of multivariate dispersion (MVDISP) test was then employed to statistically measure the relative rank dissimilarity between the samples within each method. Correspondingly, the within-group similarity was also determined for samples within each method using SIMPER analysis according to the method described in Clarke and Gorley (2006).

Furthermore, to compare the patterns of macroinvertebrates community structure among different kick sampling methods, we used the RELATE routine in PRIMER 6 (Clarke & Gorley, 2006) which compares the relationship between two different resemblance matrices or MDS plots by calculating Spearman's rank correlation coefficients ρ (rho) from relative abundances of macroinvertebrates based on 999 permutations to test the (dis)similarity of the methods. A Spearman's rank correlation coefficient ($\rho = 1$) indicates a perfect similarity between method pairs.

Finally, classification strength-sampling method comparability (CS-SMC) (Cao et al., 2005) was computed based on the formula proposed by several authors (Cao & Hawkins, 2011; Cao et al., 2005; Van Sickle, 1997). The CS-SMC allows the direct comparison of benthic macroinvertebrate community structure similarity obtained by kick sampling methods (Cao et al., 2005). All six pairs of the sampling methods were compared from the $\log(x + 1)$ transformed Bray–Curtis similarities of taxonomic assemblages (SIMPER) to calculate the CS-SMC using the following equation (Cao et al., 2005; Van Sickle, 1997): $CS - SMC = \frac{2\bar{S}_b}{\bar{S}_{w1} + \bar{S}_{w2}} \times 100$, where \bar{S}_b denotes the mean between-group similarity, \bar{S}_{w1} and \bar{S}_{w2} denotes the mean within-group similarity for any possible pairs of macroinvertebrate kick sampling methods from MH and RH samples. The CS-SMC measures similarity between methods relative to that between replicates within a method Cao et al. (2005) and indicates that methods are similar if the average within-method similarity is equal to the average within-method similarity of the other method. The CS-SMC is a useful measure of comparability as it is independent of site differences and sampling effort (Cao et al., 2005; Feeley et al., 2012).

Results

A total of 56,930 macroinvertebrates belonging to 78 families and 15 orders were collected using the four methods. Ephemeroptera 29,817 (52.37%), Coleoptera 6,243 (10.97%), Trichoptera 6150 (10.80%), and Diptera 5,405 (9.47%) were the four most abundant orders of macroinvertebrates present. The various metrics calculated are given in Table 3 below.

The Kruskal–Wallis test did not show any significant difference (Fig. 2, Table 3) among the methods for all metrics except total abundance ($H=42.4$, $p=0.001$, $df=3$) and Ephemeroptera abundance ($H=19.34$, $p=0.001$, $df=3$) (Fig. 3, Table 3). For example, the mean (\pm standard error) for richness (Fig. 2) did not show any significant difference among the methods.

For significantly different metrics, the pairwise post hoc test indicated significant differences in total abundance among all method pairs except 2-min RH and the 2-min MH pair (Fig. 3, Table 4).

Similarly, the post hoc test on Ephemeroptera abundance showed significant differences between 2- and 3-min RH, 2-min MH and 3-min RH, and 2- and 3-min MH (Fig. 3, Table 4).

The ANOSIM test based on the relative abundance of macroinvertebrates also showed no significant difference among the four kick sampling methods ($R=0.048$, $p=0.972$, ANOSIM). Furthermore, the two-way ANOSIM test, sites versus methods, did not show any significant difference ($p=0.214$) although the global R was very low ($R=0.022$; ANOSIM). Similarly, the MDS ordination plot (Fig. 4) from the relative abundance of macroinvertebrates with a high stress value of 0.24 (Clarke, 1993) as well showed no separation of the samples according to the method applied.

However, the MVDISP (index of multivariate dispersion) given in Table 5 showed the variations in the relative abundances of macroinvertebrate communities among the methods. The dispersion sequence of 0.34, 0.341, 0.945, and 1.197 for 3-min RH, 2-min RH, 3-min MH, and 2-min MH, respectively, showed that the average rank dissimilarity is almost 3.5 times higher within 2-min MH samples than for 2- and 3-min RH samples. The 2- and 3-min RH replicates were almost similar and the least dispersed compared to both 2- and 3-min multihabitat replicates. Likewise, the similarity percentages were similar for

2- and 3-min riffle habitats (Table 5). The lower the MVDISP value, the less dispersed the samples within the factor, and the greater the SIMPER, the more similar the samples within the factor (Wasserman et al., 2015).

The RELATE analysis computed from the relative abundance of macroinvertebrate communities further indicated a similarity in macroinvertebrate community relative abundances recorded between all pairs of methods with a Spearman rank correlation coefficient (ρ) of greater than 0.5 and $p=0.001$. Among all pairs, 2- and 3-min RH samples and 2- and 3-min MH samples showed the highest similarity in macroinvertebrate community structures recorded (Table 6).

Finally, the analysis of similarities between benthic macroinvertebrates community structure had a classification strength-sampling method comparability (CS-SMC) close to 100% for all pairs of methods (Fig. 5) which highlighted that the similarity of samples between methods was very high. The trend line in Fig. 5 illustrated that method pairs such as 2- and 3-min RH, 2-min RH, and 2-min MH, and 2-min MH and 3-min RH had better CS-SMC scores (close to 100%) compared to the other method pairs. Method pairs between 2-min MH and 3-min MH showed a CS-SMC value greater than 100% (Fig. 5) indicating a greater similarity between the methods than within the methods.

Discussion

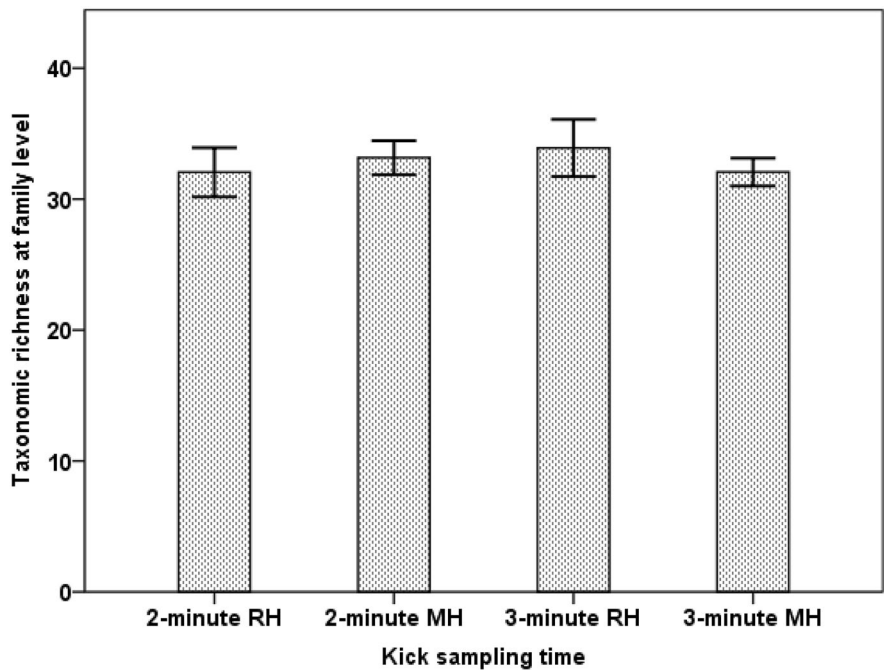
This study set out to develop a fixed time sampling protocol as a method for bioassessment of streams and rivers using macroinvertebrates in Ethiopia although it has not been tested yet for its utility in determining water quality conditions at sites. Comparisons among methods can be made at several levels of data organization: taxonomic composition, relative abundances, metrics, indices, and bioassessment endpoints (e.g., good-fair-poor) (Blocksom et al., 2008). Four kick sampling methods (2- and 3-min RH, 2- and 3-min MH) were selected and compared using various statistical analysis methods. Based on the Kruskal Wallis test (H), the majority of metrics compared were similar probably because of some reasons. According to Blocksom et al. (2008), the predominant habitat tends to be riffles in higher gradient streams and a sample that is collected using the single habitat method that

Table 3 Descriptive summary, mean \pm SE, mean ranks (MR), and Kruskal–Wallis (K-W) test with its p -value of macroinvertebrate metrics for each kick sampling method. SE standard errors, H' Shannon's index, D Simpson's index, $ASPT_{BMWP}$ average score per taxon from BMWP, FBI family biotic index, *Eph*, Ephemeroptera

Metrics	Descriptive summary for each method										MR for each method and K-W test (H) and p -values									
	2-min RH ($n=20$)		2-min MH ($n=60$)		3-min RH ($n=20$)		3-min MH ($n=60$)		2-min RH		2-min MH		3-min RH		3-min MH		H ($df=3$)		p -value	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	H	p
Richness (family)	32.05 \pm 0.94		33.17 \pm 0.65		33.9 \pm 1.09		32.07 \pm 0.53		77.28	94.93	82.87	74.4	3.25	0.355						
Abundance	347.2 \pm 4.42		334.4 \pm 3.08		392.55 \pm 6.4		367.88 \pm 4.55		72.33	126.58	55.38	92.98	42.4	0.001*						
EPT richness	6.8 \pm 0.2		7.55 \pm 0.18		7.05 \pm 0.21		7.42 \pm 0.15		60.28	71.83	86.43	84.2	6.33	0.097						
EPT (%)	66.9 \pm 1.04		66.5 \pm 0.85		66.7 \pm 1.14		67.48 \pm 0.72		78.3	76.58	78.83	84.21	0.65	0.885						
Eph. richness	4.05 \pm 0.05		4.1 \pm 0.04		4.1 \pm 0.07		4.1 \pm 0.04		77	81	81	81	0.51	0.916						
Eph. abundance	181.75 \pm 5		175.4 \pm 3.05		205.7 \pm 6.33		192.4 \pm 4.13		75.13	111.1	63.36	89.23	19.34	0.001*						
BMWP	150.05 \pm 4.1		158.2 \pm 2.86		161.95 \pm 4.9		153.95 \pm 2.58		70.4	96.6	84.12	74.88	4.62	0.202						
ASPT _{BMWP}	4.69 \pm 0.05		4.79 \pm 0.04		4.79 \pm 0.04		4.81 \pm 0.03		65.33	82.28	79.26	86.21	3.13	0.372						
Evenness	0.77 \pm 0.00		0.78 \pm 0.01		0.77 \pm 0.00		0.77 \pm 0.00		77.13	78.28	82.89	79.98	0.32	0.956						
Shannon's index	2.66 \pm 0.03		2.71 \pm 0.03		2.71 \pm 0.03		2.67 \pm 0.02		74.83	89.73	82.97	76.85	1.64	0.651						
Simpson's index	0.90 \pm 0.00		0.90 \pm 0.00		0.91 \pm 0.00		0.90 \pm 0.00		77.58	91.03	78.78	79.69	1.21	0.75						
ETHbios	129.3 \pm 4.63		130 \pm 2.72		136.9 \pm 4.8		127.53 \pm 2.34		82.6	97.5	79.62	75.02	3.6	0.308						
ASPT _{ETHbios}	4.03 \pm 0.06		3.93 \pm 0.05		4.04 \pm 0.06		3.98 \pm 0.03		84.58	91.45	75.67	80.33	1.93	0.588						
FBI	4.76 \pm 0.04		4.78 \pm 0.03		4.72 \pm 0.03		4.70 \pm 0.02		84.95	73.25	90	71.93	5.25	0.155						
Predators (%)	17.35 \pm 0.65		19.22 \pm 0.53		18.66 \pm 0.66		19.32 \pm 0.58		57.95	76.85	83.13	86.61	6.1	0.107						
Scrapplers (%)	21.44 \pm 0.49		20.7 \pm 0.48		21.44 \pm 0.38		21.48 \pm 0.47		88.3	86.93	72.64	83.62	2.95	0.4						
Collectors (%)	47.17 \pm 0.68		46.23 \pm 0.72		46.31 \pm 0.68		44.94 \pm 0.59		94.33	84.65	82.43	72.58	3.8	0.284						
Filterer (%)	13.57 \pm 0.32		13.24 \pm 0.33		13.12 \pm 0.35		13.76 \pm 0.39		88.58	77.03	75.66	83.81	1.68	0.641						
Shredders (%)	0.48 \pm 0.1		0.64 \pm 0.08		0.48 \pm 0.11		0.50 \pm 0.06		77.3	71.93	87.8	77.13	2.64	0.45						

*Significant differences at $p < 0.05$, df degree of freedom

Fig. 2 Comparison of the mean (\pm standard error) of selected benthic macroinvertebrate metric (family-level taxonomic richness) for the four kick sampling methods using bar charts and standard error bars examined in the dry season, 2019



focuses on riffle habitats should be very similar to the sample that is collected using the multiple habitat method, in which habitats are sampled according to their proportional representation in the stream.

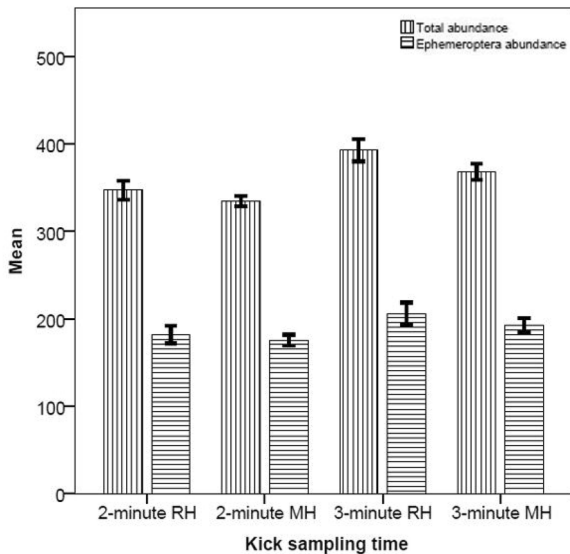


Fig. 3 Comparison of the mean (\pm standard error) of total abundance and Ephemeroptera abundance for the four kick sampling methods using bar charts and standard error bars examined in the dry season, 2019

Also, the lack of differences among the methods could be related to inadequate sampling and also the level of taxonomic resolution. The only exceptions included total abundance and Ephemeroptera abundance, which differed significantly between methods. The differences may probably be due to the larger total number of macroinvertebrates and Ephemeroptera families in the 3-min RH and 3-min MH samples.

Similarly, diversity indices, biotic scores, and functional feeding group attributes evaluated for the variability are in the support of future bioassessments using these 4 methods. This finding was in agreement with Friberg et al. (2006) who found that a sampling method with a lower sampling effort, such as the 2-min, achieves equal sampling effort as 3-min in terms of macroinvertebrate community structure. Furthermore, Mykra et al. (2006) indicated that similarities in community structure among different sample sets collected for 1-, 2-, 3-, and 4-min showing that even 1-min samples would yield adequate and comparable taxonomic structures. Similarly, Feeley et al. (2012) found that 20- and 60-s MH kick sampling methods displayed a reasonable representation of the taxa in terms of various metrics tested. Likewise, the ANOSIM test did not show any overall difference among the four methods and highlighted the high

Table 4 *p*-values in the pairwise post hoc test calculated for macroinvertebrate total abundance (lower left) and Ephemeroptera abundance (upper right) having significant differences in the Kruskal–Wallis test

		2-min RH	2-min MH	3-min RH	3-min MH		
Tot abun	2-min RH	–	0.241	0.006**	0.22	2-min RH	Eph abun
	2-min MH	0.051	–	<0.001**	0.003**	2-min MH	
	3-min RH	<0.001*	<0.001*	–	0.081	3-min RH	
	3-min MH	0.033*	<0.001*	0.002*	–	3-min MH	

*Significant differences in total abundance (Tot abun)

**Significant differences in Ephemeroptera abundance (Eph abun)

similarity in macroinvertebrate community structure across all sites sampled using the methods tested. In contrast, a related study (Feeley et al., 2012) found a high similarity but significantly different between 20- and 60-s kick sampling methods.

The RELATE analysis on relative abundances of macroinvertebrate communities showed that samples from similar habitats but collected by different kick sampling methods had higher correlation coefficients than those sampled from different habitats by either different or similar kick sampling methods. For example, 2- and 3-min RH pair matrices and 2- and 3-min MH pair matrices had Pearson correlation coefficients of 0.76 and 0.80, respectively. This probably highlights that more similar macroinvertebrate taxa had been collected from similar habitats regardless of the different methods used. This relationship had been well described by Khudhair et al. (2019) where sediment type, water flow, presence of

aquatic vegetation types directly affect the assemblage structure of benthic macroinvertebrates. The other reason for this high similarity among the methods may be related to the family level identification used. Although the family-level identification is adequate for bioassessment of water quality (Chessman, 1995; Corkum, 1989; Wright et al., 2000), Bailey et al. (2001) concluded that organisms identified to genus or species level provide a significantly more informative description of conditions in the stream than higher taxonomic levels.

The CS-SMC measure also showed high similarity scores (close to 100%) among the method pairs indicating that all kick sampling methods represented almost identical macroinvertebrate community structures from MH and RH samples. This finding was in contradiction to Cao et al. (1998) and Cao et al. (2002) who found that sample size is important but they identified some taxa to lower levels (species,

Fig. 4 MDS plot of the macroinvertebrate community structure of the samples from the various sampling methods based on $\log(x+1)$ -transformed Bray–Curtis similarity matrix

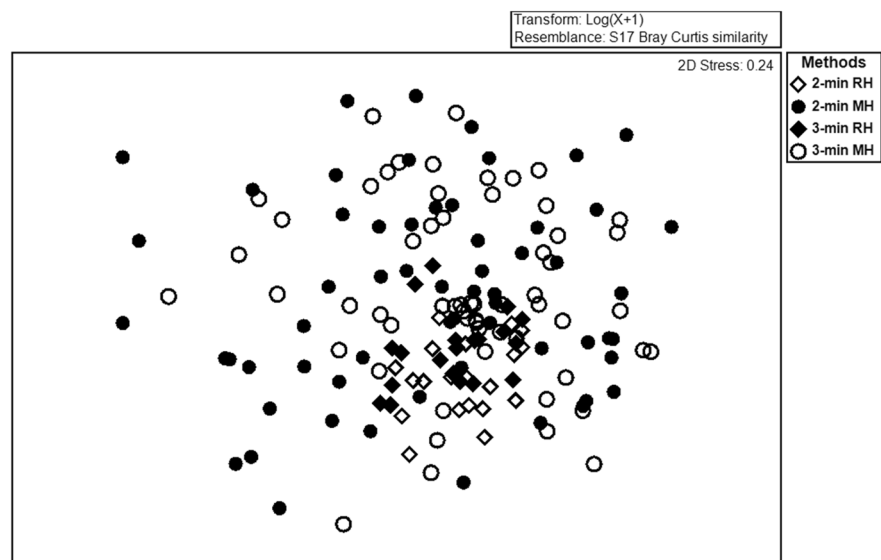


Table 5 MVDISP (index of multivariate dispersion) and SIMPER test results, respectively, representing the relative dispersion and similarity of samples within each method (factor)

Method	MVDISP	SIMPER (%)
3-min RH	0.34	82.43
2-min RH	0.341	82.21
3-min MH	0.945	77.51
2-min MH	1.197	75.42

genus, and family levels). They noted that increasing sample size can affect the relative abundance of macroinvertebrates in a sample and thus the analysis of community structure. However, Mykra et al. (2006), who sampled only riffle habitats, showed that sample size is more important to detect rare and endangered macroinvertebrate taxa and concluded that the 2-min samples are sufficient for most biodiversity purposes.

In the current study, several analytical tests, for example, the Kruskal–Wallis, ANOSIM, RELATE, and CS-SMC showed no significant differences among the four methods indicating that the low sampling effort, such as those of the 2-min RH and 2-min MH method, perform equally and often better, in terms of the numbers of taxa and abundances recorded compared to that of methods with a higher sampling effort (3-min RH and 3-min MH). However, although the stress value of the MDS ordination plot among the four methods was higher compared to the Clarke (1993) cutoff value, it appears that the sample replicates using the MH methods (2- and 3-min) were highly dispersed compared to the sample replicates using the RH methods (2- and 3-min). The differences were well reflected by the numerical analyses of MVDISP and SIMPER in terms of dispersion and similarity percentage of macroinvertebrate samples. The MVDISP analysis showed that the 2- and 3-min RH have an equal and low index of multivariate dispersion (IMD) and high within similarity (SIMPER). In contrast, high IMD and lower within similarity (SIMPER) were observed in multihabitat replicate

samples collected by 2- and 3-min (Table 5). According to Warwick and Clarke (1993), there are two potential sources of increased variability among replicate samples: (1) because of an increase in the variability of abundances of the same set of taxa, and (2) due to changes in species identities.

As detected in MDS and from the MVDISP analysis, it appears that the 2- and 3-min RH kick samples were much less variable and this would enhance bioassessments by being better able to detect impacted sites as opposed to the highly variable 2- and 3-min MH samples. Based on the findings in this study, 2- and 3-min RH kick sampling methods in terms of time and habitat type for the collection of macroinvertebrate data would be key considerations in the country. However, the extra time spent for kicking such as 3-min in RH was an effort spent increasing the abundances of macroinvertebrate taxa already sampled. This unnecessary increase in the abundance of macroinvertebrates collected by the extra time for each sample will increase the identification effort (Feeley et al., 2012) and costs which are directly related to the amount of material and the number of individuals collected (Barbour & Gerritsen, 1996). Although not examined in this study, the taxonomic resolution also plays a major role in the costs associated with sample processing (Vlek et al., 2006).

In summary, if the focus is not on rare taxa and the required information is not dependent on additional evidence provided by the use of lower taxonomic levels of identification, the results in this study support the use of the shorter 2-min RH kick samples, also used by other studies (Heino et al., 2003; Mykra et al., 2006; Paavola et al., 2003), for bioassessment of wadeable rivers and streams in Ethiopia. Future studies should explore the effectiveness of lower sampling time, such as 1-min kick samples across a wider range of wadeable stream and river types, and also the effect of variable levels of taxonomic resolution, not just for bioassessment of water quality but also for aquatic invertebrate biodiversity assessment, an issue that is becoming increasingly important

Table 6 The RELATE analysis results indicating Spearman’s rank correlation coefficients (ρ) between different biotic resemblance matrices

	2-min RH	3-min RH	2-min MH	3-min MH
2-min RH	–			
3-min RH	0.76 ($p=0.001$)	–		
2-min MH	0.50 ($p=0.001$)	0.61 ($p=0.001$)	–	
3-min MH	0.52 ($p=0.001$)	0.63 ($p=0.001$)	0.80 ($p=0.001$)	–

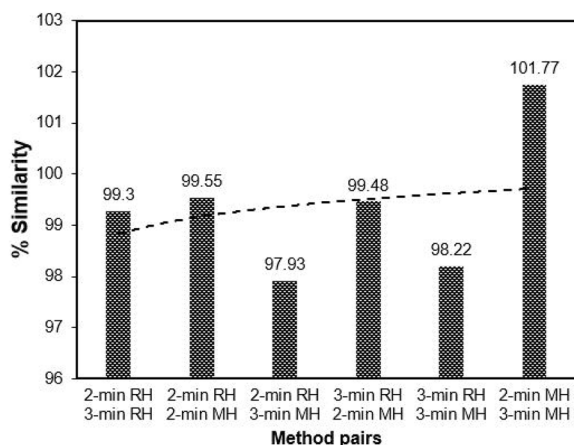


Fig. 5 The similarity between replicate samples within each method relative to the similarity of samples between methods as calculated by CS-SM

with alarming losses of freshwater biodiversity in many parts of the world (Reid et al., 2019). Furthermore, we would suggest future studies to include impacted sites to see how the methods perform there as well.

Conclusions

A method that takes the minimum duration of time and effort but produces an equitable representation of the taxa available would be eventually adopted in the assessment of streams and rivers (Buss & Borges, 2008; Feeley et al., 2012; Mykra et al., 2006) based on the ecological contexts of the region. The unnecessary increase in abundance, as found with 3-min RH kick samples, will considerably increase the identification effort for each sample. Sampling for a longer time duration increases effort and fatigue (Feeley et al., 2012; Mykra et al., 2006) and reduces concentration and the productivity of the operators, in turn potentially affecting the results (Feeley et al., 2012). Consequently, the optimum macroinvertebrate kick sampling method which enables rapid but accurate bioassessment of water quality in terms of time and habitat type is an important consideration in Ethiopia. Based on the results in the present study, the shorter 2-min RH kick sampling method could be a good candidate for the bioassessment of wadeable rivers and streams in Ethiopia but further testing across a pollution gradient is recommended.

Acknowledgements This study was partly funded by Addis Ababa University (AAU), Ethiopia. The logistic support including different dissection microscopes and expert training on data management and macroinvertebrate identification by the University College Dublin (UCD), Ireland, has also been acknowledged. The Ethiopian Institute of Water Resources (EIWR), Jimma University (JU), and Wollo University (WU), Ethiopia, also supported the transportation service and several other logistics during the data collection season. The authors also wish to thank the staff at UCD, AAU, JU, and WU for their helpful cooperation in the arrangement of laboratory facilities and offices.

Data availability The dataset generated and/or analyzed during the present study is available from the corresponding author.

Declarations

Conflict of interest The authors declare no competing interests.

References

- Alemneh, T., Ambelu, A., Bahrndorff, S., Mereta, S. T., Pertoldi, C., & Zaitchik, B. F. (2017). Modeling the impact of highland settlements on ecological disturbance of streams in Choke Mountain Catchment: Macroinvertebrate assemblages and water quality. *Ecological Indicators*, 73, 452–459. <https://doi.org/10.1016/j.ecolind.2016.10.019>
- Alemneh, T., Ambelu, A., Zaitchik, B. F., Bahrndorff, S., Mereta, S. T., & Pertoldi, C. (2019). A macroinvertebrate multi-metric index for Ethiopian highland streams. *Hydrobiologia*, 843(1), 125–141. <https://doi.org/10.1007/s10750-019-04042-x>
- Ambelu, A., Mekonen, S., & G, Silassie, A., Malu, A., & Karunamoorthi, K. (2013). Physicochemical and Biological Characteristics of Two Ethiopian Wetlands [journal article]. *Wetlands*, 33(4), 691–698. <https://doi.org/10.1007/s13157-013-0424-y>
- Ambelu, A., Mekonen, S., Koch, M., Addis, T., Boets, P., Everaert, G., & Goethals, P. (2014). The application of predictive modeling for determining bio-environmental factors affecting the distribution of blackflies (Diptera: Simuliidae) in the Gilgel Gibe watershed in southwest Ethiopia. *PLoS One*, 9(11), e112221. <https://doi.org/10.1371/journal.pone.0112221>
- Armitage, P., Gunn, R., Furse, M., Wright, J., & Moss, D. (1987). The use of prediction to assess macroinvertebrate response to river regulation. *Hydrobiologia*, 144(1), 25–32. <https://doi.org/10.1007/BF00008048>
- Aschalew, L., & Moog, O. (2015). Benthic macroinvertebrates based new biotic score “ETHbios” for assessing ecological conditions of highland streams and rivers in Ethiopia. *Limnologia*, 52, 11–19. <https://doi.org/10.1016/j.limno.2015.02.002>

Awulachew, S. B., Yilma, A. D., Loulseged, M., Loiskandl, W., Ayana, M., & Alamirew, T. (2007). *Water resources and irrigation development in Ethiopia*, 123. IWMI.

Bailey, R. C., Norris, R. H., & Reynoldson, T. B. (2001). Taxonomic resolution of benthic macroinvertebrate communities in bioassessments. *Journal of the North American Benthological Society*, 20(2), 280–286. <https://doi.org/10.2307/1468322>

Barbour, M. T., & Gerritsen, J. (1996). Subsampling of benthic samples: A defense of the fixed-count method. *Journal of the North American Benthological Society*, 15(3), 386–391. <https://doi.org/10.2307/1467285>

Barbour, M. T., Gerritsen, J., Snyder, B., & Stribling, J. (1999). Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish. 339. Washington, DC: US Environmental Protection Agency, Office of Water.

Barbour, M. T., Stribling, J. B., & Verdonshot, P. F. (2006). The multihabitat approach of USEPA’s rapid bioassessment protocols: benthic macroinvertebrates. *Limnetica*, 25(3), 839–850. <https://doi.org/10.23818/limn.25.58>

Berhanu, B., Seleshi, Y., & Melesse, A. M. (2014). Surface water and groundwater resources of Ethiopia: potentials and challenges of water resources development. In: Melesse A., Abteu W., Setegn S. (eds) Nile River basin. Springer, Cham (pp. 97–117). https://doi.org/10.1007/978-3-319-02720-3_6

Beyene, E. G., & Meissner, B. (2010). Spatio-temporal analyses of correlation between NOAA satellite RFE and weather stations’ rainfall record in Ethiopia. *International Journal of Applied Earth Observation and Geoinformation*, 12, S69–S75. <https://doi.org/10.1016/j.jag.2009.09.006>

Block, P. J. (2008). An assessment of investments in agricultural and transportation infrastructure, energy, and hydroclimatic forecasting to mitigate the effects of hydrologic variability in Ethiopia. Challenge Program on Water and Food, Colombo, Sri Lanka: CGIAR. <https://hdl.handle.net/10568/21543>

Blocksom, K. A., Autrey, B. C., Passmore, M., & 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(3), 577–593. <https://doi.org/10.1111/j.1752-1688.2008.00183.x>

Bradley, D. C., & Ormerod, S. J. (2002). Evaluating the precision of kick-sampling in upland streams for assessments of long-term change: The effects of sampling effort, habitat and rarity. *Archiv Für Hydrobiologie*, 199–221. <https://doi.org/10.1127/archiv-hydrobiol/155/2002/199>

Buss, D. F., & Borges, E. L. (2008). Application of rapid bioassessment protocols (RBP) for benthic macroinvertebrates in Brazil: Comparison between sampling techniques and mesh sizes. *Neotropical Entomology*, 37(3), 288–295. <https://doi.org/10.1590/s1519-566x2008000300007>

Buss, D. F., Carlisle, D. M., Chon, T.-S., Culp, J., Harding, J. S., Keizer-Vlek, H. E., Robinson, W. A., Strachan, S., Thirion, C., & Hughes, R. M. (2015). Stream bio-monitoring using macroinvertebrates around the globe: A comparison of large-scale programs. *Environmental Monitoring and Assessment*, 187(1), 4132. <https://doi.org/10.1007/s10661-014-4132-8>

Buttigieg, P. L., & Ramette, A. (2014). A guide to statistical analysis in microbial ecology: A community-focused, living review of multivariate data analyses. *FEMS Microbiology Ecology*, 90(3), 543–550. <https://doi.org/10.1111/1574-6941.12437>

Callanan, M., Baars, J. R., & Kelly-Quinn, M. (2008). Critical influence of seasonal sampling on the ecological quality assessment of small headwater streams. *Hydrobiologia*, 610(1), 245. <https://doi.org/10.1007/s10750-008-9439-4>

Camargo, J. A., Alonso, A., & Salamanca, A. (2005). Nitrate toxicity to aquatic animals: A review with new data for freshwater invertebrates. *Chemosphere*, 58(9), 1255–1267. <https://doi.org/10.1016/j.chemosphere.2004.10.044>

Cao, Y., & Hawkins, C. P. (2011). The comparability of bioassessments: A review of conceptual and methodological issues. *Journal of the North American Benthological Society*, 30(3), 680–701. <https://doi.org/10.1899/10-067.1>

Cao, Y., Hawkins, C. P., & Storey, A. W. (2005). A method for measuring the comparability of different sampling methods used in biological surveys: Implications for data integration and synthesis. *Freshwater Biology*, 50(6), 1105–1115. <https://doi.org/10.1111/j.1365-2427.2005.01377.x>

Cao, Y., Williams, D. D., & Larsen, D. P. (2002). Comparison of ecological communities: The problem of sample representativeness. *Ecological Monographs*, 72(1), 41–56. [https://doi.org/10.1890/0012-9615\(2002\)072\[0041:COECTP\]2.0.CO;2](https://doi.org/10.1890/0012-9615(2002)072[0041:COECTP]2.0.CO;2)

Cao, Y., Williams, D. D., & Williams, N. E. (1998). How important are rare species in aquatic community ecology and bioassessment? *Limnology and Oceanography*, 43(7), 1403–1409. <https://doi.org/10.4319/lo.1998.43.7.1403>

Carlisle, D. M., Hawkins, C. P., Meador, M. R., Potapova, M., & Falcone, J. (2008). Biological assessments of Appalachian streams based on predictive models for fish, macroinvertebrate, and diatom assemblages. *Journal of the North American Benthological Society*, 27(1), 16–37. <https://doi.org/10.1899/06-081.1>

Chapman, M., & Underwood, A. (1999). Ecological patterns in multivariate assemblages: Information and interpretation of negative values in ANOSIM tests. *Marine Ecology Progress Series*, 180, 257–265. <https://doi.org/10.3354/meps180257>

Cheshmedjiev, S., Soufi, R., Vidinova, Y., Tyufekchieva, V., Yaneva, I., Uzunov, Y., & Varadinova, E. (2011). Multi-habitat sampling method for benthic macroinvertebrate communities in different river types in Bulgaria. *Water Research and Management*, 1(3), 55–58. https://www.academia.edu/download/45659865/Multi-habitat_sampling_method_for_benthi20160516-29022-1wj5yt5.pdf

Chessman, B. C. (1995). Rapid assessment of rivers using macroinvertebrates: A procedure based on habitat-specific sampling, family level identification and a biotic index. *Australian Journal of Ecology*, 20(1), 122–129. <https://doi.org/10.1111/j.1442-9993.1995.tb00526.x>

Clarke, K., & Green, R. (1988). Statistical design and analysis for a ‘biological effects’ study. *Marine Ecology Progress Series*, 46(1/3), 213–226. Retrieved May 26, 2021 from <http://www.jstor.org/stable/24827586>

- Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, 18(1), 117–143. <https://doi.org/10.1111/j.1442-9993.1993.tb00438.x>
- Clarke, K. R., & Gorley, R. N. (2006). PRIMER v6: User Manual/Tutorial. PRIMER-E Ltd, Plymouth, UK.
- Clarke, R., Furse, M., Gunn, R., Winder, J., & Wright, J. (2002). Sampling variation in macroinvertebrate data and implications for river quality indices. *Freshwater Biology*, 47(9), 1735–1751. <https://doi.org/10.1046/j.1365-2427.2002.00885.x>
- Corkum, L. D. (1989). Patterns of benthic invertebrate assemblages in rivers of northwestern North America. *Freshwater Biology*, 21(2), 191–205. <https://doi.org/10.1111/j.1365-2427.1989.tb01358.x>
- Cumming, G., Fidler, F., & Vaux, D. L. (2007). Error bars in experimental biology. *The Journal of Cell Biology*, 177(1), 7–11. <https://doi.org/10.1083/jcb.200611141>
- De Bikuna, B. G., López, E., Leonardo, J., Arrate, J., Martínez, A., Agirre, A., & Manzano, A. (2015). Reduction of sampling effort assessing macroinvertebrate assemblages for biomonitoring of rivers. *Knowledge and Management of Aquatic Ecosystems*, (416), 8. <https://doi.org/10.1051/kmae/2015004>
- Fadiran, A., Dlamini, S., & Mavuso, A. (2008). A comparative study of the phosphate levels in some surface and ground-water bodies of Swaziland. *Bulletin of the Chemical Society of Ethiopia*, 22(2). <https://doi.org/10.4314/bcese.v22i2.61286>
- Feeley, H. B., Woods, M., Baars, J.-R., & Kelly-Quinn, M. (2012). Refining a kick sampling strategy for the bioassessment of benthic macroinvertebrates in headwater streams. *Hydrobiologia*, 683(1), 53–68. <https://doi.org/10.1007/s10750-011-0940-9>
- Friberg, N., Sandin, L., Furse, M., Larsen, S., Clarke, R., & Haase, P. (2006). Comparison of macroinvertebrate sampling methods in Europe. *Hydrobiologia*, 566, 365–378. https://doi.org/10.1007/978-1-4020-5493-8_26
- Gabriels, W., Lock, K., De Pauw, N., & Goethals, P. L. (2010). Multimetric Macroinvertebrate Index Flanders (MMIF) for biological assessment of rivers and lakes in Flanders (Belgium). *Limnologia-Ecology and Management of Inland Waters*, 40(3), 199–207. <https://doi.org/10.1016/j.limno.2009.10.001>
- Geist, J. (2011). Integrative freshwater ecology and biodiversity conservation. *Ecological Indicators*, 11(6), 1507–1516. <https://doi.org/10.1016/j.ecolind.2011.04.002>
- Getachew, M., Ambelu, A., Tiku, S., Legesse, W., Adugna, A., & Kloos, H. (2012). Ecological assessment of Cheffa Wetland in the Borkena Valley, northeast Ethiopia: Macroinvertebrate and bird communities. *Ecological Indicators*, 15(1), 63–71. <https://doi.org/10.1016/j.ecolind.2011.09.011>
- Getachew, M., Mulat, W. L., Mereta, S. T., Gebrie, G. S., & Kelly-Quinn, M. (2020). Challenges for water quality protection in the greater metropolitan area of Addis Ababa and the upper Awash basin, Ethiopia—Time to take stock. *Environmental Reviews*. <https://doi.org/10.1139/er-2020-0042>
- Gratwicke, B. (1998). The effect of season on a biotic water quality index: A case study of the Yellow Jacket and Mazowe Rivers, Zimbabwe. *Southern African Journal of Aquatic Science*, 24, 24–35. <https://doi.org/10.1080/10183469.1998.9631409>
- Haase, P., Lohse, S., Pauls, S., Schindehütte, K., Sundermann, A., Rolaufts, P., & Hering, D. (2004a). Assessing streams in Germany with benthic invertebrates: Development of a practical standardised protocol for macroinvertebrate sampling and sorting. *Limnologia*, 34(4), 349–365. [https://doi.org/10.1016/S0075-9511\(04\)80005-7](https://doi.org/10.1016/S0075-9511(04)80005-7)
- Haase, P., Pauls, S., Sundermann, A., & Zenker, A. (2004b). Testing different sorting techniques in macroinvertebrate samples from running waters. *Limnologia*, 34(4), 366–378. [https://doi.org/10.1016/S0075-9511\(04\)80006-9](https://doi.org/10.1016/S0075-9511(04)80006-9)
- Hawkes, H. A. (1998). Origin and development of the biological monitoring working party score system. *Water Research*, 32(3), 964–968. [https://doi.org/10.1016/S0043-1354\(97\)00275-3](https://doi.org/10.1016/S0043-1354(97)00275-3)
- Heino, J., Muotka, T., Mykrä, H., Paavola, R., Hämäläinen, H., & Koskenniemi, E. (2003). Defining macroinvertebrate assemblage types of headwater streams: Implications for bioassessment and conservation. *Ecological Applications*, 13(3), 842–852. [https://doi.org/10.1890/1051-0761\(2003\)013\[0842:DMATOH\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2003)013[0842:DMATOH]2.0.CO;2)
- Heino, J., Muotka, T., Paavola, R., Hämäläinen, H., & Koskenniemi, E. (2002). Correspondence between regional delineations and spatial patterns in macroinvertebrate assemblages of boreal headwater streams. *Journal of the North American Benthological Society*, 21(3), 397–413. <https://doi.org/10.2307/1468478>
- KDOW. (2015). Methods for sampling benthic macroinvertebrate communities in wadeable waters. Kentucky Department for Environmental Protection, Kentucky Division of Water (KDOW), Frankfort, Kentucky.
- Khudhair, N., Yan, C., Liu, M., & Yu, H. (2019) Effects of habitat types on macroinvertebrates assemblages structure: A case study of Sun Island Bund Wetland BioMed Research International. <https://doi.org/10.1155/2019/2650678>
- Koester, M., & Gergs, R. (2017). Laboratory Protocol for genetic gut content analyses of aquatic macroinvertebrates using group-specific rDNA primers. *Journal of Visualized Experiments*, (128). <https://doi.org/10.3791/56132>
- Lakew, A. (2017). Reference Criteria For Ecological River Health Assessment in Ethiopian Highlands. *Ethiopian Journal of Environmental Studies & Management*, 10(7), 942–957. <http://ejesm.org/wp-content/uploads/2017/10/ejesm.v10i7.10.pdf>
- Lakew, A., & Moog, O. (2015). Benthic macroinvertebrates based new biotic score “ETHbios” for assessing ecological conditions of highland streams and rivers in Ethiopia. *Limnologia-Ecology and Management of Inland Waters*, 52, 11–19. <https://doi.org/10.1016/j.limno.2015.02.002>
- Larsen, S., Mancini, L., Pace, G., Scalici, M., & Tancioni, L. (2012). Weak concordance between fish and macroinvertebrates in Mediterranean streams. *PLoS ONE*, 7(12), e51115. <https://doi.org/10.1371/journal.pone.0051115>
- Luo, K., Hu, X., He, Q., Wu, Z., Cheng, H., Hu, Z., & Mazumder, A. (2018). Impacts of rapid urbanization on the water quality and macroinvertebrate communities of streams: A case study in Liangjiang New Area, China. *Science of the Total Environment*, 621, 1601–1614. <https://doi.org/10.1016/j.scitotenv.2017.10.068>

- Mabidi, A., Bird, M. S., & Perissinotto, R. (2017). Distribution and diversity of aquatic macroinvertebrate assemblages in a semi-arid region earmarked for shale gas exploration (Eastern Cape Karoo, South Africa). *PLoS One*, *12*(6). <https://doi.org/10.1371/journal.pone.0178559>
- Mandaville, S. (2002). Benthic macroinvertebrates in freshwaters: Taxa tolerance values, metrics, and protocols. 128. Citeseer.
- Mereta, S. T., Boets, P., Bayih, A. A., Malu, A., Ephrem, Z., Sisay, A., Endale, H., Yitbarek, M., Jemal, A., & De Meester, L. (2012). Analysis of environmental factors determining the abundance and diversity of macroinvertebrate taxa in natural wetlands of Southwest Ethiopia. *Ecological Informatics*, *7*(1), 52–61. <https://doi.org/10.1016/j.ecoinf.2011.11.005>
- Mereta, S. T., Boets, P., De Meester, L., & Goethals, P. L. (2013). Development of a multimetric index based on benthic macroinvertebrates for the assessment of natural wetlands in Southwest Ethiopia. *Ecological Indicators*, *29*, 510–521. <https://doi.org/10.1016/j.ecolind.2013.01.026>
- MOA. (1998). *Agroecological zones of Ethiopia. Natural resources management and regulatory department (NRMRD) of Ministry of Agriculture (MOA), Addis Ababa, Ethiopia*. Available from <http://hdl.handle.net/123456789/2517> [accessed 1 June 2021].
- Montana, D. (2012). Sample collection, sorting, taxonomic identification, and analysis of benthic macroinvertebrate communities standard operating procedure. Helena: Montana Department of Environmental Quality. Available from <http://mtwatersheds.org/app/wp-content/uploads/2017/11/Taxonomic-Identification.pdf> [accessed 1 June 2021].
- Mykra, H., Ruokonen, T., & Muotka, T. (2006). The effect of sample duration on the efficiency of kick-sampling in two streams with contrasting substratum heterogeneity. *Internationale Vereinigung Für Theoretische Und Angewandte Limnologie: Verhandlungen*, *29*(3), 1351–1355. <https://doi.org/10.1080/03680770.2005.11902901>
- Oluyemi, E., Adekunle, A., Adenuga, A., & Makinde, W. (2010). Physico-chemical properties and heavy metal content of water sources in Ife North Local Government Area of Osun State, Nigeria. *African Journal of Environmental Science and Technology*, *4*(10), 691–697. Available from <http://www.academicjournals.org/AJEST> [accessed 27 May 2020].
- Ostermiller, J. D., & Hawkins, C. P. (2004). Effects of sampling error on bioassessments of stream ecosystems: Application to RIVPACS-type models. *Journal of the North American Benthological Society*, *23*(2), 363–382. [https://doi.org/10.1899/0887-3593\(2004\)023%3c0363:EOSEOB%3e2.0.CO;2](https://doi.org/10.1899/0887-3593(2004)023%3c0363:EOSEOB%3e2.0.CO;2)
- Paavola, R., Muotka, T., Virtanen, R., Heino, J., & Kreivi, P. (2003). Are biological classifications of headwater streams concordant across multiple taxonomic groups? *Freshwater Biology*, *48*(10), 1912–1923. <https://doi.org/10.1046/j.1365-2427.2003.01131.x>
- Pinna, M., Marini, G., Mancinelli, G., & Basset, A. (2014). Influence of sampling effort on ecological descriptors and indicators in perturbed and unperturbed conditions: A study case using benthic macroinvertebrates in Mediterranean transitional waters. *Ecological Indicators*, *37*, 27–39. <https://doi.org/10.1016/j.ecolind.2013.09.038>
- Reid, A. J., Carlson, A. K., Creed, I. F., Eliason, E. J., Gell, P. A., Johnson, P. T., Kidd, K. A., MacCormack, T. J., Olden, J. D., & Ormerod, S. J. (2019). Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological Reviews*, *94*(3), 849–873. <https://doi.org/10.1111/brv.12480>
- Resh, V. H., & Rosenberg, D. M. (1993). Freshwater bio-monitoring and benthic macroinvertebrates. (504.4 FRE). <https://doi.org/10.2307/2404174>
- Sanchez-Montoya, M. D. M., Arce, M. I., Vidal-Abarca, M. R., Suárez, M. L., Prat, N., & Gómez, R. (2012). Establishing physico-chemical reference conditions in Mediterranean streams according to the European Water Framework Directive. *Water Research*, *46*(7), 2257–2269. <https://doi.org/10.1016/j.watres.2012.01.042>
- Stark, J., Boothroyd, I., Harding, J., Macted, J., & Scarsbrook, M. (2001). Protocols for sampling macroinvertebrates in wadeable streams. New Zealand Macroinvertebrate Working Group Report No. 1. Prepared for the Ministry for the Environment. Sustainable Management Fund Project No. 5103. p. 57. <https://riversgroup.org.nz/wp-content/uploads/2018/06/4.1.3-macroinvertebrate-sampling.pdf>
- Sickle Van, J. (1997) Using mean similarity dendrograms to evaluate classifications. *Journal of Agricultural, Biological, and Environmental Statistics*, 370–388. <https://doi.org/10.2307/1400509>
- Vlek, H. E., Sporka, F., Krno, I., & j. (2006). Influence of macroinvertebrate sample size on bioassessment of streams. *Hydrobiologia*, *566*, 523–542. https://doi.org/10.1007/978-1-4020-5493-8_35
- Wang, X., Cai, Q., Tang, T., Yang, S., & Li, F. (2012). Spatial distribution of benthic macroinvertebrates in the Erhai basin of southwestern China. *Journal of Freshwater Ecology*, *27*(1), 89–96. <https://doi.org/10.1080/02705060.2011.617041>
- Warwick, R. M., & Clarke, K. R. (1993). Increased variability as a symptom of stress in marine communities. *Journal of Experimental Marine Biology and Ecology*, *172*(1), 215–226. [https://doi.org/10.1016/0022-0981\(93\)90098-9](https://doi.org/10.1016/0022-0981(93)90098-9)
- Wasserman, R. J., Matcher, G. F., Vink, T. J. F., & Froneman, P. W. (2015). Preliminary evidence for the organisation of a bacterial community by zooplanktivores at the top of an estuarine planktonic food web. *Microbial Ecology*, *69*(2), 225–256. <https://doi.org/10.1007/s00248-014-0505-3>
- Wright, J. F., Sutcliffe, D. W., & Furse, M. T. (2000). Assessing the biological quality of freshwaters. RIVPACS and other techniques. Freshwater Biological Association, Ambleside, England. <http://aquaticcommons.org/id/eprint/5341>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.