Stream biomonitoring using macroinvertebrates around the globe: a comparison of large-scale programs

Citation	Buss, D. F., Carlisle, D. M., Chon, T. S., Culp, J., Harding, J. S., Keizer-Vlek, H. E., ... & Hughes, R. M. (2014). Stream biomonitoring using macroinvertebrates around the globe: a comparison of large-scale programs. Environmental Monitoring and Assessment, 187(1), 4132. doi:10.1007/s10661-014-4132-8
DOI	10.1007/s10661-014-4132-8
Publisher	Springer
Version	Version of Record
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsofuse

Environ Monit Assess (2015) 187:4132
DOI 10.1007/s10661-014-4132-8

Stream biomonitoring using macroinvertebrates around the globe: a comparison of large-scale programs

Daniel F. Buss · Daren M. Carlisle · Tae-Soo Chon · Joseph Culp · Jon S. Harding · Hanneke E. Keizer-Vlek · Wayne A. Robinson · Stephanie Strachan · Christa Thirion · Robert M. Hughes

Received: 10 April 2014 / Accepted: 28 October 2014 / Published online: 9 December 2014

Abstract Water quality agencies and scientists are increasingly adopting standardized sampling methodologies because of the challenges associated with interpreting data derived from dissimilar protocols. Here, we compare 13 protocols for monitoring streams from different regions and countries around the globe. Despite the spatially diverse range of countries assessed, many aspects of bioassessment structure and protocols were similar, thereby providing evidence of key characteristics that might be incorporated in a global sampling methodology. Similarities were found regarding sampler type, mesh size, sampling period, subsampling methods, and taxonomic resolution. Consistent field and laboratory methods are essential for merging data sets collected by multiple institutions to enable large-scale comparisons. We discuss the similarities and differences among protocols and present current trends and future recommendations for monitoring

D. F. Buss (✉)
Laboratório de Avaliação e Promoção da Saúde Ambiental, Instituto Oswaldo Cruz, FIOCRUZ,
Rio de Janeiro, RJ, Brazil
e-mail: dbuss@ioc.fiocruz.br

D. M. Carlisle
U.S. Geological Survey,
Reston, VA, USA
e-mail: dcarlisle@usgs.gov

T.-S. Chon
Department of Biological Sciences,
Pusan National University,
Busan, Pusan 609-735, Republic of Korea
e-mail: tschon@pusan.ac.kr

J. Culp
Environment Canada and Canadian Rivers Institute,
Department of Biology, University of New Brunswick,
Fredericton, NB, Canada
e-mail: joseph.culp@ec.gc.ca

J. S. Harding
School of Biological Sciences, University of Canterbury,
Private Bag 4800, Christchurch, New Zealand
e-mail: jon.harding@canterbury.ac.nz

H. E. Keizer-Vlek
Department of Freshwater Ecology, Alterra, Wageningen University and Research Centre,
Wageningen, The Netherlands
e-mail: hanneke.keizer-vlek@wur.nl

W. A. Robinson
School of Environmental Sciences, Charles Sturt University,
Thurgoona, Australia
e-mail: wrobinson@csu.edu.au

S. Strachan
Environment Canada,
#201-401 Burrard Street, Vancouver, BC, Canada
e-mail: stephanie.strachan@ec.gc.ca

C. Thirion
Resource Quality Information Services, Department Water and Sanitation,
Private Bag X313, Pretoria 0001, South Africa
e-mail: thirionc@dwa.gov.za

R. M. Hughes
Amnis Opes Institute and Department of Fisheries and Wildlife, Oregon State University,
Corvallis, OR, USA
e-mail: hughes.bob@amnisopes.com

programs, especially for regions where large-scale protocols do not yet exist. We summarize the current state in one of these regions, Latin America, and comment on the possible development path for these techniques in this region. We conclude that several aspects of stream biomonitoring need additional performance evaluation (accuracy, precision, discriminatory power, relative costs), particularly when comparing targeted habitat (only the commonest habitat type) versus site-wide sampling (multiple habitat types), appropriate levels of sampling and processing effort, and standardized indicators to resolve dissimilarities among biomonitoring methods. Global issues such as climate change are creating an environment where there is an increasing need to have universally consistent data collection, processing and storage to enable large-scale trend analysis. Biomonitoring programs following standardized methods could aid international data sharing and interpretation.

Keywords Biomonitoring protocols · Standardization · Biological assessment · Subsampling taxonomic resolution · River management

Introduction

Biological monitoring protocols began being used systematically for determining the ecological condition of water bodies since the development of the saprobic system, more than 100 years ago (Rosenberg and Resh 1993). In general, biomonitoring is used for characterizing ecological condition of waterways and the existence, extent, and severity of biological degradation, as well as for long-term trend analysis. Biomonitoring aids the identification of regional biotic attributes and pattern and potential sources and causes of degradation. Additionally, it can evaluate the effectiveness of pollution control and remediation activities and can be used to detect and assess cumulative impacts (Barbour et al. 1999; Hering et al. 2006; Paulsen et al. 2008). Although, many biological assemblages (e.g., bacteria, algae, fish) are used for assessing the ecological condition of rivers and streams, benthic macroinvertebrates are the most common bioindicator (Hellawell 1986; Rosenberg and Resh 1993).

Ideally, macroinvertebrate biomonitoring protocols should be efficient, cost-effective, easy to use, and sensitive to impacts (Resh and Jackson 1993; Resh et al. 1995). Moreover, they need to be consistently applied at large spatial scales (Barbour et al. 1999; Hering et al. 2006; Paulsen et al. 2008; Stoddard et al. 2008), as this will improve the ability to undertake meaningful regional to national bioassessment comparisons. This is why consistent field and laboratory methods are essential for making spatially extensive comparisons and for merging data sets collected by multiple institutions (Hughes and Peck 2008; Bonar et al. 2009). However, internationally accepted standard methods do not yet exist for collecting and processing benthic macroinvertebrate samples. Although many national agencies and institutions have developed standardized biomonitoring protocols, they often are not consistently applied at local, national, or continental scales—even in their home countries. The reasons why this has not occurred are various, not only often associated with lack of logistics and funding, particularly when upscaling from smaller programs, but also with a reluctance to change established techniques or gear, the existence of large historical databases acquired via specific methods, or the fact that locally developed methods sometimes yield more accurate results than regionally applicable methods. In geographic regions where large-scale protocols do not yet exist, such as Latin America and South-east Asia, the application of internationally accepted sampling methods may be especially useful for rapidly creating credible bioassessment programs. Also, biomonitoring programs following standard biomonitoring methods could aid international data sharing and interpretation.

Consistent field and laboratory methods are essential for making spatially extensive comparisons and for merging data sets collected by multiple institutions (Hughes and Peck 2008; Bonar et al. 2009). Thus, the objective of this paper is to compare and contrast key aspects of spatially extensive biomonitoring protocols in the USA, Canada, Europe (England and Wales, Germany, Flanders (Belgium), The Netherlands, Slovakia, and Spain), South Africa, Republic of Korea (South Korea), Australia, and New Zealand and to discuss the barriers and research gaps that need to be overcome to develop "globally" transferable standard methods. The protocols analyzed in this paper were selected for the assessment based on the availability of researchers familiar with the methods used in those countries. For Europe, a set of questions was sent to researchers in countries representing different backgrounds and history in monitoring macroinvertebrates. All European countries that responded to the questions were included in this study.

Biomonitoring program objectives and brief history

The history of many biomonitoring programs reviewed in this paper is similar. Biomonitoring protocols were developed first at small spatial scales—in a province, state, or small- to medium-sized river basins leading to many protocols being developed independently, often as products of tradition or convenience (Carter and Resh 2001). In the case of large countries, where many agencies developed their own methods (e.g., the USA), or multinational basins within Europe, it became necessary to intercalibrate data from the different protocols or to standardize a protocol to produce scientifically valid information for basin management (Blocksom et al. 2008; Clarke and Hering 2006). Otherwise, in situations where rivers border or pass through multiple states/provinces/countries, one is not able to determine the degree to which assessments from differing agencies result from different sampling and analytical protocols, different ecological conditions, or both.

Here, we analyze large-scale biomonitoring programs implemented in the USA, Canada, European Union (EU), South Africa, Australia, New Zealand, and South Korea. These programs represent a range of legislative or legal mandates as well as a variety of governmental funding formulae (Table 1). Those programs have generally been designed to monitor and assess biological status, patterns, and trends in lotic ecosystems with the overarching goal of providing information for freshwater policy and management.

In the USA, two major national biomonitoring programs exist and are funded through the US Environmental Protection Agency (USEPA) and the US Geological Survey (USGS). The objective of the USEPA's National Aquatic Resources Survey (NARS; previously called EMAP) is to assess the ecological status and trends in all US surface waters (lakes, streams, rivers, wetlands, coastal). The NARS was developed as a means of fulfilling the USEPA's requirements to report on the status and trends of US waters under the Clean Water Act of 1972. The assessment is based on a probability sample of all possible freshwaters because a census of all waters would be fiscally and logistically infeasible, and previous attempts to merge disparate non-standardized state data produced imprecise and inaccurate reports (Hughes et al. 2000). In contrast, the USGS' National Water Quality Assessment (NAWQA) assesses the effects of major land use types (e.g., agriculture, urbanization) on streams and ground water through use of sites selected along an anthropogenic disturbance gradient. Both NAWQA and NRSA (the National Rivers and Streams Assessment of NARS) assess physical habitat, water chemistry, algal, macroinvertebrate, and fish assemblages.

In Canada, the Canadian Aquatic Biomonitoring Network (CABIN) was developed by Environment Canada to promote interagency collaboration and data sharing to achieve comparable and consistent reporting on freshwater ecosystem health. The program evolved from research conducted by Environment Canada in the Great Lakes (Reynoldson et al. 1995) and in the Fraser River Basin, British Columbia (Reynoldson et al. 1997). As a result, routine biological monitoring was applied regionally in these areas and the CABIN national biomonitoring strategy was instituted in 1999 (Reynoldson et al. 1999). This strategy relies on the Reference Condition Approach for assessment (Bailey et al. 2004), which required the establishment and continued maintenance of a large reference database. In a country the size of Canada, interagency collaboration is the most efficient and cost-effective means to acquire reference data whether the agency is federal, provincial, municipal, First Nation, community watershed group, university, or industry. This is true, even if the agencies have differing goals, legal obligations, funding sources, and extent of interagency collaboration. At the national level, Environment Canada maintains a common CABIN Website, database, and training program to support the standardized collection, assessment, reporting, and distribution of biological monitoring information by all agencies using nationally comparable standards.

In Europe, since the introduction of the European Water Framework Directive (WFD) in 2000, a legal structure exists for a common approach to the management and protection of freshwater ecosystems. The objective of the WFD is to monitor and assess the ecological status of surface waters and to maintain or reach good ecological status of EU surface waters by 2015 (European Commission 2000). The European Union has funded multiple research projects in an attempt to unify biological assessment and monitoring efforts. All countries (Austria, Czech Republic, Greece, Italy, the Netherlands, Portugal, Germany, Sweden) participating in the AQEM project (The Development and Testing of an Integrated Assessment System for the Ecological Quality of Streams and Rivers throughout Europe using Benthic Macroinvertebrates; Hering et al. 2006), applied a standardized sampling and sample processing

Table 1 Protocols/programs for monitoring streams from different regions and countries around the globe

	Protocol/program analyzed	Protocol/program since	Sampling device	Mesh size (μm)	Sampled habitats	Sampling effort	Sampling extent	Subsampling
USA/USEPA	NARS	1992–	D-frame	500	Multiple: sampled in proportion to coverage	Samples—eleven 0.09 m^2 samples (0.99 m^2)	40× mean wetted width	Fixed count—500 individuals +large/rare taxa
USA/USGS	NAWQA	1991–	Slack sampler/ D-frame	500	Riffles or snags+qualitative multiple samples in proportion to coverage	Samples—five 0.25 m^2 samples in riffles/snags (1.25 m^2); Time—1 h qualitative samples	Fixed 150–300 m	Fixed count—300 individuals (or 8 h)+large/rare taxa
Canada	CABIN	2006–	Kick-net	400	Riffle/run	Time—3 min kicking	6× bankfull width (streams), fixed 3 min in one margin (large rivers)	Fixed count—300 individuals (if not achieved in 50 % sample, pick whole sample)
England and Wales	National monitoring program	Information not available	Kick-net/dredge	1000	Multiple: sampled in proportion to coverage	Time—3 min +1 min manual search	Fixed 50 m	Entire sample
Germany	PERLODES protocol	2004–	Handnet	500	Multiple (those >5 % coverage) sampled in proportion to coverage	Samples—twenty 25 cm samples (1.25 m^2)	Fixed—20–50 or 50–100 m (depending on stream size)	Fixed count/fixed area—350 individuals (in at least 1/6 of the sample)
Belgium (Flanders)	National monitoring program	Information not available	Handnet	500	Multiple: sampled in proportion to coverage	Time—3–5 min	Fixed—10–20 m	Entire sample
Netherlands	Handboek Hydrobiolgie (protocol)	2006–	Handnet	500	Multiple: sampled in proportion to coverage	Length—5–10 m	Fixed—50–100 m	Minimum of 1925 individuals, group specific; subsampling is optional
Slovakia	National monitoring program	2006-	Kick-net	500	Multiple (those >5 % coverage sampled in proportion to coverage)	Samples—twenty 25 cm samples (1.25 m^2)	Fixed—100 m	Fixed count/fixed area—500 individuals (in at least 1/6 of the sample)
Spain	National monitoring program	2006-	D-frame	500	Multiple (those >5 % coverage sampled in proportion to coverage)	Samples—twenty 50 cm samples (2.50 m^2)	Fixed—100 m	Whole sample (>5 mm fraction)/ Fixed count (optional)
South Africa	National River Health monitoring program (SASS and MIRAI)	SASS 1996-; MIRAI 2005-	Kick-net	1000	Multiple: biotopes: stones, vegetation, gravel sand, and mud	Time—2–5 min; length—2 m (marginal vegetation); area—1 m^2 (aquatic vegetation)	Fixed—30–100 m	Fixed time - 15 min per biotope group
South Korea	NAEMP	2008–	Surber	1000	Riffles (if depth ≤50 cm); otherwise, riffles and runs	Samples—3 Surber samples	Fixed—50–100 m	Entire sample
Australia	NRHP, SRA	1997–2002, 2004–2013	Kick-net (sweep net in larger rivers)	250	Edge (riffles where they occur)	Length—10 m (riffle, main channel or edge)	10× stream width	Fixed count—200 individuals; fixed area—10 % sample in South Australia
New Zealand	Sampling macroinvertebrates in wadeable streams	2001–	Kick-net/Surber	500	Riffle/run (hard-bottom); Edge/ macrophytes (soft–bottom)	Samples—3–10 samples depending on stream typology	Fixed—single riffle	Fixed count—200 individuals; full count coded abundance

Table 1 (continued)

	Subsampling device or sorting procedures	Sorting aids	Taxonomic level	Biological index used	Sampling season	QA/QC		Agency responsible for monitoring
						Crew training and certification for field procedures	Crew training and taxonomic certification	
USA/USEPA	Sample splitter	Stereomicroscope	Genus: mostly (lowest possible)	Multimetric, predictive O/E	Low flow; Jun-Sep, depending on region	Field training; public field protocols	Taxonomic certification of sample processing labs	State EPAs, contractors
USA/USGS	Gridded tray	Stereomicroscope	Genus/species (including Chironomidae)	Multimetric, predictive O/E	Low flow; Apr-May and/or Sep-Oct, depending on region	Field training; public field protocols	Taxon-based QC; taxonomic certification	USGS
Canada	Marchant box	Stereomicroscope	Family	Multivariate, predictive (multimetric if limited data)	Low flow; Jun-Nov	Online training and field certification	Taxonomic certification; 10 % samples sent to CABIN lab	Varied
England and Wales	Coded abundance	Stereomicroscope	Family (but moving to Species from 2004-)	Biotic, predictive	Mar-May and Sep-Nov	Training and certification; internal audits	Training and certification; external audits	National database; users upload data with quality check
Germany	Live pick, gridded tray	Naked eye	Species (except Diptera and Oligochaeta)	Multimetric	Feb-Apr	Not obligatory, but training provided	No	Regional
Belgium (Flanders)	None	Naked eye	Genus/family	Multimetric	Apr-Nov	Training and certification	Training and certification	Central (Flanders)
Netherlands	Subsampling optional	Naked eye	Species mostly (lowest possible)	Biotic	Mar-Jun	Not obligatory, but training provided	Not obligatory, but training provided	Regional
Slovakia	Live pick, gridded tray	Naked eye	Species: mostly (lowest possible)	Multimetric	Apr-Nov	Training provided	Training and certification	National
Spain	Volumetric subsample (100 individuals minimum + large/ rare taxa)	Naked eye	Family (some in order / class)	Multimetric	Jun-Sep	Training provided	Training provided	National and regional
South Africa	None	Hand lens and naked eye	Family (some in order / class)	Biotic	Low flow; May-Sep or Oct-Apr, depending on region	National auditing and accreditation valid for 3 years	National auditing and accreditation valid for 3 years	Provincial gov, consultants; National Dept. Water Affairs
South Korea	None	Naked eye	Species (genus/family for Chironomidae)	Biotic (Won et al. 2006)	Mar-Apr and Sep-Oct	Field training; sampling by same team	Random inspection of samples; nat'l codes and keys	Min. Environment Nat'l Inst. Env. Research
Australia	Live pick (field) or Marchant box (lab)	Stereomicroscope (lab), magnifiers (field live pick)	Family (some in order/ class)	Predictive, filters O/E	Low flow; Apr-May and Sep-Oct	Field training and certification	Random inspection of samples; nat'l codes and keys	State EPAs
New Zealand	None	Stereomicroscope	Family (genus for EPT)	Biotic (qualitative, semi-quantitative and quantitative)	Base-flow; Sep-Feb, depending on region	Full-time staff	10 % samples checked (<10 % inaccurate taxonomy and count needed)	Regional gov't, consultants; no central gov't department

Table 1 (continued)

	No. of sites assessed each year nationally	Public participation in monitoring	Reporting facilities	Specific legislation	Estimated cost for 1 sample	Field protocols	Best journal references to date
USA/USEPA	1000 sites; rotating panel every 5 years with 10 % revisit	Limited to a few states, municipalities and NGOs	National database publicly available; report via annual reports	Clean Water Act (1972); report to Congress on nation's waters every 2 years	US$5000 (physical habitat, algae, fish, fish tissue, water quality, crew, travel)	Peck et al. (2006) and Hughes and Peck (2008)	Hughes and Peck (2008), Paulsen et al. (2008), and Stoddard et al. (2008)
USA/USGS	50–100 sites; variable number in regional studies	None	Data storage; data open to public; internet application to input data	Not applicable	US$5000–10,000 (physical habitat, algae, fish, water quality, crew, travel, data process)	Moulton et al. (2002)	http://water.usgs.gov/nawqa/ecology/
Canada	>800 sites; 10 % of reference sites revisited	Limited to a few NGOs where multiagency collaboration exists	Central database management; requires training to access data	Canada Water Act; Canadian Envir. Protect. Act; Canadian Envir. Assess. Act	CAN$ 450–1300 depending on access	Environment Canada site[a]	Reynoldson et al. (1995, 1997, 2001)
England and Wales	~10,000 sites sampled every 3 years; rotating panel	Yes, but data do not contribute to official monitoring	Report to the European Commission	European Legislation–Water Framework Directive; national	£225 (sampling, sorting and identification)	Environment Agency (2012a, b)	Information not available
Germany	Information not available	None	Report to the European Commission	European Legislation–Water Framework Directive	Information not available	Haase and Sundermann (2004) and Meier et al. (2006)	Kail et al. (2012)
Belgium (Flanders)	400 sites	None	Report to the European Commission	European Legislation–Water Framework Directive	€300 (sampling, sorting and identification)	Gabriels et al. (2010)	Gabriels et al. (2010)
Netherlands	Variable across 26 water authorities	None	Report to the European Commission	European legislation; Water Framework Directive	€1500 (sampling, sorting and identification)	Splunder et al. (2006) and Bijkerk (2010)	Information not available
Slovakia	200 sites; rotating panel	None	Report to the European Commission	European Legislation–Water Framework Directive	€450 (sampling, sorting and identification)	AQEM Consortium (2002) and Makovinská et al (2008)	Mišíková et al. 2010
Spain	Variable; no set number of sites	None	Report to the European Commission	European Legislation–Water Framework Directive	€400 (sampling, sorting and identification)	Pardo et al. (2010)	Bennett et al. (2011)
South Africa	Variable; no set number of sites	Schools and public, but data do not contribute to official monitoring	National Rivers database publicly available	National Water Act (Act 36 of RSA, 1998)	Variable, depends on the sampler used	Dickens and Graham (2002) and Thirion (2008)	Dallas (1997), Ollis et al. (2006), and Dallas (2004)
South Korea	Same 960 sites	None	Web/GIS database; available online	Ecological Conservation Act	US$400–500	Cho et al. (2011) and Jun et al. (2011)	Cho et al. (2011), Bae et al. (2011), and Jun et al. (2011; 2012)
Australia	NRHP: >1000 sites; SRA: 400 sites rotating panel with 25 % revisit	Waterwatch, but data do not contribute to official monitoring	Indicators and summary of results published; available online	National Water Initiative (2007), legislated in Victoria and ACT	US$1000–3000	Davies (1994); but states have own manuals[b]	Davies 2000), Simpson and Norris (2000) and Davies (et al. 2010)
New Zealand	Variable across 17 Regional Councils	Stream & Water Care Groups, but data do not contribute to official monitoring	Primarily internal reports	Resource Management Act (1991); Regional Gov't water plans include biotic indices	US$300–400	Stark et al. (2001), but Councils use own modifications	Clapcott et al. (2012)

[a] Available from www.ec.gc.ca/rcba-cabin

[b] Available from http://ausrivas.ewater.com.au/index.php/manuals-a-datasheets

methodology. In the STAR project (Standardization of River Classifications: Framework method for calibrating different biological survey results against ecological quality classifications to be developed for the Water Framework Directive), several additional countries (UK, France, Poland Slovakia, Denmark, Latvia, Italy) applied a modified version of the AQEM methodology (i.e., the AQEM-STAR methodology; Clarke et al. 2006). Despite these efforts, a "pan-European protocol" for sampling, sample processing, and data analysis of macroinvertebrate assemblages does not yet exist. Nevertheless, there are two important European standards (EN) related to the monitoring of benthic macroinvertebrates: (1) EN-ISO 10870:2012: Water quality—guidelines for the selection of sampling methods and devices for benthic macroinvertebrates in freshwaters and (2) EN 16150:2012: Water quality—guidance on pro-rata Multi-Habitat sampling of benthic macroinvertebrates from wadeable rivers.

In South Africa, the National Aquatic Ecosystem Health Monitoring Program (NAEHMP) is managed by the Department of Water and Sanitation and aims to assess the status and trends of the inland water bodies in terms of ecosystem status. To date, only the program for assessing the ecosystem status of rivers (The National River Health Programme (RHP)) has been fully designed and implemented. The RHP focuses not only on aquatic macroinvertebrates but also includes other biota such as fish and riparian vegetation, as well as measures of habitat quality and quantity. The Resource Quality Information Services Directorate of the Department of Water and Sanitation is responsible for the RHP but relies on partnerships with local, provincial and regional government departments, Water Boards, NGOs, and academic institutions to assist with the sampling. There is currently a process to rationalize the >600 national sites spread throughout South Africa.

In Australia, there are two major programs: (1) the National River Health Program (NRHP), which included a one-time national environmental assessment of all catchments conducted between 1997 and 2002, and (2) the Sustainable Rivers Audit (SRA), which was an ongoing (2004–2013), large-scale (1 million ha) program monitoring river health bi-annually in 23 catchments across five states. The Australian River Assessment System (AUSRIVAS) sampling protocols (Simpson and Norris 2000) were developed as part of the NRHP (Schiller 2003) and allow for site-based assessments, which are summarized at regional scales. The SRA selected sites using a probabilistic sampling design specifically targeted at catchment scale reporting that allows site or regional scale reference conditions to be used for reporting (Davies et al. 2010).

In New Zealand, no national monitoring program exists, although the National Institute of Water and Atmospheric Research (NIWA) have a program called the National River Water Quality Network (NRWQN). This network consists of 77 sites (primarily single sites on larger rivers), which are annually sampled for water quality and macroinvertebrates. NIWA has developed its own protocols for this program. Most of New Zealand biomonitoring of the state of streams, rivers, and lakes occurs at the regional (provincial) level, and approximately 17 Regional Councils undertake annual stream health monitoring. This monitoring is required under the Resource Management Act of 1991. Generally, sites are sampled once per year and each Council selects sites based on differing criteria and may use slightly different sampling protocols. The total number of sites sampled annually varies depending on each Council's priorities and budgets, but the total number of sites assessed nationally is around 1500–2000. The New Zealand Ministry for the Environment, which is primarily a policy agency, attempts to pool Regional Council data to extrapolate national patterns and trends and several national online databases have been developed although inconsistencies between different council's protocols remain a barrier.

In South Korea, there are two national-scale macroinvertebrate biomonitoring programs. The National Ecosystem Survey is conducted to monitor species diversity in both terrestrial and aquatic ecosystems. Since 1986, in every 5-year cycle, 3500 sites are surveyed one to two times per year. For the period of 1997–2002, 1875 sites were reported regarding freshwater ecosystems (Park et al. 2004). Aiming specifically to assess aquatic ecosystems, the Nationwide Aquatic Ecological Monitoring Program (NAEMP) was launched in 2008. It is being conducted in five main river basins, where 640–960 sites are selected for surveying diatoms, macroinvertebrates, and fish two times per year. Multitaxa community organization and water quality according to the Korean Saprobic Index (Bae et al. 2011; Cho et al. 2011; Jun et al. 2012) are reported based on survey results. In this review, we analyze the NAEMP, which focus specifically on assessing the ecological health of South Korean streams.

Sampling considerations

Macroinvertebrates can be found in virtually all aquatic habitats—from tree-borne epiphytic bromeliads to lake hypolimnia, from intermittent spring creeks to the brackish estuaries of great rivers—and these very diverse environments may necessitate different sampling strategies. Most biological monitoring protocols require that sampling methods provide biologically meaningful information for management, yet be relatively rapid and low cost (Hughes and Peck 2008). Additionally, sampling efficiency and effort strongly influence species richness estimates (Li et al. 2001; Vlek et al. 2006; Cao and Hawkins 2011; Qu et al. 2013), biotic index scores (Angermeier and Karr 1986; Simon and Sanders 1999; Reynolds et al. 2003; Vlek et al. 2006; Hughes and Herlihy 2007; Cao and Hawkins 2011), and multivariate analysis results (Cao et al. 2002; Cao and Hawkins 2011; Qu et al. 2013). These factors need to be considered when choosing sampling protocols for a large-scale macroinvertebrate biomonitoring program. Several other issues must be considered to ensure a robust program, including deciding on the sampling device, mesh size, sampling effort (number of replicates and sampling area), sampled habitats and site length/extent, sampling period, subsampling and sorting procedures, level of taxonomic identification, and assessment indicator (e.g., organisms, metrics, indices). A robust protocol should also include quality control and quality assurance of field sampling, processing, identification, and data input and storage.

Sampling device

Despite the many sampling methods available, of the 13 biomonitoring protocols we compared, 12 programs opted to use kick samplers, whereas South Korea used the Surber sampler (Table 1). In New Zealand, standardized protocols exist for both kick-net and Surber sampler (Stark et al. 2001). In the USA, kick-net devices (including D-frame and hand nets) were used by >60 % of the State/Federal biomonitoring protocols, whereas fixed-area samplers such as dredges, Surber and Hess samplers were used by ~9 % and artificial substrates by ~13 % (Resh and Jackson 1993; Carter and Resh 2001; Carter and Resh 2013).

Borisko et al. (2007) suggested that the sampling method was not as important a source of variation in index values relative to other factors such as the stream types or annual variation. Brua et al. (2011) found that kick- and U-net (similar to the Surber sampler) produced similar results and concluded that benthic macroinvertebrate data collected by these methods could be combined for data analysis and bioassessments, given that mesh size of the sample nets is similar. Other studies found that kick samplers collected more taxa and enabled more accurate index values to be calculated than Surber samplers (Mackey et al. 1984; Buss and Borges 2008). The primary advantage of kick-nets is their usefulness in sampling a variety of habitats including deep- and non-flowing water and coarse and heterogeneous substrates (Hughes and Peck 2008).

Mesh size

By definition, macroinvertebrates are those visible to the naked eye, and thus, a net mesh size of approximately 0.5 mm (500 μm) might be used to sample those groups. Nevertheless, there has been much debate regarding the most appropriate mesh size, with the general consensus being that mesh size choice depends upon the objectives and constraints of the biomonitoring program (Bowman and Bailey 1997). Choice of mesh size is a critical decision point for stream biomonitoring programs because it determines the smallest size of the organisms collected and can change biotic metrics if abundance data are included because finer mesh will capture more organisms. Mesh size also influences the amount of backwash at the net opening, and the amount of fine detritus that must be processed.

Most biomonitoring protocols we evaluated used 500 μm or larger mesh sizes (exceptions were the Canadian and the Australian protocol which use 400 and 250 μm mesh, respectively; Table 1). In Canada, the 400 μm mesh was chosen because field tests indicated the number of individuals collected and laboratory processing time were significantly higher with samples collected with 200 μm mesh nets. Furthermore, the number of taxa found in the smaller mesh was not significantly different from the larger mesh (Rosenberg et al. 1999). A smaller mesh size is often applied under site conditions where there are few taxa and more microinvertebrates, as is the case of Australia. Similar to our results, Carter and Resh (2001) found that >80 % of biomonitoring protocols in the USA used 500–600 μm mesh. A mesh size of approximately 500 μm appears to be most cost-effective because it retains the most macroinvertebrate genera per unit of effort

(Rosenberg et al. 1999; Buss and Borges 2008), despite losing smaller specimens and earlier instars that are often difficult to identify accurately. Finally, most taxonomic keys are based on late instar nymphs or larvae, and thus smaller or early instars can create serious taxonomic issues for sample processors, as these early life stages are difficult to identify beyond family level (Winterbourn et al. 2006).

Sampling effort

The sampling effort (number of replicates, sampled area, and site extent) varied considerably among the programs we reviewed and reflect differences among programs designs and objectives. For example, some countries using the Reference Condition Approach (Bailey et al. 2004) incorporate the stream reach habitat as the sampling unit. In this case, different streams within a particular stream order or ecodistrict/ecozone are used for replication. In contrast, the South Korean and New Zealand protocols consider samples within a stream reach as replicates. These aspects need more research. Some insights for that are provided by the USEPA, USGS, and CABIN/Canada protocols that repeat samples to assess the effects of sampling, processing, month and year, and differing field crews on indicator variance (Rosenberg et al. 1999; Stoddard et al. 2008; Zuellig et al. 2012; Table 1).

Sampled area The protocols we analyzed adopted various levels of sampling effort, and they could be divided into three groups: (1) fixed sample number (USGS, USEPA, and South Korea), (2) fixed sampling length/area (Australia, New Zealand, Germany, The Netherlands, Slovakia, and Spain), and (3) fixed sampling time (England and Wales, Flanders, Canada, and South Africa). Although all three approaches intend to standardize sampling effort, there was a considerable difference of effort among them (Table 1).

Sampled habitats and site extent Sampling protocols varied from highly prescriptive to more flexible, depending on the program. Protocols from Canada, Australia, and South Korea focus entirely on sampling riffle/run habitats, while the Australian programs also include both edge and riffles, if they occur. New Zealand has different protocols for hard- or soft-bottom streams, although hard-bottom stream riffle/runs generally are sampled (Stark et al. 2001; Stark and Maxted 2010). In contrast, the USGS, USEPA, and the six European countries require sampling of all habitat types, and protocols from the USGS, USEPA, England and Wales, Germany, Spain, and Slovakia require habitats to be sampled based on their relative occurrence in the reach. Other studies also show disagreements between protocols regarding the habitats sampled. For example, the majority (67 %) of the state environmental agencies in the USA focus macroinvertebrate sampling only on the most diverse habitats (e.g., riffles; Carter and Resh 2001). In contrast, Barbour et al. (1999) and Hering et al. (2006) recommended 20 samples distributed proportionately by major habitat types. Gerth and Herlihy (2006) reported markedly fewer species and biased results from targeted habitats when riffles were rare or unrepresentative of an entire site. Other studies found that sampling only in riffles may produce lower taxonomic richness because most taxa seem to be strongly associated with a specific substrate type (Parsons and Norris 1996; Buss et al. 2004; Gerth and Herlihy 2006). In addition, Blocksom et al. (2008) concluded that in areas with a wide variety of stream types, the multiple habitat method may be more desirable than sampling riffles only.

We found considerable differences in the sampling reach length and how a site was defined. Most protocols used a fixed length (USGS, South Africa, England and Wales, Germany, Flanders, The Netherlands, Slovakia, and Spain), whereas others used a multiple of stream wetted width (USEPA, Australia, and New Zealand) or bankfull width (Canada). Sampling site varied from a single riffle (New Zealand) to 40 times the mean wetted channel width, with a minimum of 150 m (USEPA) (Table 1). The reach length in the latter protocol was determined from US field studies and designed to maximize physical habitat variability and fish and macroinvertebrate taxa richness within a reasonable level of effort (Li et al. 2001; Cao et al. 2002; Hughes and Peck 2008).

Sampling season

All programs and protocols we analyzed sampled predominantly during low-flow or dry season periods, when flows were most stable and conditions were safest for crews (Plafkin et al. 1989; Hering et al. 2006; Hughes and Peck 2008). Even though this more stable period varies among latitudes and continents, most protocols reported sampling sometime between April and November (Table 1). Many studies reported that macroinvertebrate biomonitoring data were sensitive to sampling season (Reece et al. 2001; Hawkins 2006; Šporka et al. 2006; Chen et al. 2014). This can create a problem because monitoring should be performed only in the season during which the protocol was developed. Alternatively, Cao and Hawkins (2011) suggest three options to circumvent this limitation: (1) standardize sampling on a short window of time (as indicated by Hose et al. 2004), (2) aggregate samples across seasons (e.g., Furse et al. 1984; Humphrey et al. 2000), or (3) adjust for the effect of seasonal variation on assemblage composition by modeling (e.g., Hawkins 2006).

Subsampling and sorting procedures

Most biomonitoring programs acknowledge a trade-off between efficiency and sensitivity for large-scale monitoring, thus many strategies for reducing processing time and cost have been implemented. Frequently, programs use subsampling procedures to reduce the amount of sample processed, speeding up the reporting of results and decreasing the costs associated with sample sorting. Common subsampling strategies use fixed area (e.g., Walsh 1997) and fixed sampling time (e.g., Environment Agency 2012a, b) and sort a fixed number of individuals, even though the sufficient number of individuals is a subject of ongoing debate (e.g., Barbour et al. 1996; Somers et al. 1998; Norris et al. 1995; Stark et al. 2001; King and Richardson 2002). In addition, some programs use rarefaction techniques, although some studies reported that such procedure may lead to misleading estimates of the true differences in taxa richness among sites (e.g., Cao et al. 2002; Ligeiro et al. 2013b). Others favor sorting the samples entirely (e.g., Courtemanch 1996; Doberstein et al. 2000), which may consume a great amount of time and increase costs—but presumably increases the capability to detect anthropogenic disturbance.

Most protocols we reviewed offer options for analyzing data using a fixed-count method. However, the number of individuals sampled in each protocol varied: 200 (Australia, New Zealand including scanning the whole sample for rare taxa), 300 (USGS, Canada), and 500 (USEPA). New Zealand offers several protocols (depending on the monitoring aim), including an option to undertake full counts. In such cases, however, the use of the coded abundance method may reduce the time and costs of sample processing. In Europe, the AQEM-STAR methodology, which has been adopted in some national protocols (Table 1) developed a standardized method for subsampling such that one sixth of the sample must be sorted with a minimum of 700 individuals. This was especially true for countries lacking a long bioassessment history or a national bioassessment program (e.g., Germany, Slovakia, and Spain), as opposed to those with decades of data and vested interests in a traditional approach. Carter and Resh (2001) stated that in the USA, most State protocols required sorting entire samples, but among those that used subsampling, 53 % sampled only 100 individuals.

All protocols we analyzed sample animals after preservation. Some studies have shown live sorting may be faster than lab sorting, but small and cryptic animals are often missed and it adds in-field complexity, producing higher variability in assessment results (Haase et al. 2004) and is more dependent on field team skills (Carter and Resh 2001).

Given the constraints of many biomonitoring programs (i.e., budgetary pressures, technical training, need for rapid response) and the relative efficiency of these methods (Barbour and Gerritsen 1996; Somers et al. 1998), subsampling has been recognized as having an acceptable cost/benefit ratio. Clarke et al. (2006) and Ligeiro et al. (2013a) recommended using a fixed count because of the effect of the number of individuals on richness metrics and also suggested tracking the number of subsamples so taxonomic densities can be estimated from quantitative samples. To implement large-scale stream and river surveys by employing subsampling, a number of programs include a robust quality control and use a consistent count (Cao and Hawkins 2011). Cao et al. (2002) showed that 500 individuals maximized the discriminatory power of similarity indices, while Ligeiro et al. (2013a) reported that samples with <300 individuals processed had less precision than those with 300 individuals. However, the accuracy and precision of the processing may also depend on the stream type, the metric being used (e.g., species richness versus a functional metric or a multimetric index; Clarke et al. 2006; Petkovska and Urbanič 2010; Marzin et al. 2012), and the level of taxonomic identification (Whittier and Van Sickle 2010).

While sorting macroinvertebrates from debris, most protocols use stereomicroscopes at low magnification (maximum ×10), but South Korea and EU countries sorted samples by eye. Sorting by eye may be more practical and faster, but it will result in missing smaller taxa and abundance numbers being lower. Metrics based on relative percentages, and those based on groups with many small organisms or during seasons dominated by early in stars are strongly influenced by the sorting strategy. Additionally, the sorting strategy needs to consider, if possible, to confidently identify small specimens and earlier instars at the chosen taxonomic level, because most benthic invertebrate keys are designed for late instar organisms. Specimens that are sorted and not identified are omitted in data analyses, thereby increasing processing time and cost, without improving data quality.

Taxonomic sufficiency

Taxonomic sufficiency, defined as the necessary taxonomic resolution to satisfy the objectives of a study (Ellis 1985), is a critical component of biomonitoring and is determined by considering the trade-offs associated with different levels of resolution. Although the species level holds benefits (e.g., Resh and McElravy 1993; Lenat and Resh 2001), others question if this level of detail is needed for a biomonitoring program. This is because biomonitoring datasets are usually summarized in indices, which do not necessarily require species data and are often robust to taxonomic aggregation (Vlek et al. 2006; Whittier and Van Sickle 2010).

Across the range of programs we evaluated, biomonitoring protocols could be divided between genus/species or lowest possible taxonomic level (USGS, USEPA, Germany, The Netherlands, Slovakia, and South Korea), a mixture of genus/family level, depending on known taxonomy (New Zealand and Flanders), and family-level assessment (Australia, Canada, South Africa, England and Wales, and Spain).

Resh and McElravy (1993), examining 45 published lotic biomonitoring studies, reported that many insect groups (e.g., Ephemeroptera, Plecoptera, Trichoptera, Coleoptera), as well as Platyhelminthes and Crustacea, were commonly identified to genus or species. Lesser known–or otherwise more challenging to identify taxa (e.g., Nematoda, Annelida, and Hydrachnidia)—were most often assigned to family or higher taxonomic groups. Similar results were found by Carter and Resh (2001) when analyzing biomonitoring protocols in the USA: crustaceans, mites, oligochaetes, and mollusks were generally identified more coarsely than the Ephemeroptera, Plecoptera, Trichoptera, and Chironomidae.

Some argue that the default taxonomic level should be species (Jones 2008). However, some countries and regions lack basic taxonomic information at lower levels (e.g., genus, species) or resources to train staff and undertake rigorous quality control of identification. Nonetheless, in some countries higher taxonomic levels (e.g., family) may provide similar bioassessment information as lower levels (e.g., genus, species; Furse et al. 1984; Marchant et al. 1995; Bowman and Bailey 1997; Wright et al. 2000; Reynoldson et al. 2001; Schmidt-Kloiber and Nijboer 2004; Buss and Vitorino 2010; Whittier and Van Sickle 2010), being less expensive to conduct (Vlek et al. 2006).

Approaches to biological assessment

In the 1980s, most biotic indices were based on subjective scoring systems (based on the presumed sensitivity, or resistance, of each taxon to impairment) such as the Biotic Index (Chutter 1972), Biotic Condition Index (Winget and Mangum 1979), Biological Monitoring Working Party system (BMWP; Armitage et al. 1983), Indice Biotico Esteso (IBE; Ghetti 1997), Family Biotic Index (Hilsenhoff 1977, 1987), Macroinvertebrate Community Index (Stark 1985, 1998), and many more (see reviews in Metcalfe 1989 and Rosenberg and Resh 1993). Since then, more objective, quantitative and precise indices have been developed, such as the SingScore in Singapore (Blakely et al. 2014). From the protocols we analyzed, England and Wales, Netherlands, South Africa, South Korea, and New Zealand are currently using biotic indices in their monitoring protocols.

Another approach that has been routinely used in monitoring programs is based on multivariate analysis, which compares the macroinvertebrate fauna observed at a site with a prediction of the fauna expected at that site in the absence of major environmental stress (Clarke et al. 2003). This method, also known as "predictive modeling," has been used for nearly three decades. This approach makes no a priori assumptions about the expected similarity of communities at different sites based on physical or chemical descriptors. The expected similarity is modeled based on a large collection of reference conditions. From the protocols we analyzed, England and Wales, Australia, Canada, and the USA have developed approaches based on predictive models of taxon occurrence (RIVPACS—Clarke et al. 2003; Wright et al. 2000; 1984; AUSRIVAS—Simpson and Norris 2000; benthic assessment of sediment (BEAST)—Reynoldson et al. 1995; Reference Condition–Bailey et al. 1998; O/E–Hawkins et al. 2000; Paulsen et al. 2008). One regional predictive model has been tested in New Zealand (Joy and Death 2003), but it is not currently being used in routine programs.

A third approach that has been widely used is based on multimetric indices. Based on Karr's (1981) conceptual model, a multimetric index is a combination of individual metrics that, together, represent a range of assemblage responses to human impact. Often, such indices incorporate responses of biotic indices and/or other biomonitoring approaches like taxonomic richness, assemblage composition and functional feeding groups. Multimetric approaches for benthic macroinvertebrates are the most widely used approach for water-quality assessments (Bonada et al. 2006). Large-scale multimetric indices have been developed for macroinvertebrates in many countries and continents (Klemm et al. 2003; Hering et al. 2006; Baptista et al. 2007; Stoddard et al. 2008; Moya et al. 2011; Cho et al. 2011; Jun et al. 2012). The USEPA-NRSA uses both multimetric and O/E indices for assessing all wadeable streams in the conterminous USA (Paulsen et al. 2008; Stoddard et al. 2008; USEPA 2013). Germany, Belgium, Slovakia, and Spain are currently employing multimetric indices in their monitoring programs.

Several other approaches have been tested, but here we analyzed only the current national (or large-scale) programs. For example, Pont et al. (2006, 2009), Moya et al. (2011), and USEPA (2013) developed predictive multimetric indices at a national-scale, calibrating reference sites for natural variables. That may be a new approach in biomonitoring programs, given that extensive sampling is conducted in reference sites. Other quantitative multivariate techniques employ artificial neural networks (i.e., self-organizing maps) by extracting complexity residing in community and metric data (Park et al. 2004; Bae et al. 2011; Cho et al. 2011; Li et al. 2012; Chon et al. 2013). Also, the use of multiple biological traits (Statzner et al. 2001; Menezes et al. 2010; Marzin et al. 2012) has been tested and this approach has advantages for large-scale applicability because aquatic invertebrates worldwide can be described and compared on the same scale for a given trait (Statzner et al. 1997). However, like other biomonitoring approaches, its application is hindered by the lack of knowledge about these traits in many regions of the world.

Bonada et al. (2006) analyzed ten biomonitoring approaches using 12 criteria that might provide an "ideal" biomonitoring protocol. For each protocol they addressed: (1) rationale (derived from sound theoretical concepts in ecology; a priori predictive; potential to assess ecological processes; potential to discriminate overall human impact; potential to discriminate different types of human impact); (2) implementation (costs for field sampling and sorting or standardized laboratory experimentation; simplicity of sampling protocol; cost for non-specialist taxonomic identification); and (3) performance (applicability across ecoregions or biogeographic provinces; reliability of indication of changes in overall human impact; reliability of indication of changes in different types of human impact; human impact indication on a linear scale). They found that no approach met all criteria, but multimetric indices, bioassays, multiple biological traits, and leaf-litter decay rates scored higher (10 out of 12 criteria).

All three major types of biomonitoring indices currently used in large-scale programs described in this paper (biotic, predictive, or multimetric indices) are based on the establishment of reference conditions at minimally or least-disturbed sites (sensu Hughes 1995; Stoddard et al. 2006) and data comparisons from test or impaired sites.

Defining reference conditions

Knowledge of benchmark or reference conditions is essential for developing and testing metrics and indices and for making rigorous biological assessments. Those conditions are based on data from sets of minimally or least-disturbed or "best practice" regional reference sites (Hughes 1995; Bailey et al. 1998; Stoddard et al. 2006; Whittier et al. 2007; Herlihy et al. 2008) and have been adopted in legislation in several countries (e.g., the Clean Water Act in the USA and the Water Reform Framework in Australia). As an example, the European Union Water Framework Directive (WFD Directive, 2000/60/EC) defines reference condition as sites with "no or minimal anthropogenic stress" and satisfying the following criteria: (1) reflecting totally, or nearly, undisturbed conditions for hydromorphological elements, general physicochemical elements, and biological-quality elements; (2) having concentrations of specific synthetic pollutants close to zero or below the limit of detection of the most advanced analytical techniques in general use; and (3) exhibiting concentrations of specific non-synthetic pollutants within the range normally associated with background levels.

The use of regional reference sites has advantages for large-scale monitoring programs to represent a range of values (for any given index or metric) resulting from sampling error and natural variability, both in time and in space (Stoddard et al. 2006). In situations without minimally disturbed sites, empirical models derived from associations between biological indicators and human-disturbance gradients can be extrapolated to infer conditions in the absence of human disturbance (e.g., Hughes 1995; Karr and Chu 1999).

Many factors can influence biological assemblages and should be considered when establishing reference condition. These may be both large-scale patterns like ecoregions (a priori regional patterns based on land-surface form, soil, potential natural vegetation, and land use; sensu Omernik 1987; Omernik and Griffith 2014) and smaller-scale characteristics, such as watershed area and stream order (Barbour et al. 1999), stream typology (Verdonschot and Nijboer 2004), and altitude (e.g., Bailey et al. 2004). Also, any set of sites—even undisturbed ones—vary over time, given the potential for influence of large-scale factors such as climate change, atmospheric contaminants, and land use (Nichols et al. 2010; Wang et al. 2011). Therefore, definitions of ecological status need to be viewed more as probability density functions than as discrete contiguous entities (Jones et al. 2010). In line with that view, approaches such as those described in Pont et al. (2006, 2009) and Chen et al. (2014) aim to adjust metrics and indices to account for natural variability and use residuals distributions to select metrics that discriminate between reference and disturbed sites.

All national biomonitoring protocols we analyzed use a priori criteria for reference site selection. A priori approaches use biological data from sets of least-disturbed reference sites in ecoregions (in the USA) or aggregate regions for setting index expectations; then sites are screened through use of abiotic and catchment criteria. In the case of the BEAST approach in Canada, reference sites are determined a priori and grouped according to similar assemblages (Reynoldson et al. 1997). Then, reference condition models are developed to relate habitat attributes to the biological assemblage; these models are used to determine with which reference group a test site will be compared. A posteriori systems use biological data to define biological expectations, which incorporates the problem of biological and logical circularity. Nonetheless, a priori reference sites can be influenced by unknown stressors such as migration barriers, alien species, anomalous physical and/or chemical habitat conditions, or the legacy effects of past impacts (Harding et al. 1998; Hughes 1995; Whittier et al. 2007; Zhang et al. 2009).

Developing large-scale biomonitoring programs elsewhere—Latin America

National-scale biomonitoring programs are less advanced in Latin America, as well as much of Africa, Asia (Morse et al. 2007), and Eastern Europe. Here, we focus on Latin America as a case study.

In Latin America, interest in developing and testing rapid biomonitoring tools has increased in the last decade. Several authors have described the effects on macroinvertebrate fauna of environmental variables or anthropogenic activities (e.g., Marques and Barbosa 2001; Buss et al. 2002; Fenoglio et al. 2002; Couceiro et al. 2007; Miserendino et al. 2008), but few studies have tested methods and developed indices, which are central for developing a systematic and effective biomonitoring program. Assessment of biomonitoring protocols has been conducted in Brazil, Mexico, Argentina, Bolivia, and Ecuador, with most studies applying slightly modified versions of biotic indices generated in Europe, such as the BMWP and/or Average Score Per Taxon (ASPT or BMWP/taxon richness) index to detect impairment (e.g., Jacobsen 1998; Tarras-Wahlberg et al. 2001). Other studies have adapted and tested biotic indices (e.g., Capítulo et al. 2001; Mugnai et al. 2008; Junqueira et al. 2010), multimetric indices (Weigel et al. 2002; Baptista et al. 2007; Ferreira et al. 2011; Oliveira et al. 2011b), multivariate models (Moreno et al. 2009), and predictive multimetric models (Moya et al. 2011) for basin or regional use.

Although some studies in the region demonstrate the benefits of using species-level taxonomy (Buss and Salles 2007), many indices based on family-level taxonomy provided similar discriminatory power as genus-level resolution for biomonitoring purposes (Buss and Vitorino 2010). Continentally, taxonomic knowledge of immature insects is scarce because many species are still undescribed. Taxonomy is based on adults and the correlation of adults with the immature forms is hindered by the lack of rearing studies. Therefore, the use of family-level taxonomy facilitates the integration of information within and among Latin American countries because this approach aids data and methods comparisons. However, it is still necessary to test family-level in a biomonitoring program for streams in the whole region, considering there is substantial difference of biota among the biomes. Finally, few studies in Latin America have dealt with developing and testing of other important aspects of biomonitoring protocols, such as sampling procedures and mesh sizes (Buss and Borges 2008), sample size (Schneck and Melo 2010), subsampling methods (Oliveira et al. 2011a; Ligeiro et al. 2013a), and taxonomic sufficiency (e.g., Melo 2005; Buss and Vitorino 2010).

In Brazil, national reports on water quality reveal that while more than 3000 sites/rivers are monitored each year, little biological information is gathered. Human resources for assessing biological condition are focused on the southeastern region of the country, which has 45 % of the population, 10 % of the territory, and only 6 % of surface waters. Most macroinvertebrate studies (67 %) are focused on taxonomy, auto-ecology or surveys, and it is estimated that more species are still unknown than described so far. To establish a national biomonitoring program, a multi-stakeholder panel was formed to discuss biomonitoring methods; mainstreaming biomonitoring in high-school, undergraduate, and graduate programs; building strategies to include public participation in sampling; data analysis and raising awareness; and creating legislative and funding mechanisms. Among the participating institutions were four ministries, the Brazilian Society of Limnology (ABLimno), the National Agency of Water (ANA), state environmental agencies, and academic institutions from 12 states. This ongoing process has encouraged the creation of a database with information on macroinvertebrates sampled in more than 2500 streams and rivers, the development of technical courses on biomonitoring in areas where this information is less seldom applied, and new regional taxonomic keys (e.g., Mugnai et al. 2010; Hamada et al. 2014).

Final remarks

Although field sampling and sample processing methods differed somewhat among the macroinvertebrate protocols we compared, there are many underlying similarities. Most countries collect composite samples from habitats present at a site during either base-flow or low-flow using a kick-net, with a 500 μm or larger mesh. Although some countries sort the entire sample, many employ fixed-count subsampling (varying from 200 to 500 individuals), use a stereomicroscope for magnification, and identify organisms to genus and/or family level. Most protocols used a priori reference sites. Bioassessment protocols were fairly evenly divided between biotic indices, multimetric indices, predictive models, or a combination of those to assess site condition.

Several factors dictate the limitations to adopting standardized biomonitoring protocols. First, because of climatic differences among countries it is unlikely that the same season, month, or flow condition will be appropriate for sampling in all regions; however, sampling during base-flow or low-flow appears to be a common sampling strategy. Secondly, when there is insufficient taxonomy or taxonomic expertise to identify specimens to species or genus, family-level identifications is the only option for bioassessment. We strongly support continued taxonomic studies to further develop taxonomic knowledge but believe that countries lacking this knowledge should not be discouraged from conducting biomonitoring in their water basin assessment plans. Above, we noted that family-level assessment can provide scientifically valid data for management purposes. A third factor that hinders standardization of bioassessment protocols is the fact that most biological indicators were developed for specific geographic regions, states, or countries. For example, several European countries developed multimetric indices to assess the ecological status of their national waters only. Some countries adopted the AQEM-STAR methodology for sampling and sample processing, but most developed or retained their own methods, with no truly unified method for European countries. Examples of international integration exist, including the European Fish Index (EFI, EFI+) project, which was successful in building a unified method because a standard protocol existed for electrofishing (CEN), sampling occurred during the summer low-flow period, the entire sample was processed to species, and a single index was developed collaboratively by the international research team funded by the EU (Pont et al. 2006). A similar approach was taken in the climatically and hydromorphologically diverse USA by the USEPA: standard field and laboratory methods were developed by a collaborative multiinstitutional research team (Hughes and Peck 2008); all samples were processed by accredited taxonomy laboratories; and a research team developed national and regional biological indices (Paulsen et al. 2008; Stoddard et al. 2008; USEPA 2013; Esselman et al. 2013). Similar to the EFI project, the collaborators were united in the goal of developing national methods and indicators, and the USEPA provided the funding for the research, monitoring, data management, index development, and reporting. A similar field approach, with minor modifications, is being tested in basin-scale pilot studies in China (Li et al. 2014) and Brazil (Ligeiro et al. 2013a; Callisto et al. 2014; Jiménez-Valencia et al. 2014).

We emphasize that in situations like Latin America, Asia, and Africa, where a long history in comprehensive biological monitoring is lacking, standardization may be an easier process than it has been to date for macroinvertebrate assessments in Europe, where many have applied different sampling, sample processing, and indicator methods for decades.

In our view, it is most important to reduce variability through standardization, provided the methods are tested for precision and accuracy. Thus, in situations where a long history in biological monitoring and vested interests are lacking, we stress the importance of a more pragmatic approach to standardization. Moreover, we recommend adopting or adapting existing national methods described herein that have been implemented across hydromorphologically and climatically diverse regions and states, preceded by a minimal number of pilot studies to ensure their applicability—especially in tropical settings (e.g., Callisto et al. 2014; Jiménez-Valencia et al. 2014). Together with sound scientific research, it is imperative for countries to develop specific legislation and have mandated agencies, with proper training and funding to implement biomonitoring and bioassessment.

In summary, additional performance evaluations (accuracy, precision, discriminatory power, relative costs) are needed regarding targeted habitat (only the richest habitat type) versus site-wide sampling (multiple habitat types), appropriate levels of sampling and processing effort, and standardized indicators to resolve dissimilarities among biomonitoring methods. If universally standardized methods are proposed, some form of calibration is required to maximize the use of historical data generated using sampling and sample processing methods deviating from that standard. Despite the spatially and ecologically diverse range of countries assessed, the structure of sampling methodologies is quite similar and provides confirmation of the key components of a universal sampling methodology.

Global issues such as climate change are creating an environment where there is an increasing need to have universally consistent data collection, processing and storage to enable large-scale trend analysis. We hope this review will provide useful insights for researchers to develop standardized protocols.

Acknowledgments This paper was stimulated by a Fulbright-Brazil grant to R.M. Hughes during a visit to FIOCRUZ, Brazil, and a grant to D.F. Buss (CNPq/PROEP No 400107/2011-2) and generated out of a special symposium at the 2011 annual meeting of the Society for Freshwater Science, at which the coauthors willingly committed their time and travel funds. We are very grateful to Tim Jones, Isabel Pardo, Emilia Mišíková Elexová, Wim Gabriels, and Christian Feld for providing us with the essential information on their national monitoring protocols.

References

Angermeier, P. L., & Karr, J. R. (1986). Applying an index of biotic integrity based on stream-fish communities: considerations in sampling and interpretation. *North American Journal of Fisheries Management, 6*, 418–429.

AQEM Consortium. (2002). Manual for the application of the AQEM method. A comprehensive method to assess European streams using benthic macroinvertebrates, developed for the purpose of the Water Framework Directive. Version 1.0. http://www.eu-star.at/pdf/AqemMacroinvertebrateSamplingProtocol.pdf.

Armitage, P. D., Moss, D., Wright, J. F., & Furse, M. T. (1983). The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. *Water Research, 17*, 333–347.

Bae, M.-J., Kwon, Y., Hwang, S.-J., Chon, T.-S., Yang, H.-J., Kwak, I.-S., Park, J.-H., Ham, S.-A., & Park, Y.-S. (2011). Relationships between three major stream assemblages and their environmental factors in multiple spatial scales. *Annales de Limnologie International Journal of Limnology, 47*, S91–S105.

Bailey, R. C., Kennedy, M. G., Dervish, M. Z., & Taylor, R. M. (1998). Biological assessment of freshwater ecosystems using a reference condition approach: comparing predicted and actual benthic invertebrate communities in Yukon streams. *Freshwater Biology, 39*, 765–774.

Bailey, R. C., Norris, R. H., & Reynoldson, T. B. (2004). *Bioassessment of freshwater ecosystems using the reference condition approach* (p. 170). Norwell: Kluwer Academic Publishers.

Baptista, D. F., Buss, D. F., Egler, M., Giovanelli, A., Silveira, M. P., & Nessimian, J. L. (2007). A multimetric index based on benthic macroinvertebrates for evaluation of Atlantic forest streams of Rio de Janeiro state, Brazil. *Hydrobiologia, 575*, 83–94.

Barbour, M. T., & Gerritsen, J. (1996). Subsampling of benthic samples: a defense of the fixed organism method. *Journal of the North American Benthological Society, 15*, 386–392.

Barbour, M. T., Gerritsen, J., Griffith, G. E., Frydenbourg, R., McCarron, E., White, J. S., & Bastian, M. L. (1996). A framework for biological criteria for Florida streams using benthic macroinvertebrates. *Journal of the North American Benthological Society, 15*, 185–211.

Barbour, M. T., Gerritsen, J., Snyder, B. D., & Stribling, J. B. (1999). *Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish*. 2nd edition. EPA 841-B-99-002. US Environmental Protection Agency, Office of Water, Washington, DC.

Bennett, C., Owen, R., Birk, S., Buffagni, A., Erba, S., Mengin, N., Murray-Bligh, J., Ofenböck, G., Pardo, I., van de Bund, W., Wagner, F., & Wasson, J. G. (2011). Bringing European river quality into line: an exercise to intercalibrate macro-invertebrate classification methods. *Hydrobiologia, 667*, 31–48.

Bijkerk, R. (editor). (2010). *Handboek hydrobiologie: biologisch onderzoek voor de ecologische beoordeling van Nederlandse zoete en brakke oppervlaktewateren*. Stichting Toegepast Onderzoek Waterbeheer, Amersfoort. (Available from: http://handboekhydrobiologie.stowa.nl/Het_Handboek/Het_Handboek.aspx?pId=110).

Blakely, T. J., Eikaas, H. S., & Harding, J. S. (2014). The SingScore: a macroinvertebrate biotic index for assessing the health of Singapore's streams and canals. *Raffles Bulletin of Zoology, 62*, 540–548.

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*, 1–17.

Bonada, N., Prat, N., Resh, V. H., & Statzner, B. (2006). Developments in aquatic insect biomonitoring: a comparative analysis of recent approaches. *Annual Review of Entomology, 51*, 495–523.

Bonar, S., Hubert, W., & Willis, D. (Eds.). (2009). *Standard methods for sampling North American freshwater fishes*. Bethesda: American Fisheries Society.

Borisko, J. P., Kilgour, B. W., Stanfield, L. W., & Jones, F. C. (2007). An evaluation of rapid bioassessment protocols for stream benthic invertebrates in Southern Ontario. *Water Quality Research Journal of Canada, 42*, 184–193.

Bowman, M. F., & Bailey, R. C. (1997). Does taxonomic resolution affect the multivariate structure of freshwater benthic macroinvertebrate communities? *Canadian Journal of Fisheries and Aquatic Sciences, 54*, 1802–1807.

Brua, R. B., Culp, J. M., & Benoy, G. A. (2011). Comparison of benthic macroinvertebrate communities by two methods: kick- and U-net sampling. *Hydrobiologia, 658*, 293–302.

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*, 288–295.

Buss, D. F., & Salles, F. F. (2007). Using Baetidae species as biological indicators of environmental degradation in a Brazilian River Basin. *Environmental Monitoring and Assessment, 130*, 365–372.

Buss, D. F., & Vitorino, A. (2010). Rapid bioassessment protocols using benthic macroinvertebrates in Brazil: evaluation of taxonomic sufficiency. *Journal of the North American Benthological Society, 29*, 562–571.

Buss, D. F., Baptista, D. F., Silveira, M. P., Nessimian, J. L., & Dorvillé, L. F. M. (2002). Influence of water chemistry and environmental degradation on macroinvertebrate assemblages in a river basin in south-east Brazil. *Hydrobiologia, 481*, 125–136.

Buss, D. F., Baptista, D. F., Nessimian, J. L., & Egler, M. (2004). Substrate specificity, environmental degradation and disturbance structuring macroinvertebrate assemblages in Neotropical streams. *Hydrobiologia, 518*, 179–188.

Callisto, M., Hughes, R. M., Lopes, J. M., & Castro, M. A. (Eds). (2014). *Ecological conditions in watersheds of hydropower dams*. Série Peixe Vivo 2, Companhia Energética de Minas Gerais, Belo Horizonte, Minas Gerais.

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*, 680–701.

Cao, Y., Larsen, D. P., Hughes, R. M., Angermeier, P. L., & Patton, T. M. (2002). Sampling effort affects multivariate comparisons of stream communities. *Journal of the North American Benthological Society, 21*, 701–714.

Capítulo, A. R., Tangorra, M., & Ocón, C. (2001). Use of benthic macroinvertebrates to assess the biological status of Pampean streams in Argentina. *Aquatic Ecology, 35*, 109–119.

Carter, J. L., & Resh, V. H. (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.

Carter, J. L., & Resh, V. H. (2013). Analytical approaches used in stream benthic macroinvertebrate biomonitoring programs of State agencies in the United States. *U.S. Geological Survey Open-File Report, 1129*, 50.

Chen, K., Hughes, R. M., Xu, S., Zhang, J., Cai, D., & Wang, B. (2014). Evaluating performance of macroinvertebrate-based predictive and null modeled multimetric indices (MMI) using multi-season and multi-year samples. *Ecological Indicators, 36*, 142–151.

Cho, W.-S., Park, Y.-S., Park, H.-K., Kong, H.-Y., & Chon, T.-S. (2011). Ecological informatics approach to screening of integrity metrics based on benthic macroinvertebrates in streams. *Annales de Limnologie International Journal of Limnology, 47*, S51–S62.

Chon, T.-S., Qu, X., Cho, W.-S., Hwang, H.-J., Tang, H., Liu, Y., Choi, J.-H., Jung, M., Chung, B. S., & Lee, H. Y. (2013). Evaluation of stream ecosystem health and species association based on multi-taxa (benthic macroinvertebrates, algae, and microorganisms) patterning with different levels of pollution. *Ecological Informatics, 17*, 58–72.

Chutter, F. M. (1972). An empirical biotic index of the quality of the water in South African streams and rivers. *Water Resources, 6*, 19–30.

Clapcott, J. E., Collier, K. J., Death, R. G., Goodwin, E. O., Harding, J. S., Kelly, D., & Young, R. G. (2012). Quantifying relationships between land-use gradients and structural and functional indicators of stream ecological integrity. *Freshwater Biology, 57*(1), 74–90.

Clarke, R. T., & Hering, D. (2006). Errors and uncertainty in bioassessment methods—major results and conclusions from the STAR project and their applications using STARBUGS. *Hydrobiologia, 566*, 433–439.

Clarke, R. T., Wright, J. F., & Furse, M. T. (2003). RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecological Modelling, 160*, 219–233.

Clarke, R. T., Lorenz, A., Sandin, L., Schmidt-Kloiber, A., Strackbein, J., Kneebone, N. T., & Haase, P. (2006). Effects of sampling and sub-sampling variation using the STAR-AQEM sampling protocol on the precision of macroinvertebrate metrics. *Hydrobiologia, 566*, 441–459.

Couceiro, S. R. M., Hamada, N., Luz, S. L. B., Forsberg, B. R., & Pimentel, T. P. (2007). Deforestation and sewage effects on aquatic macroinvertebrates in urban streams in Manaus, Amazonas, Brazil. *Hydrobiologia, 575*, 271–284.

Courtemanch, D. L. (1996). Commentary on the subsampling procedures used for rapid bioassessments. *Journal of the North American Benthological Society, 15*, 381–385.

Dallas, H. F. (1997). A preliminary evaluation of aspects of SASS (South African Scoring System) for the rapid bioassessment of water quality in Rivers with particular reference to the incorporation of SASS in a national Biomonitoring Programme. *Southern African Journal of Aquatic Sciences, 23*(1), 79–94.

Dallas, H. F. (2004). Seasonal variability of macroinvertebrate assemblages in two regions of South Africa: implications for aquatic bioassessment. *African Journal of Aquatic Science, 29*(2), 173–184.

Davies, P. E. (1994). *Monitoring river health initiative: river bioassessment manual*. Freshwater Systems, University of Tasmania, Sandy Bay, Tasmania.

Davies, P. E. (2000). Development of a national river bioassessment system (AUSRIVAS) Australia. In J. F. Wright, D. W. Sutcliffe, & M. T. Furse (Eds.), *Assessing the biological quality of freshwaters: RIVPACS and other techniques* (pp. 113–124). Cumbria: Freshwater Biological Association.

Davies, P. E., Harris, J., Hillman, T., & Walker, K. (2010). The Sustainable Rivers Audit: assessing river ecosystem health in the Murray-Darling Basin, Australia. *Marine and Freshwater Research, 61*, 764–777.

Dickens, C. W. S., & Graham, P. M. (2002). The South African Scoring System (SASS) version 5 rapid bioassessment method for rivers. *African Journal of Aquatic Science, 27*, 1–10.

Doberstein, C. P., Karr, J. R., & Conquest, L. L. (2000). The effect of fixed-count subsampling on macroinvertebrate biomonitoring in small streams. *Freshwater Biology, 44*, 355–371.

Ellis, D. (1985). Taxonomic sufficiency in pollution assessment. *Marine Pollution Bulletin, 16*, 459.

Environment Agency. (2012a). *Freshwater macro-invertebrate sampling in rivers.* Operational instruction 018_08.

Environment Agency. (2012b). *Freshwater macro-invertebrate analysis of riverine samples*. Operational instruction 024_08.

Esselman, P. C., Infante, D. M., Wang, L., Cooper, A. R., Wieferich, D., Tsang, Y., Thornbrugh, D. J., & Taylor, W. W. (2013). Regional fish community indicators of landscape disturbance to catchments of the conterminous United States. *Ecological Indicators, 26*, 163–173.

European Commission. (2000). *The EU Water Framework Directive - integrated river basin management for Europe*. http://ec.europa.eu/environment/water/water-framework/index_en.html.

Fenoglio, S., Badino, G., & Bona, F. (2002). Benthic macroinvertebrate communities as indicators of river environment quality: an experience in Nicaragua. *Revista de Biología Tropical, 50*, 1125–1131.

Ferreira, W. R., Paiva, L. T., & Callisto, M. (2011). Development of a benthic multimetric index for biomonitoring of a neotropical watershed. *Brazilian Journal of Biology, 71*, 15–25.

Furse, M. T., Moss, D., Wright, J. F., & Armitage, P. D. (1984). The influence of seasonal and taxonomic factors on the ordination and classification of running-water sites in Great Britain and on the prediction of their macro-invertebrate communities. *Freshwater Biology, 14*, 257–280.

Gabriels, W., Lock, K., de Pauw, N., & Goethals, P. L. M. (2010). Multimetric Macroinvertebrate Index Flanders (MMIF) for biological assessment of rivers and lakes in Flanders (Belgium). *Limnologica, 40*, 199–207.

Gerth, W. J., & Herlihy, A. T. (2006). Effect of sampling different habitat types in regional macroinvertebrate bioassessment surveys. *Journal of the North American Benthological Society, 25*, 501–512.

Ghetti, P. F. (1997). *Manuale di applicazione indice biotico esteso (I.B.E.)*. Trento: Provincia Autonoma di Trento.

Haase, P., & Sundermann, A. (2004). *Standardisierung der erfassungs- und auswertungsmethoden von makrozoobenthos-untersuchungen in fließgewässern (FKZ O 4.02)*. Forschungsinstitut Senckenberg, Biebergemünd. http://www.perlodes.de/download/probenahme-sortierung/

Haase, P., Pauls, S., Sundermann, A., & Zenker, A. (2004b). Testing different sorting techniques in macroinvertebrate samples from running waters. *Limnologica, 34*, 366–378.

Hamada, N., Nessimian, J. L., & Querino, R. B. (Eds.). (2014). *Insetos aquáticos na Amazônia brasileira: taxonomia, biologia e ecologia*. Manaus: Editora INPA.

Harding, J. S., Benfield, E. F., Bolstad, P. V., Helfman, G. S., & Jones, E. B. D., III. (1998). Stream biodiversity: the ghost of land use past. *Proceedings of the National Academy of Science, 95*, 14843–14847.

Hawkins, C. P. (2006). Quantifying biological integrity by taxonomic completeness: evaluation of a potential indicator for use in regional- and global-scale assessments. *Ecological Applications, 16*, 1277–1294.

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

Hellawell, J. M. (1986). *Biological indicators of freshwater pollution and environmental management*. London: Elsevier Applied Science.

Hering, D., Feld, C. K., Moog, O., & 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.

Herlihy, A. T., Paulsen, S. G., Van Sickle, J., Stoddard, J. L., Hawkins, C. P., & Yuan, L. L. (2008). Striving for consistency in a national assessment: the challenges of applying a reference condition approach at a continental scale. *Journal of the North American Benthological Society, 27*, 860–877.

Hilsenhoff, W. L. (1977). *Use of arthropods to evaluate water quality of streams. Technical bulletin 100*. Madison: Department of Natural Resources.

Hilsenhoff, W. L. (1987). An improved biotic index of organic stream pollution. *Great Lakes Entomologist, 20*, 31–39.

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

Hughes, R. M. (1995). Defining acceptable biological status by comparing with reference conditions. In W. Davis & T. Simon (Eds.), *Biological assessment and criteria: tools for water resource planning and decision making* (pp. 31–47). Michigan: Lewis, Chelsea.

Hughes, R. M., & Herlihy, A. T. (2007). Electrofishing distance needed to estimate consistent IBI scores in raftable Oregon rivers. *Transactions of the American Fisheries Society, 136*, 135–141.

Hughes, R. M., & Peck, D. V. (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.

Hughes, R. M., Paulsen, S. G., & Stoddard, J. L. (2000). EMAP-surface waters: a national, multiassemblage, probability survey of ecological integrity. *Hydrobiologia, 422*(423), 429–443.

Humphrey, C. L., Storey, A. W., & Thurtell, L. (2000). AUSRIVAS: operator sample processing errors and temporal variability—implications for model sensitivity. In J. F. Wright, D. W. Sutcliffe, & M. T. Furse (Eds.), *Assessing the biological quality of fresh waters* (pp. 143–165). Ambleside: Freshwater Biological Association.

Jacobsen, D. (1998). The effect of organic pollution on the macroinvertebrate fauna of Ecuadorian highland streams. *Archives of Hydrobiology, 143*, 179–195.

Jiménez-Valencia, J., Kaufmann, P. R., Sattamini, A., Mugnai, R., & Baptista, D. F. (2014). Assessing the ecological condition of streams in a southeastern Brazilian basin using a probabilistic monitoring design. *Environmental Monitoring and Assessment, 186*, 4685–95.

Jones, F. C. (2008). Taxonomic sufficiency: the influence of taxonomic resolution on freshwater bioassessments using benthic macroinvertebrates. *Environmental Reviews, 16*, 45–69.

Jones, J. I., Davy-Bowker, J., Murphy, J. F., & Pretty, J. L. (2010). Ecological monitoring and assessment of pollution in rivers. In L. Batty (Ed.), *Ecology of industrial pollution: remediation, restoration and preservation*. UK: Cambridge Press.

Joy, M. K., & Death, R. G. (2003). Biological assessment of rivers in the Manawatu-Wanganui region of New Zealand using a predictive macroinvertebrate model. *New Zealand Journal of Marine and Freshwater Research, 37*, 367–379.

Jun, Y.-C., Kim, N.-Y., Kwon, S.-J., Han, S.-C., Hwang, I.-C., Park, J.-H., Won, D.-H., Byun, M.-S., Kong, H.-Y., Lee, J.-E., & Hwang, S.-J. (2011). Effects of land use on benthic macroinvertebrate communities: comparison of two mountain streams in Korea. *Annales de Limnologie International Journal of Limnology, 47*, S35–S49.

Jun, Y.-C., Won, D.-H., Lee, S.-H., Kong, D.-S., & Hwang, S.-J. (2012). A multimetric benthic macroinvertebrate index for the assessment of stream biotic integrity in Korea. *International Journal of Environmental Research and Public Health, 9*, 3599–3628.

Junqueira, M. V., Friedrich, G., & Pereira de Araujo, P. R. (2010). A saprobic index for biological assessment of river water quality in Brazil (Minas Gerais and Rio de Janeiro states). *Environmental Monitoring and Assessment, 163*(1–4), 545–554.

Kail, J., Arle, J., & Jähnig, S. C. (2012). Limiting factors and thresholds for macroinvertebrate assemblages in European rivers: empirical evidence from three datasets on water quality, catchment urbanization, and river restoration. *Ecological Indicators, 18*, 63–72.

Karr, J. R. (1981). Assessment of biotic integrity using fish communities. *Fisheries, 6*(6), 21–27.

Karr, J. R., & Chu, E. W. (1999). *Restoring life in running waters: better biological monitoring*. Washington, D.C.: Island Press.

King, R. S., & Richardson, C. J. (2002). Evaluating subsampling approaches and macroinvertebrate taxonomic resolution for wetland bioassessment. *Journal of the North American Benthological Society, 21*, 150–171.

Klemm, D. J., Blocksom, K. A., Fulk, F. A., Herlihy, A. T., Hughes, R. M., Kaufmann, P. R., Peck, D. V., Stoddard, J. L., Thoeny, W. T., Griffith, M. B., & Davis, W. S. (2003). Development and evaluation of a macroinvertebrate biotic integrity index (MBII) for regionally assessing Mid-Atlantic Highlands streams. *Environmental Management, 31*, 656–669.

Lenat, D. R., & Resh, V. H. (2001). Taxonomy and stream ecology - the benefits of genus- and species-level identifications. *Journal of the North American Benthological Society, 20*, 287–298.

Li, J., Herlihy, A. T., Gerth, W., Kaufmann, P. R., Gregory, S. V., Urquhart, S., & Larsen, D. P. (2001). Variability in stream macroinvertebrates at multiple spatial scales. *Freshwater Biology, 46*, 87–97.

Li, F., Cai, Q., Qu, X., Tang, T., Wu, N., Fu, X., Duan, S., & Jähnig, S. C. (2012). Characterizing macroinvertebrate communities across China: large-scale implementation of a self-organizing map. *Ecological Indicators, 23*, 394–401.

Li, L., Liu, L., Hughes, R. M., Cao, Y., & Wang, X. (2014). Towards a protocol for stream macroinvertebrate sampling in China. *Environmental Monitoring and Assessment, 186*, 469–479.

Ligeiro, R., Ferreira, W., Hughes, R. M., & 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.

Ligeiro, R., Hughes, R. M., Kaufmann, P. R., Macedo, D. R., Firmiano, K. R., Ferreira, W., Oliveira, W. D., Melo, A. S., & 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.

Mackey, A. P., Cooling, D. A., & Berrie, A. D. (1984). An evaluation of sampling strategies for qualitative surveys of macro-invertebrates in rivers, using pond nets. *Journal of Applied Ecology, 21*, 515–534.

Makovinská, J., Tóthová, L., Baláži, P., Hlúbiková, D., Mišíková Elexová, E., Šporka, F., Ftorková, L. (2008). *STN 75 7715. Kvalita vody. Biologický rozbor povrchovej vody - Slovak National Standard 75 7715. Water quality. Biological analysis of surface water*. Water Research Institute. Slovak Office of Standards, Metrology And Testing. Bratislava, Slovak Republic.

Marchant, R., Barmuta, L. A., & Chessman, B. C. (1995). Influence of sample quantification and taxonomic resolution on the ordination of macroinvertebrate communities from running waters in Victoria, Australia. *Marine and Freshwater Research, 46*, 501–506.

Marques, M. M., & Barbosa, F. (2001). Biological quality of waters from an impacted tropical watershed (middle Rio Doce basin, southeast Brazil), using benthic macroinvertebrate communities as an indicator. *Hydrobiologia, 457*, 69–76.

Marzin, A., Archaimbault, V., Belliard, J., Chauvin, C., Delmas, F., & Pont, D. (2012). Ecological assessment of running waters: do macrophytes, macroinvertebrates, diatoms and fish show similar responses to human pressures? *Ecological Indicators, 23*, 56–65.

Meier, C., Hering, D., Biss, R., Bohmer, J., Rawer-Jost, C., Zenker, A., Haase, P., Scholl, F., Rolauffs, P., & Sundermann, A. (2006). *Weiterentwicklung und anpassung des nationalen bewertungssystems für makrozoobenthos an neue internationale vorgaben*. Essen: University of Duisburg-Essen.

Melo, A. S. (2005). Effects of taxonomic and numeric resolution on the ability to detect ecological patterns at a local scale using stream macroinvertebrates. *Archives für Hydrobiologie, 164*, 309–323.

Menezes, S., Baird, D. J., & Soares, A. M. V. M. (2010). Beyond taxonomy: a review of macroinvertebrate trait-based community descriptors as tools for freshwater biomonitoring. *Journal of Applied Ecology, 47*, 711–719.

Metcalfe, J. L. (1989). Biological water quality assessment of running waters based on macroinvertebrate communities: history and present status in Europe. *Environmental Pollution, 60*, 101–139.

Miserendino, M. L., Brand, C., & Di Prinzio, C. Y. (2008). Assessing urban impacts on water quality, benthic communities and fish in streams of the Andes Mountains, Patagonia (Argentina). *Water, Air, and Soil Pollution, 194*, 91–110.

Mišíková, E. E., Haviar, M., Lešťáková, M., & Ščerbáková, S. (2010). Checklist of taxa examined at localities monitored in the Slovak surface water bodies - Benthic invertebrates. *Acta Environmentalica Universitatis Comenianeae (Bratislava), 18*(1), 1–335.

Moreno, P., Franca, J. S., Ferreira, W. R., Paz, A. D., Monteiro, I. M., & Callisto, M. (2009). Use of the BEAST model for biomonitoring water quality in a neotropical basin. *Hydrobiologia, 630*, 231–242.

Morse, J. C., Bae, Y. J., Munkhjargal, G., Sangpradub, N., Tanida, K., Vshivkova, T. S., Wang, B., Yang, L., & Yule, C. M. (2007). Freshwater biomonitoring with macroinvertebrates in East Asia. *Frontiers in Ecology and the Environment, 5*, 33–42.

Moulton, S. R., II, Kennan, J. G., Goldstein, R. M., & Hambrook, J. A. (2002). *Revised protocols for sampling algal, invertebrate, and fish communities as part of the National Water-Quality Assessment Program. Open-File Report 02-150. U.S.* Reston: Geological Survey.

Moya, N., Hughes, R. M., Dominguez, E., Gibon, F. M., Goita, E., & Oberdorff, T. (2011). Macroinvertebrate-based multimetric predictive models for measuring the biotic condition of Bolivian streams. *Ecological Indicators, 11*, 840–847.

Mugnai, R., Oliveira, R. B. S., Carvalho, A. L., & Baptista, D. F. (2008). Adaptation of the indice biotico esteso (IBE) for water quality assessment in rivers of Serra do Mar, Rio de Janeiro State, Brazil. *Tropical Zoology, 21*, 57–74.

Mugnai, R., Nessimian, J.L., & Baptista, D.F. (2010). *Manual de identificação de macroinvertebrados aquáticos do Estado do Rio de Janeiro*. Rio de Janeiro: Editora Technical Books. 176p.

Nichols, S. J., Robinson, W. A., & Norris, R. H. (2010). Using the reference condition maintains the integrity of a bioassessment program in a changing climate. *Journal of the North American Benthological Society, 29*, 1459–1471.

Norris, R. H., McElravy, E. P., & Resh, V. H. (1995). The sampling problem. In P. Calow & G. E. Petts (Eds.), *The rivers handbook* (pp. 282–306). Oxford: Blackwell Scientific Publications.

Oliveira, R. B. S., Mugnai, R., Castro, C. M., & Baptista, D. F. (2011a). Determining subsampling effort for the development of a rapid bioassessment protocol using benthic macroinvertebrates in streams of southeastern Brazil. *Environmental Monitoring and Assessment, 175*, 75–85.

Oliveira, R. B. S., Mugnai, R., Castro, C. M., Baptista, D. F., & Hughes, R. M. (2011b). Towards a rapid bioassessment protocol for wadeable streams in Brazil: development of a multimetric index based on benthic macroinvertebrates. *Ecological Indicators, 11*, 1584–1593.

Ollis, D. J., Dallas, H. F., Esler, K. J., & Boucher, C. (2006). Bioassesment of the ecological integrity of river ecosystems using aquatic macroinvertebrates: an overview with a focus on South Africa. *African Journal of Aquatic Science, 31*(2), 205–227.

Omernik, J. M. (1987). Ecoregions of the conterminous United States. *Annals of the Association of American Geographers, 77*(1), 118–125.

Omernik, J. M., & Griffith, G. E. (2014). Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework. *Environmental Management*. doi:10.1007/s00267-014-0364-1.

Pardo, I., García, L., Delgado, C., Costas, N., & Abraín, R. (2010). *Protocolos de muestreo de comunidades biológicas acuáticas fluviales en el ámbito de las Confederaciones Hidrográficas del Miño-Sil y Cantábrico*. Convenio entre la Universidad de Vigo y las Confederaciones Hidrográficas del Miño-Sil y Cantábrico. 68pp. NIPO 783-10-001-8.

Park, Y.-S., Chon, T.-S., Kwak, I.-S., & Lek, S. (2004). Hierarchical community classification and assessment of aquatic ecosystems using artificial neural networks. *Science of the Total Environment, 327*, 105–22.

Parsons, M., & Norris, R. H. (1996). The effect of habitat specific sampling on biological assessment of water quality using a predictive model. *Freshwater Biology, 36*, 419–434.

Paulsen, S. G., Mayio, A., Peck, D. V., Stoddard, J. L., Tarquinio, E., Holdsworth, S. M., Van Sickle, J., Yuan, L. L., Hawkins, C. P., Herlihy, A., Kaufmann, P. R., Barbour, M. T., Larsen, D. P., & Olsen, A. R. (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.

Peck, D. V., Herlihy, A. T., Hill, B. H., Hughes, R. M., Kaufmann, P. R., Klemm, D. J., Lazorchak, J. M., McCormick, F. H., Peterson, S. A., Ringold, P. L., Magee, T., & Cappaert, M. (2006). *Environmental monitoring and assessment program—surface waters western pilot study: field operations manual for Wadeable streams. EPA 620/R-06/003. U.S.* Washington: Environmental Protection Agency, Office of Research and Development.

Petkovska, V., & Urbanič, G. (2010). Effect of fixed-fraction subsampling on macroinvertebrate bioassessment of rivers. *Environmental Monitoring and Assessment, 169*, 179–201.

Plafkin, J. L., Barbour, M. T., Porter, K. D., Gross, S. K., & Hughes, R. M. (1989). *Rapid bioassessment protocols for use in streams and rivers: benthic macroinvertebrates and fish* (EPA/444/4-89/001). Washington, DC: U.S. Environmental Protection Agency.

Pont, D., Hugueny, B., Beier, U., Goffaux, D., Melcher, A., Noble, R., Rogers, C., Roset, N., & Schmutz, S. (2006). Assessing river biotic condition at the continental scale: a European approach using functional metrics and fish assemblages. *Journal of Applied Ecology, 43*, 70–80.

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

Qu, X. D., Bae, M.-J., Chon, T.-S., & Park, Y.-S. (2013). Evaluation of subsampling efforts in estimating community indices and community structures. *Ecological Informatics, 17*, 3–13.

Reece, P. F., Reynoldson, T. B., Richardson, J. S., & Rosenberg, D. M. (2001). Implications of seasonal variation for biomonitoring with predictive models in the Fraser River catchment, British Columbia. *Canadian Journal of Fisheries and Aquatic Sciences, 58*, 1411–1418.

Resh, V. H., & Jackson, J. K. (1993). Rapid assessment approaches to biomonitoring using benthic macroinvertebrates. In D. M. Rosenberg & V. H. Resh (Eds.), *Freshwater biomonitoring and benthic macroinvertebrates* (pp. 195–233). New York: Chapman and Hall.

Resh, V. H., & McElravy, E. P. (1993). Contemporary quantitative approaches to biomonitoring using benthic macroinvertebrates. In D. M. Rosenberg & V. H. Resh (Eds.), *Freshwater biomonitoring and benthic macroinvertebrates* (pp. 159–194). New York: Chapman and Hall.

Resh, V. H., Norris, R. H., & Barbour, M. T. (1995). Design and implementation of rapid assessment approaches for water resource monitoring using benthic macroinvertebrates. *Australian Journal of Ecology, 20*, 108–121.

Reynolds, L., Herlihy, A. T., Kaufmann, P. R., Gregory, S. V., & Hughes, R. M. (2003). Electrofishing effort requirements for assessing species richness and biotic integrity in western Oregon streams. *North American Journal of Fisheries Management, 23*, 450–461.

Reynoldson, T. B., Bailey, R. C., Day, K. E., & Norris, R. H. (1995). Biological guidelines for freshwater sediment based on benthic assessment of sediment (the BEAST) using a multivariate approach for predicting biological state. *Australian Journal of Ecology, 20*, 198–219.

Reynoldson, T. B., Norris, R. H., Resh, V. H., Day, K. E., & Rosenberg, D. M. (1997). The reference condition approach: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinveterbates. *Journal of the North American Benthological Society, 16*, 833–852.

Reynoldson, T. B., Bombardier, M., Donald, D. B., O'Neill, H., Rosenberg, D. M., Shear, H., Tuominen, T. M., & Vaughan, H. H. (1999). *Strategy for a Canadian aquatic biomonitoring network*. National Water Research Institute, Environment Canada, Burlington, ON. NWRI Contribution No. 99-248. 24 p.

Reynoldson, T. B., Rosenberg, D. M., & Resh, V. H. (2001). Comparison of models predicting invertebrate assemblages for biomonitoring in the Fraser River catchment, British Columbia. *Canadian Journal of Fisheries and Aquatic Sciences, 58*, 1395–1410.

Rosenberg, D. M., & Resh, V. H. (Eds.). (1993). *Freshwater biomonitoring and benthic macroinvertebrates*. New York: Chapman and Hall.

Rosenberg, D. M., T. B. Reynoldson, and V. H. Resh. (1999). *Establishing reference condition for benthic invertebrate monitoring in the Fraser River catchment, British Columbia, Canada*. Environment Canada, Vancouver BC. DOE-FRAP 1998-32. 149 p.

Schiller, C. (2003). *AUSRIVAS protocol development and testing*. Water Ecoscience Report Number 3044/2003. Water Ecoscience, Monitoring river health initiative technical report number 36, Department of the Environment and Heritage.

Schmidt-Kloiber, A., & Nijboer, R. C. (2004). The effect of taxonomic resolution on the assessment of ecological water quality classes. *Hydrobiologia, 516*, 269–283.

Schneck, F., & Melo, A. S. (2010). Reliable sample sizes for estimating similarity among macroinvertebrate assemblages in tropical streams. *Annals of Limnology International Journal of Limnology, 46*, 93–100.

Simon, T. P., & Sanders, R. E. (1999). Applying an index of biotic integrity based on great-river fish communities: considerations in sampling and interpretation. In T. P. Simon (Ed.), *Assessing the sustainability and biological integrity for water resources using fish communities* (pp. 474–505). Boca Raton: CRC Press.

Simpson, J., & Norris, R. H. (2000). Biological assessment of water quality: development of AUSRIVAS models and outputs. In J. F. Wright, D. W. Sutcliffe, & M. T. Furse (Eds.), *RIVPACS and similar techniques for assessing the biological quality of freshwaters* (pp. 125–142). Ableside: Freshwater Biological Association and Environment Agency.

Somers, K. M., Reid, R. A., & David, S. M. (1998). Rapid biological assessments: how many animals are enough? *Journal of the North American Benthological Society, 17*, 348–358.

Splunder, I. van, Pelsma, T. A. H. M., & Bak, A. (editors). (2006). *Richtlijnen monitoring oppervlaktewater: Europese Kaderrichtlijn Water*, versie 1.3. Landelijk Bestuurlijk Overleg Water.

Šporka, F., Vlek, H. E., Bulánková, E., & Krno, I. (2006). Influence of seasonal variation on bioassessment of streams using macroinvertebrates. *Hydrobiologia, 566*, 543–555.

Stark, J. D. (1985). *A macroinvertebrate community index of water quality for stony streams. Water & Soil Miscellaneous Publication 87* (p. 53). Wellington: National Water and Soil Conservation Authority.

Stark, J. D. (1998). SQMCI: a biotic index for freshwater macroinvertebrate coded abundance data. *New Zealand Journal of Marine and Freshwater Research, 32*, 55–66.

Stark, J. D., & Maxted, J. R. (2010). A biotic index for New Zealand's soft-bottomed streams. *New Zealand Journal of Marine and Freshwater Research, 41*, 43–61.

Stark, J. D., Boothroyd, I. K., Harding, J. S., Maxted, J. R., & Scarsbrook, M. R. (2001). *Protocols for sampling macroinvertebrates in wadeable streams. New Zealand Macroinvertebrate Working group report no. 1* (p. 57). Wellington: Ministry for the Environment.

Statzner, B., Hoppenhaus, K., Arens, M. F., & Richoux, P. (1997). Reproductive traits, habitat use and templet theory: a synthesis of world-wide data on aquatic insects. *Freshwater Biology, 38*, 109–35.

Statzner, B., Hildrew, A. G., & Resh, V. H. (2001). Species traits and environmental constraints: entomological research and the history of ecological theory. *Annual Review of Entomology, 46*, 291–316.

Stoddard, J. L., Larsen, D. P., Hawkins, C. P., Johnson, R. K., & Norris, R. H. (2006). Setting expectations for the ecological condition of streams: the concept of reference condition. *Ecological Applications, 16*, 1267–1276.

Stoddard, J. L., Herlihy, A. T., Peck, D. V., Hughes, R. M., Whittier, T. R., & 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.

Tarras-Wahlberg, N. H., Flachier, A., Lang, S. N., & Sangfors, O. (2001). Environmental impacts and metal exposure of aquatic ecosystems in rivers contaminated by small scale gold mining: the Puyango River basin, southern Ecuador. *Science of the Total Environment, 278*, 239–261.

Thirion, C. (2008). *River ecoclassification: manual for ecostatus determination (version2) module E: Macroinvertebrate Response Assessment Index (MIRAI) (Volume1)*. Report to the Water Research Commission. WRC report no TT332/08.

USEPA U.S. Environmental Protection Agency. (2013). *National rivers and streams assessment 2008-2009: a collaborative survey*. EPA/841/D-13/001. Office of wetlands, oceans and watersheds and office of research and development, Washington, DC.

Verdonschot, P. F. M., & Nijboer, R. C. (2004). Testing the European stream typology of the water framework directive for macroinvertebrates. *Developments in Hydrobiology, 175*, 35–54.

Vlek, H. E., Šporka, F., & Krno, I. (2006). Influence of macroinvertebrate sample size on bioassessment of streams. *Hydrobiologia, 566*, 523–542.

Walsh, C. J. (1997). A multivariate method for determining optimal subsample size in the analysis of macroinvertebrate samples. *Marine and Freshwater Research, 48*, 241–248.

Wang, L., Infante, D., Esselman, P., Cooper, A., Wu, D., Taylor, W., Beard, D., Whelan, G., & Ostroff, A. (2011). A hierarchical spatial framework and database for the national river fish habitat condition assessment. *Fisheries, 36*, 436–449.

Weigel, B. M., Henne, L. J., & Martínez-Riveira, L. M. (2002). Macroinvertebrate-based index of biotic integrity for protection of streams in west-central Mexico. *Journal of the North American Benthological Society, 21*, 686–700.

Whittier, T. R., & Van Sickle, J. (2010). Macroinvertebrate tolerance levels and an assemblage tolerance index (ATI) for western USA streams and rivers. *Journal of the North American Benthological Society, 29*, 852–866.

Whittier, T. R., Stoddard, J. L., Larsen, D. P., & Herlihy, A. T. (2007). Selecting reference sites for stream biological assessment: best professional judgement or objective criteria. *Journal of the North American Benthological Society, 26*, 349–360.

Winget, R. N., & Mangum, F. A. (1979). *Biotic condition index: integrated biological, physical, and chemical stream parameters for management*. Ogden: U.S. Forest Service, Intermountain Region.

Winterbourn, M. J, Gregson, K. L. D., & Dolphin, C. H. (2006). *Guide to the aquatic insects of New Zealand*. Fourth Edition. Bulletin of the Entomological Society of New Zealand 14. 108p.

Won, D.-H., Jun, Y.-C., Kwon, S.-J., Hwang, S.-J., Ahn, K.-G., & Lee, J. K. (2006). Development of Korean Saprobic Index using benthic macroinvertebrates and its application to biological stream environment assessment. *Journal of Korean Society on Water Quality, 22*, 768–783 (in Korean).

Wright, J. F., Moss, D., Armitage, P. D., & Furse, M. T. (1984). A preliminary classification of running-water sites in Great Britain based on macroinvertebrate species and the prediction of community type using environmental data. *Freshwater Biology, 14*, 221–256.

Wright, J. F., Sutcliffe, D. W., & Furse, M. T. (Eds.). (2000). *Assessing the biological quality of fresh waters—RIVPACS and other techniques*. Ambleside: Freshwater Biological Association.

Zhang, Y., Richardson, J. S., & Pinto, X. (2009). Catchment-scale effects of forestry practices on benthic invertebrate communities in Pacific coastal streams. *Journal of Applied Ecology, 46*, 1292–1303.

Zuellig, R. E., Carlisle, D. M., Potapova, M., & Meador, M. R. (2012). Variance partitioning of stream diatom, fish, and invertebrate indicators of biological condition. *Freshwater Science, 31*, 182–190.