Hydrobiologia (2006) 566:311–324
M.T. Furse, D. Hering, K. Brabec, A. Buffagni, L. Sandin & P.F.M. Verdonschot (eds), The Ecological Status of European Rivers: Evaluation and Intercalibration of Assessment Methods
DOI 10.1007/s10750-006-0087-2

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

Daniel Hering[1,*], Christian K. Feld[1], Otto Moog[2] & Thomas Ofenböck[2]

[1] *Department of Hydrobiology, University of Duisburg Essen, D-45117, Essen, Germany*

[2]*Department of Water, Atmosphere & Environment, BOKU – University of Natural Resources and Applied Life Sciences, Vienna, Max Emanuel-Strasse 17, 1180, Vienna, Austria*

*(*Author for correspondence: E-mail daniel.hering@uni-essen.de)*

Key words: biological quality elements, macroinvertebrates, bioassessment, ecological status, stressor-specific Multimetric Index, Water Framework Directive

Abstract

The requirements of the European Water Framework Directive (WFD), aimed at an integrative assessment methodology for evaluating the ecological status of water bodies are frequently being achieved through multimetric techniques, i.e. by combining several indices, which address different stressors or different components of the biocoenosis. This document suggests a normative methodology for the development and application of Multimetric Indices as a tool with which to evaluate the ecological status of running waters. The methodology has been derived from and tested on a European scale within the framework of the AQEM and STAR research projects, and projects on the implementation of the WFD in Austria and Germany. We suggest a procedure for the development of Multimetric Indices, which is composed of the following steps: (1) selection of the most suitable form of a Multimetric Index; (2) metric selection, broken down into metric calculation, exclusion of numerically unsuitable metrics, definition of a stressor gradient, correlation of stressor gradients and metrics, selection of candidate metrics, selection of core metrics, distribution of metrics within the metric types, definition of upper and lower anchors and scaling; (3) generation of a Multimetric Index (general or stressor-specific approach); (4) setting class boundaries; (5) interpretation of results. Each step is described by examples.

Introduction

The "ecological status" of rivers, which is mainly based on their biotic components, is an important parameter for European water management (European Water Framework Directive 2000/60/EC; WFD). To assess the ecological status of a water body the taxonomic composition, abundance, ratio of disturbance sensitive taxa to insensitive taxa, and the diversity of biological indicators, have to be considered and compared to respective target values under reference conditions. This ensures the adaptation of the assessment models to a stream typology based on typological descriptors, such as ecoregions, bioregions, catchment size, and altitude. Thus, the WFD forces a re-orientation of existing monitoring procedures towards an integrative type- and reference-specific approach (Heiskanen et al., 2004).

Going far beyond the traditional procedure of documenting biological water quality with respect to organic pollution, the assessment of the ecological status of water bodies under the WFD has to document the relationships between aquatic biota and manifold environmental pressures, particularly the hydrological, morphological, and physical–chemical components. The experiences of the EU-funded AQEM and STAR projects show that the multimetric approach is a valuable procedure for bridging the gap between the current methodologies and future need for evaluating the ecological status of water bodies (Hering et al., 2004a; Furse et al., 2006). Similar experiences have been made particularly in the United States, where Multimetric Indices are frequently used in routine water management (Davis & Simon, 1995; Hughes et al., 1998; Barbour et al., 1999; Karr & Chu, 1999).

Multimetric Indices are now a commonly used tool in regionalised assessment systems for describing the quality of fresh- and brackish water ecosystems (rivers, lakes, transitional waters, wetlands; Hughes & Oberdorff, 1999). The multimetric approach attempts to provide an integrated analysis of the biological community of a site by deriving a variety of biological measures and knowledge of a site's fauna (Karr & Chu, 1999). Within a multimetric index, each single component metric is predictably and reasonably related to specific impacts caused by environmental alterations. For example, while the proportion of different feeding types is suited to assess the trophic integrity of an ecosystem, saprobic or acid indices provide a measure with which to directly assess the impact of certain pollutants and acidification, respectively. Thus, the Multimetric Index considers multiple impacts and combines individual metrics (e.g. saprobic indices, diversity indices, feeding type composition, current preferences, etc.) into a unitless measure, which can be used to assess a site's overall condition. By combining different categories of metrics (e.g. taxa richness, diversity measures, proportion of sensitive and tolerant species, trophic structure) reflecting different environmental conditions and aspects of the community the multimetric assessment is regarded as a more reliable tool than assessment methods based on single metrics (Barbour et al., 1995, 1999; Klemm et al., 2003). The multimetric approach was first developed by Karr (1981) using fish as indicators to describe stream quality. Since the development of Karr's Index of Biotic Integrity (IBI) numerous multimetric indices have been developed (Plafkin et al., 1989; Resh et al., 2000; Hering et al., 2004b; Ofenböck et al., 2004). In principle, Multimetric Indices can be applied to different types of ecosystem (rivers, lakes, transitional waters, wetlands, forests) and to different Biological Quality Elements (fish, benthic invertebrates, macrophytes, phytoplankton, phytobenthos, or other biota) and provide a flexible tool with regard to the set of components.

Within the AQEM and STAR projects Multimetric Indices have been developed for various river types throughout Europe. The procedure has been intensively discussed with both the project consortia and the water authorities, particularly in Germany and Austria, where Multimetric Indices are currently applied in water management. This document is based on the experiences gained during this implementation process.

The experiences of the AQEM and STAR projects clearly show that to enhance comparability between assessment systems the procedure of developing and applying a Multimetric Index needs to be standardised. Aim of this paper is to condense the experiences in developing Multimetric Indices gained in the AQEM and STAR projects into a more generally applicable approach, which may also be useful for ecosystem types others than rivers. In a continent like Europe, where both river biota and political and economic conditions are heterogeneous, a single approach for river assessment, which is likely to be used by all water managers, is unrealistic. At least there is the need to distinguish between simple, unspecific methods, which are useful as a first attempt in areas with little experiences in assessment, and more complex, stressor-specific approaches. Thus, we suggest alternatives differing in their degree of precision, but which always use the same basic steps.

The principles of developing a Multimetric Index

A "metric" is defined as a measurable part or process of a biological system empirically shown to change in value along a gradient of human influence (Karr & Chu, 1999). It reflects specific and predictable responses of the community to human activities, either to a single impact factor or to the cumulative effects of multiple human impairments within a watershed. Metrics are addressing comparable ecological aspects of a community, regardless of the stressor they are responding to. The following metric types can be distinguished:

- *Composition/abundance metrics*. All metrics giving the relative proportion of a taxon or taxonomic group with respect to its total number or abundance, respectively.
- *Richness/diversity metrics*. All metrics giving the number of species, genera, or higher taxa within a certain taxonomical entity, including the total number of taxa, all diversity indices.
- *Sensitivity/tolerance metrics*. All metrics related to taxa known to respond sensitively or tolerantly to a stressor or a single aspect of the stressor, respectively, either using presence/absence or abundance information.
- *Functional metrics*. All metrics addressing the ecological function of taxa (other than their sensitivity to stress), such as feeding types, habitat and current preferences, ecosystem type preferences, life cycle parameters, biometric parameters. They can be based on taxa abundance.

The procedure of data analysis during the development of a Multimetric Index typically involves the following steps:

- Selection of the most suitable form of a Multimetric Index
- Metric selection
- Metric calculation
  - Exclusion of numerically unsuitable metrics
  - Definition of a stressor gradient
  - Correlation of stressor gradients and metrics
  - Selection of candidate metrics
  - Selection of core metrics
  - Distribution of metrics within the metric types
  - Definition of upper and lower anchors and scaling
- Generation of a Multimetric Index
  - Development of a Multimetric Index (general approach)
  - Development of a Multimetric Index (stressor-specific approach)
- Setting class boundaries
- Interpretation of results

Aspects of methods needed to gain comparable taxa lists, i.e. sampling, sorting, and proper determination of the sampled individuals (Schmidt-Kloiber & Nijboer, 2004) is not considered further here, since this paper aims at describing the procedure that starts with metric calculation.

Selection of the most suitable form of a Multimetric Index

Depending on purpose, ecosystem type, organism group and available data Multimetric Indices may be designed differently. In many cases a reliable assessment reflecting the integrity of an ecosystem is sufficient, in other cases more specific data on which stressor causes deterioration of the biota is required. Thus, we distinguish two main forms of Multimetric Indices: (1) the general approach and (2) the stressor-specific approach. Stressor-specific Multimetric Indices can only be derived if the development data set includes environmental data reflecting different specific stress types, if different environmental gradients are present in the development data set and if the autecology of the targeted organism group is well known.

Metric selection

Metric calculation

Due to the long-term tradition of macro-invertebrate research, which has led to extensive ecological knowledge of this group of aquatic organisms, numerous metrics and indices have been developed that can simply be derived from taxa lists (e.g. Moog, 1995; Merritt & Cummins, 1996; Schmedtje & Colling, 1996; Tachet et al., 2002). Several software packages (e.g. ECO-PROF; Moog et al., 2001) aid the quick derivation of metrics from those taxa lists, among which the AQEM River Assessment Program (Hering et al., 2004a) provides a tool that is capable of calculating more than 200 macro-invertebrate metrics. Other tools are available for fish (EFI Software: Fame Consortium, 2004), macrophytes and phyto-benthos (Schaumburg et al., 2004).

Exclusion of numerically unsuitable metrics

In order to reduce the long lists of metrics that are quickly and easily processed by software packages, filter procedures have to be applied. These procedures include the identification and exclusion of numerically unsuitable measures, for example, measures with a narrow range of values or with many outliers and extreme values, which can be simply revealed by box-whisker plots (Figs. 1 and 2).

Definition of stressor gradients

It is mandatory that the data set used for development includes data on a gradient of sites, ideally including unimpacted (reference) sites and heavily degraded (poor) sites.

An environmental stressor gradient is ideally represented by a set of sites of one freshwater ecosystem type covering the whole range (high, good, moderate, poor, and bad sites) of the environmental stressor that is to be targeted by the Multimetric System. The gradient may be a continuous measure or may be classified into five classes or even into the two classes "unstressed" and "stressed", only. Stressor gradients provide an invaluable tool by which to minimize the subjective "expert judgement" in pre-classification of sites and the subsequent selection of candidate metrics, which is based on the pre-classification.

Analysis of the gradient may be restricted to a single stressor or may include the impact of multiple stressors. For description of the impact of a single stressor, physical, chemical, or hydromorphological data on the individual sites can be used. We propose to use:

- data on BOD_5 or oxygen content to describe the impact of organic pollution;
- data on BOD_5, $N–NO_2$, chloride, *Escherichia coli*, eventually combined, for a Multimetric Index addressing water pollution in general terms;

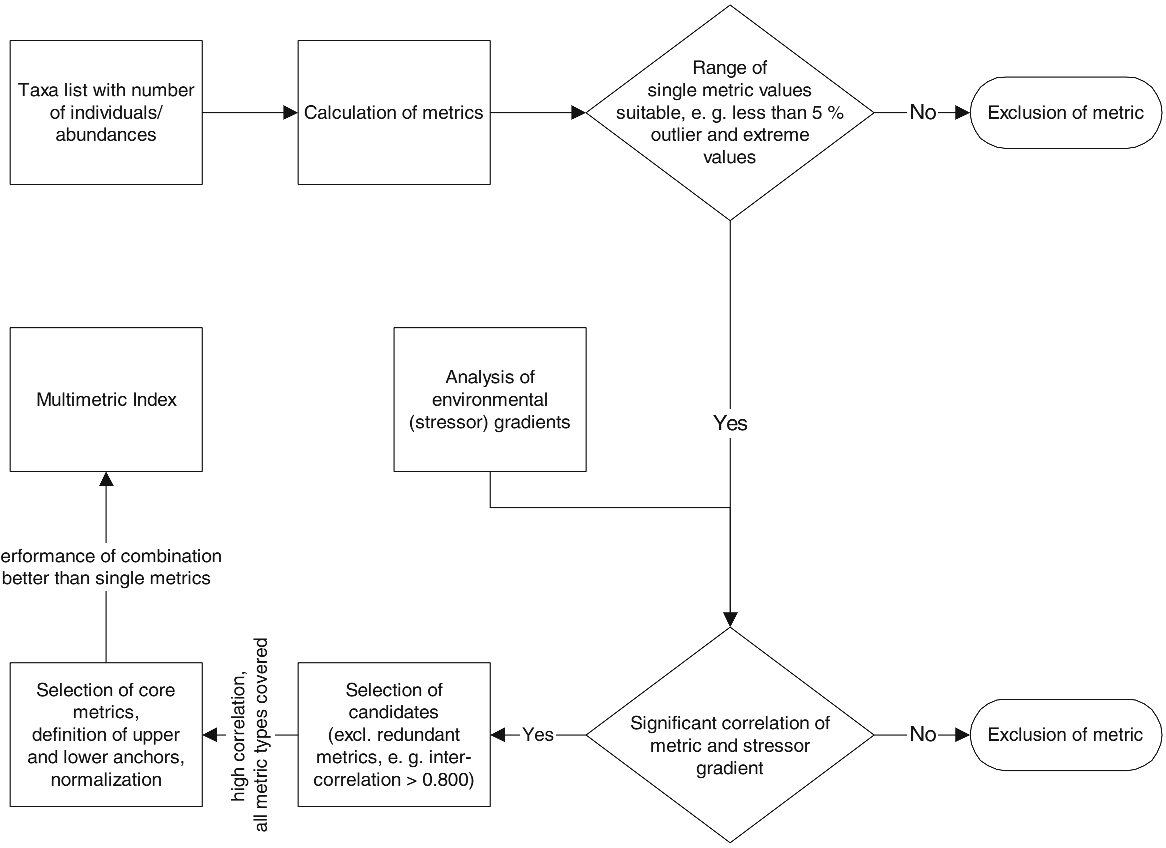

Figure 1. Schematic overview of the steps required to develop a Multimetric Index based on taxa lists.

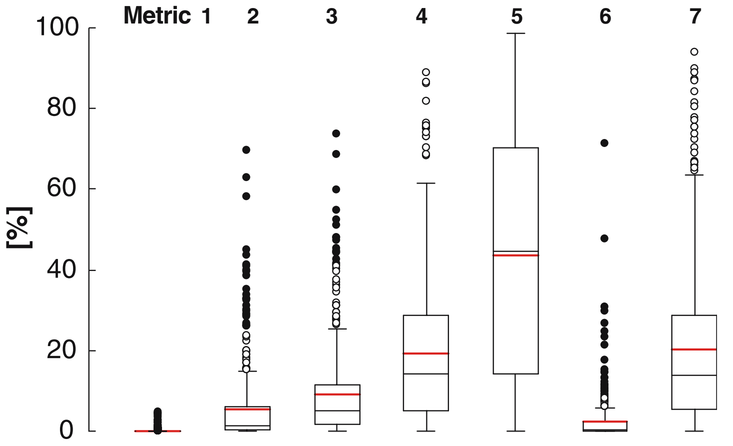

Figure 2. Example for numerically unsuitable (Metrics 1–3, 6) and suitable metrics (Metrics 4–5, 7). Circles indicate outliers (○) and extremes (●).

- data describing the trophic status of sites such as concentrations of phosphorus and nitrogen compounds;
- data that characterise the morphological situation of a site such as the German Structure-Index (Feld, 2004; Lorenz et al., 2004);
- data that characterize the hydrological and hydraulic situation of a site with information on alterations to the discharge regime, damming, residual flows, etc.;
- data on catchment land use for describing general stress gradients (Böhmer et al., 2004a);
- several of the above mentioned data to describe more general types of stress.

A statistical analysis such as PCA (Principal Component Analysis) can be used to reduce the number of variables by (i) calculating hypothetical main gradients of the environmental dataset and (ii) identifying redundant (co-correlating) variables. The direct analysis of metrics and abiotic environmental data is possible with Redundancy Analysis (RDA). The advantage of direct ordination procedures is their aim to fit the main abiotic and biotic gradients. Thus, if the existence of a strong stressor gradient is obvious, this method can be used simultaneously to identify the faunal response (Feld, 2005).

Johnson et al. (2006), and Hering et al. (submitted) defined stress gradients for a subset of the STAR sites by means of a PCA, with data indicating different sources of impairment. The selection of parameters (Table 1) was dependent on (1) their availability and completeness in the dataset (2), their relevance to the targeted stream type and (3) their relevance for the targeted stressor.

Correlation of stressor gradients and metrics

Correlating the results of a metric to the stressor gradient is a central part of the procedure, which can be processed either by looking for significant differences (*t*-test, *U*-test) or by running rank correlation analysis (e.g. Spearman, Kendall). It is also possible to use Pearson' product moment correlation in cases of large data sets, but, this coefficient is prone to partial correlation. Thus, a simple scatter plot may be used to aid the judgement on the strength and quality of metric-stressor correlations.

Selection of candidate metrics

An ideal metric should be responsive to stressors, have a low natural variability, provide a response that can be distinguished from natural variation, and be interpretable. A candidate metric's results must show a significant correlation to the stressor gradient. This correlation can be positive or negative, either across the whole stressor gradient or measured for a part thereof (e.g. only moderate to high quality sites). Metrics fulfilling this criterion are, in principal, suited to assessing the degradation of the freshwater ecosystem type and can be selected as candidate metrics. There are numerous examples

Table 1. Environmental parameters used for calculating stressor gradients

Parameter	Transformation	Lowlands				Mountains			
Pollution/eutrophication		G	P	H	M	G	P	H	M
pH		x	x			x	x		
Conductivity	log 10	x	x			x	x		
BOD5	log 10					x	x		
Oxygen [mg/l]	log 10		x				x		
Ammonium [mg/l]	log 10	x	x			x	x		
Nitrite [mg/l]	log 10					x	x		
Nitrate [mg/l]	log 10	x	x			x	x		
Ortho-phosphate [μg/l]	log 10	x	x			x	x		
Total phosphate [μg/l]	log 10		x				x		
Source pollution (yes/no)	log 10						x		
Non-source pollution (yes/no)	log 10						x		
Eutrophication (yes/no)	log 10						x		
Land use									
Forest catchment [%]	arcsin sq. root	x				x			
Urban sites catchment [%]	arcsin sq. root	x				x			
Natural grassland catchment [%]	arcsin sq. root	x				x			
Cropland catchment [%]	arcsin sq. root	x				x			
Pasture catchment [%]	arcsin sq. root	x				x			
Hydromorphology									
Shading at zenith (foliage cover)	arcsin sq. root	x		x		x		x	
Width woody rip. vegetation [m]	arcsin sq. root	x		x		x		x	
Number of debris dams		x		x		x		x	
Number of logs		x		x		x		x	
Shoreline covered with woody riparian vegetation [%]	arcsin sq. root	x		x		x		x	
No. bank fixation [%]	arcsin sq. root	x		x		x		x	
No. bed fixation [%]	arcsin sq. root	x		x		x		x	
Stagnation (yes/no)		x		x		x		x	
Straightening (yes/no)		x		x		x		x	
Microhabitats									
Hygropetric sites [%]	arcsin sq. root				x				x
Megalithal > 40 cm [%]	arcsin sq. root				x				x
Macrolithal > 20–40 cm [%]	arcsin sq. root				x				x
Mesolithal > 6–20 cm [%]	arcsin sq. root				x				x
Microlithal > 2–6 cm [%]	arcsin sq. root				x				x
Akal > 0.2–2 cm [%]	arcsin sq. root				x				x
Psammal/psammopelal [%]	arcsin sq. root				x				x
Argyllal < 6 μm [%]	arcsin sq. root				x				x
Macro-algae [%]	arcsin sq. root				x				x
Micro-algae [%]	arcsin sq. root				x				x
Submerged macrophytes [%]	arcsin sq. root				x				x
Emergent macrophytes [%]	arcsin sq. root				x				x
Living parts of ter. plants [%]	arcsin sq. root				x				x
Xylal [%]	arcsin sq. root				x				x

Continued on page 317

Table 1. (Continued)

Parameter	Transformation	Lowlands				Mountains			
Pollution/eutrophication		G	P	H	M	G	P	H	M
CPOM [%]	arcsin sq. root				x				x
FPOM [%]	arcsin sq. root				x				x
Debris [%]	arcsin sq. root				x				x

G = general degradation gradient; p = pollution/eutrophication gradient; H = hydromorphology gradient; M = microhabitat gradient (from Hering et al., submitted, altered).

in the literature for the process of selecting candidate metrics (Vlek et al., 2004; Böhmer et al., 2004b; Ofenböck et al., 2004; Buffagni et al., 2004). Three examples on how metrics relate to different stress gradients are given in Figure 3.

Numerous papers describe the possible approaches to metric selection (e.g. Holland, 1990; Barbour et al., 1992; Karr & Kerans,1992; Barbour et al., 1999; Karr & Chu, 1999; Buffagni et al., 2004; Hering et al., 2004b; Ofenböck et al., 2004; Vlek et al., 2004). Based on existing knowledge and literature information, the candidate metrics are selected on the basis of knowledge of the aquatic biota within a geographical entity, e.g. the metric "number of *Corbicula* individuals" would make no sense if this taxon does not occur in the targeted ecoregion. As another example, the inclusion of the metric "morphological deformation of chironomids" will be useless if the administrative framework would not allow financing of the necessary investigations. On the other hand, candidate metrics must fit the sampling method applied. If Chironomid pupae or Annelids are collected with 1000 μm-mesh samplers, metrics derived from those taxa are not likely to be reliable.

After having selected the candidate metrics they need to be evaluated for efficacy and validity. This means that inappropriate metrics have to be eliminated from the process. Metrics have to be considered as inappropriate if they (1) are less than robust and have a high temporal and/or spatial variability that does not allow discrimination between anthropogenic influences and natural variability, (2) do not reflect human impairment and have little relationship to the impacts, (3) are not well founded on ecological principles and understanding; for example, the correlation of land use and the feeding type miner.

Only those metrics that show a quantitative impact-response change across a stressor gradient that is reliable, interpretable and not diffused or obscured by natural variation, must be selected. Moreover, different types of metric should be considered (composition/abundance metrics, richness/diversity metrics; sensitivity/tolerance metrics; functional metrics).

Hering et al. (2004a), who aimed at designing a Multimetric Index for indicating "general degradation", restricted a more extensive list of metrics to those indices which are not explicitly designed to detect organic pollution. A further selection criterion was the taxonomic resolution needed for the metric (order/family vs. genus/species level), which should be achieved by, and comparable among, the majority of taxa lists (e.g. Eurolimpacs, 2004; Schmidt-Kloiber et al., 2006) used for the development process. These criteria resulted in restricting a list of almost 300 to 79 metrics.

Selection of core metrics

Candidate metrics, which can be identified as robust and most informative are scrutinised further in the process of selecting core metrics. To be selected as a core metric two major aspects have to be considered: (1) the metrics should cover the different metric types (Table 2) and (3 2) redundant metrics need to be excluded. Metrics that show strong inter-correlations (Spearman's $r > 0.8$) with one another are defined as redundant. The identification of redundant metrics is aided by triangular cross-correlation matrices and, in case of redundancy, the correlation of each of the pair of metrics with the other metrics is compared in order to finally omit the one that showed the higher overall mean correlation (see examples given in Table 4). For the selection of appropriate

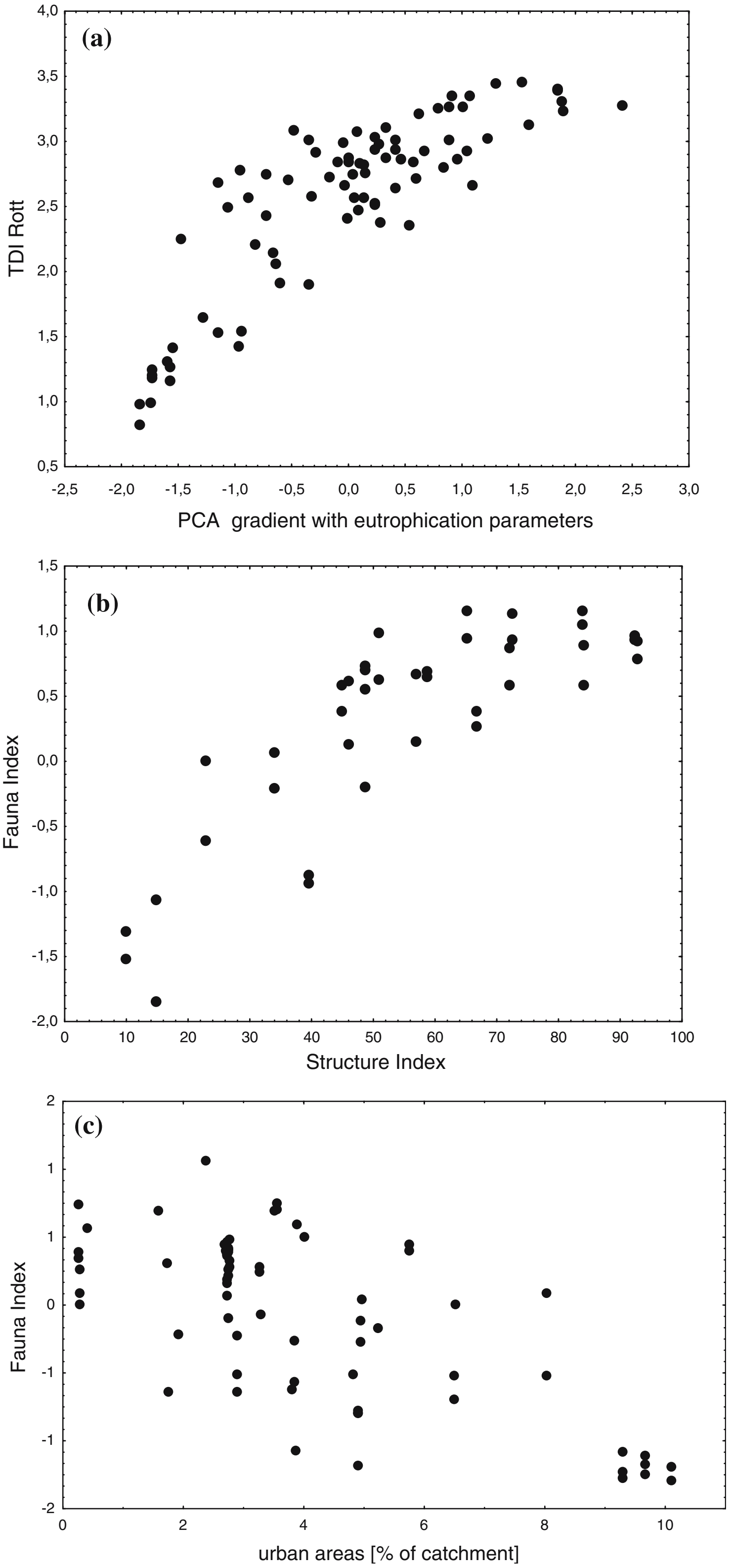
(a)
TDI Rott
4,0
3,5
3,0
2,5
2,0
1,5
1,0
0,5
-2,5
-2,0
-1,5
-1,0
-0,5
0,0
0,5
1,0
1,5
2,0
2,5
3,0
PCA gradient with eutrophication parameters
(b)
Fauna Index
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-2,0
0
10
20
30
40
50
60
70
80
90
100
Structure Index
(c)
Fauna Index
2
1
1
0
-1
-1
-2
0
2
4
6
8
10
urban areas [% of catchment]

Figure 3. (a) Correlation of a periphyton metric (Trophic Diatom Index according to Rott et al., 1999) to a eutrophication gradient. Samples from the STAR lowland rivers (data from Hering et al. in press). (b) Correlation of a benthic invertebrate metric (German Fauna Index D05 according to Lorenz et al., 2004) to hydromorphological quality measured with a structure index (data from Lorenz et al., 2004). (c) Correlation of a benthic invertebrate metric (German Fauna Index D03 according to Lorenz et al., 2004) to catchment land use (share of urban areas) in medium-sized lowland rivers in Germany.

Table 2. Examples for metrics used to assess individual Biological Quality Elements, assigned to metric types

	Composition/abundance metrics	Richness/diversity metrics	Sensitivity/tolerance metrics	Functional metrics
Fish	Population age structure Population size	Diversity (Shannon-Weaver, Margalef) Number of river type specific species	Individuals of tolerant species	Number of rheophile species Number of lithophile species
Benthic invertebrates	[%] EPT	Diversity (Shannon-Weaver Margalef)	Saprobic indices	[%] sand-preferring taxa
	[%] Trichoptera	Number of Trichoptera species	Acid Index (Henrikson & Medin, 1986) German Fauna Index (Lorenz et al., 2004)	[%] shredders, RETI (Schweder, 1992) [%] rheophile species
Macrophytes	[%] *Potamogeton pectinatus*	Diversity (Shannon-Weaver, Margalef) Number of taxa	Mean Trophic Ranking (Holmes et al., 1999)	Ellenberg et al. (1992) numbers (humidity, light, salinity)
Phytobenthos	[%] Pennales (volume) (Mischke & Behrendt, 2005)	Diversity (Shannon-Weaver, Margalef) Number diatom of taxa	Trophic Diatom Index (Kelly & Whitton, 1995) Trophic Index Austria (Rott et al., 1999)	[%] planctonic taxa
Phytoplankton	[%] Pennales (volume) (Mischke & Behrendt, 2005)	Diversity (Shannon-Weaver, Margalef,) Number of Desmid taxa	Rare taxa and indicative taxa (Coesel, 2001) "Index-20" (Mischke & Behrendt, 2005)	[%] planctonic taxa

core metrics, statistical analysis aimed at identifying those variables, which show the strongest relationship to certain environmental stressors, are recommended.

Distribution of metrics within the metric types

Well-constructed Multimetric Indices contain a suggested number of metrics from each type (Table 2) and therefore reflect multiple dimensions of biological systems (Karr & Chu, 1999). About three metrics per metric type is considered ideal. A higher (e.g. to more exhaustively describe the community attributes) or lower (e.g. if fewer suitable metrics can be identified) number of metrics can be included into a Multimetric Index. If there is at least one candidate metric of a particular metric type, then at least one of this metric type must be selected as a core metric to ensure that each metric type is represented in the Multimetric Index. This procedure makes Multimetric Indices more comparable and ensures that different aspects of the community are regarded. The possible combinations of metrics resulting from the selection of candidate metrics must be correlated to the stressor gradient used to select the candidate metrics. For this purpose, all metric results are first scaled by transformation into a score ranging from 0 to 1 (100%). This enables the calculation of means for all candidate metrics.

Those metrics whose combination results in the strongest significant correlation to the stressor gradient should be selected as core metrics.

Table 3. Example for the definition of upper anchors and lower anchors of candidate metrics in the stream type "medium-sized lowland rivers" in Germany (data from Hering et al., 2004a)

Metric	Shannon-diversity	[%] litoral preferring taxa	[%] rheophile taxa	[%] shredderes	German Fauna Index D03	[%] EPT	# Plecoptera taxa	# Trichoptera Taxa
Upper (95%) percentile	3.50	29.07	61.28	36.51	0.89	66.79	2.00	13.00
Lower (5%) percentile	1.54	2.52	6.04	2.88	−1.43	6.87	0.00	0.00
Correlation with land use index								
Correlation coefficient	−0.35	0.41	−0.54	0.39	−0.44	−0.65	−0.37	−0.48
Suggested upper anchor	3.39	−3.60	77.99	−0.40	1.26	82.12	1.69	16.01
Suggested lower anchor	2.29	24.27	11.71	17.93	−0.90	13.36	−0.40	1.71
Correlation with Structure Index								
Correlation coefficient	−0.18	0.55	−0.43	−0.44	−0.75	−0.14	−0.42	−0.53
Suggested upper anchor	2.71	8.06	46.61	20.31	0.59	38.30	0.46	7.35
Suggested lower anchor	2.50	18.95	32.05	7.59	−0.50	31.22	−0.10	3.95
Chosen upper anchor	3.50	3.00	70.00	35.00	1.50	70.00	3.00	15.00
Chosen lower anchor	1.00	25.00	10.00	3.00	−1.50	5.00	0.00	0.00

Three different methods for defining anchors have been applied: (1) 95% and 5% percentile of all data; (2) Spearman Rank Correlation with a land use index; (3) Spearman Rank Correlation with a structure index.

Table 4. Example for a correlation matrix of candidate metrics (invertebrate metrics, stream type "medium-sized lowland rivers" in Germany) (data from Hering et al., 2004a)

	Shannon diversity	[%] litoral preferring taxa	[%] rheophile taxa	[%] shredderes	German Fauna Index D03	[%] EPT	# Plecoptera taxa	# Trichoptera Taxa
Shannon diversity	1.0000							
[%] lioral preferring taxa	−0.1906	1.0000						
[%] rheophile taxa	0.1420	**−0.8505**	1.0000					
[%] shredderes	−0.3349	0.1195	−0.2393	1.0000				
German Fauna Index D03	0.2390	**−0.8020**	0.7911	−0.1247	1.0000			
[%] EPT	0.2874	−0.6495	0.7040	−0.2887	0.6855	1.0000		
# Plecoptera taxa	0.1467	−0.4450	0.4650	−0.0038	0.5168	0.5894	1.0000	
# Trichoptera taxa	0.6000	−0.5165	0.4733	−0.1627	0.6212	0.6881	0.5324	1.0000

Correlation coefficients of individual metrics are given. Bold: Correlation coefficient > 0.8 (one of these metrics needs to be excluded).

Definition of upper and lower anchors and scaling

The upper and lower anchors mark the indicative range of a metric, i.e. the values that are empirically set and defined as "1" (upper anchor) and "0" (lower anchor), respectively, to normalize a metric's result. The upper anchor corresponds to the upper limit of the metric's value under reference conditions. If data on reference sites are available, the upper anchor should be set as a percentile of all the metric values of the reference sites (e.g. 95%, 75% or median, depending on the quality of the reference sites). If few data (e.g. up to 5–10 samples) are available for reference sites, and the site classification is to some extent uncertain, the highest observed value can also be considered (excluding abundance metrics). If there are no data on reference sites but data on sites representing different degrees of stress are available, the upper anchor can be obtained by extrapolation.

The lower anchor corresponds to the lower limit of the metric's value under the worst attainable conditions. If data on sites of bad ecological quality are available, the lower anchor should be set as a percentile (e.g. 5 or 10%) of all metric values of the bad ecological quality sites, or at the lowest value obtained or obtainable. If there are no data on bad ecological quality sites but data on sites representing different degrees of stress are available, the Lower Anchor can be obtained by extrapolation. An example from the German invertebrate assessment system is given in Table 4.

The results of the various core metrics that have been selected for contributing to a Multimetric Index may vary between different ranges of values: while the metric "number of Plecoptera species" can have a value between 0 and *n*, the German Saprobic Index can range from 1.0 to 4.0 and the metric "[%] shredders" from 0 to 100. To combine these individual measures into an integrated Multimetric Index; it is essential to normalize the core metrics via transformation to unitless scores. In practice, each metric result must be translated into a value between 0 and 1 (Ecological Quality Ratio), using the following formula:

$$\text{Value} = \frac{\text{Metric\_result} - \text{Lower\_Anchor}}{\text{Upper\_Anchor} - \text{Lower\_Anchor}}$$

for metrics decreasing with increasing impairment, and

$$\text{Value} = 1 + \frac{\text{Metric\_result} - \text{Lower\_Anchor}}{\text{Upper\_Anchor} - \text{Lower\_Anchor}}$$

for metrics increasing with increasing impairment. Values > 1 are set to 1.

The resulting metric value for a given site is finally expressed as an ecological quality ratio (EQR). The EQR represents the relationship between the values of the biological parameters observed for a given body of surface water and the values for these parameters under the reference conditions applicable to that water body. The ratio is expressed as a numerical value between zero and one: high ecological status is represented by values close to one and bad ecological status by values close to zero.

Generation of a Multimetric Index

The aggregation of metric scores into an index simplifies decision making so that a single value can be used to determine the quality class of a river site. The action, which is potentially needed to improve the ecosystem (e.g. restoration, mitigation, pollution enforcement) is not inherently determined by the index value, but may be deduced from the single component metrics, in addition to the raw data, and consideration of other ecological information (Barbour et al., 1999).

We propose two ways of generating a Multimetric Index: a "general approach" and a "stressor-specific approach". In the "general approach", various metrics are calculated and the results are individually compared to the respective metric values under reference conditions. From this comparison, a score is derived for each metric. These scores are finally combined into a Multimetric Index. The "stressor-specific" approach sorts out the metrics forehand according to their ability to detect the effects of a certain stressor on the targeted biota. Thus, the scores of the metrics addressing a single stressor are first combined into a value reflecting the intensity of this stressor; the assessment results for all stressors are finally combined into the Multimetric Index.

Development of a Multimetric Index (general approach)

The aggregation of metrics into a Multimetric Index should ensure that each metric type is represented by a similar number of metrics (e.g. Karr & Chu, 1999). Nevertheless, the final selection of metrics for a Multimetric Index should produce the strongest multimetric view of biological condition. Therefore, we do not recommend a fixed number of metric types or measures per metric type.

The procedure described by Böhmer et al. (2004a) is based on the assumption that if the same number of metrics has been selected for each metric type, the Multimetric Index can be calculated as the mean of the 0–1 digit scores of all core metrics. This will attribute the same weight to each metric and metric type. If the number of core metrics belonging to different metric types is different, weighting factors can be used so that e.g.

each group of metrics (i.e. clustered within a type) has the same influence on the final Multimetric Index. If, within a metric type, the various core metrics are based on information of different confidence levels (e.g. one is based on the whole Invertebrate community, while the others on single insect orders) weighting factors can be applied to the metrics so that the more inclusive metrics contribute to a greater extent to the final score.

Development of a Multimetric Index (stressor-specific approach)

For the ''stressor-specific approach'' almost exactly the same steps as for the ''general approach'' are required. However, all the above described steps (from the generation of environmental gradients to the scaling of metrics) should be done separately for different environmental gradients, representing different stressors. This procedure results in a separate list of core metrics for each stressor, e.g. organic pollution, acidification or hydromorphological degradation.

The scores of those core metrics, which have been selected using the gradient of a single stressor, must the be combined into a Multimetric Index by calculating the mean of their 0–1 scores. This step results in a quality class for each stressor, e.g. ''organic pollution'' and ''acidification''.

If the same degree of confidence is expected for the different stressor-specific indices, the resulting stressor-specific quality classes are converted into the ecological quality class using the worst result of all stressor-specific quality classes. Otherwise, priority can be given to the most robust metric, the results of the other metric(s) being used to confirm the classification obtained. Weighting factors can be considered as explained above.

Setting Class boundaries

The final Multimetric Index provides a score that represents the overall relationship between the combined values of the biological parameters observed for a given site and the expected value under reference conditions. This score is – as for single metrics – expressed as a numerical value between zero and one. This range can be subdivided into any number of categories corresponding to various levels of impairment. Because the metrics are scaled to reference conditions and expectations for the stream classes, any decision on subdivision should reflect the distribution of the scores for the reference sites. We propose quality classes with equal ranges to provide five ordinal rating categories for assessment of impairment in accordance with the demands of the WFD, using the following scheme for setting class boundaries:

reference ≥ 0.8
good $\geq 0.6 < 0.8$
moderate $\geq 0.4 < 0.6$
poor $\geq 0.2 < 0.4$
bad < 0.2

The more metrics are included into the Multimetric Index, the more the index values under reference conditions will diverge from 1, because even under the most pristine conditions, not all metrics will reach maximum levels in a single site. Alternative: Therefore, it is recommended not to use the best available values as reference values, but e.g. to use the 25% percentile of index values from reference sites as the class boundary for reference conditions.

Interpretation of results

Multimetric Indices can be easily interpreted, which is regarded as a main advantage of this type of bioassessment. However, since European water managers have only little experience with Multimetric Indices, an aid for the interpretation of results is highly recommended, particularly if the ''general approach'' is applied, which does not inherently distinguish between stress types. An interpretation aid should include the values to be expected under reference conditions, the stress type the metric is most strongly reacting to and the restoration measures needed to improve the metric.

Conclusions

Multimetric Indices provide a valuable tool for assessing various types of freshwater ecosystems, since they integrate different stressors and different components of the community. Thus, they can be adapted to the specific conditions of a river type or lake type in an optimal way, by considering the most relevant stressors, and specific characters of the biocoenosis. However, to gain a certain level of comparability, the development of Multimetric Indices should be carried out in an analogous way. By considering the steps described in this paper, comparability of Multimetric Indices can be ensured, without loosing the degree of freedom which is necessary to cope with the natural variability of river and lake types and their communities. Thus, the procedure described here may be helpful for consideration as an international standard.

Acknowledgments

AQEM and STAR were funded by the European Commission, 5th Framework Program, Energy, Environment and Sustainable Development, Key Action Water, Contract no. EVK1-CT 1999-00027 and EVK1-CT-2001-00089. The authors acknowledge the support of all their colleagues who have contributed to the work described and want to express their special thanks to Mike Furse the terrific co-ordinator of STAR. We would also like to acknowledge Mary Burgis for linguistic help and two anonymous reviewers for valuable comments, which significantly improved the paper.

References

Barbour, M. T., J. L. Plafkin, B. P. Bradley, C. G. Graves & R. W. Wisseman, 1992. Evaluation of EPA's rapid bioassessment benthic metrics: Metric redundancy and variability among reference stream sites. Environmental Toxicology and Chemistry 11: 437–449.

Barbour, M. T., J. B. Stribling & J. R. Karr, 1995. Multimetric approach for establishing biocriteria. Biological assessment and criteria. In Davies, W. S. & T. P. Simon (eds), Tools for water resource planning and decision making. CRC Press, Boca Raton: 63–77.

Barbour, M. T., J. Gerritsen, B. D. Snyder & J. B. Stribling, 1999. Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish (2nd edn.nd ed.). U.S. EPA. Office of Water, Washington, DC, EPA/841-B-98-010.

Böhmer, J., C. Rawer-Jost & A. Zenker, 2004a. Multimetric assessment of data provided by water managers from Germany: assessment of several different types of stressors with macrozoobenthos communities. Hydrobiologia 516: 215–228.

Böhmer, J., C. Rawer-Jost, A. Zenker, C. Meier, C. K. Feld, R. Biss & D. Hering, 2004b. Assessing streams in Germany with benthic invertebrates: Development of a multimetric invertebrate based assessment system. Limnologica 34(4): 416–432.

Buffagni, A., S. Erba, M. Cazzola & J. L. Kemp, 2004. The AQEM multimetric system for the southern Italian Apennines: assessing the impact of water quality and habitat degradation on pool macroinvertebrates in Mediterranean rivers. Hydrobiologia 516: 315–331.

Coesel, P. F. M. 2001. A method for quantifying conservation value in lentic freshwater habitats using desmids as indicator organisms. Biodiversity and Conservation 10(2): 177–187.

Davis, W. S. & T. P. Simon, 1995. Biological Assessment and Criteria. Tools for Water Resource Planning and Decision Making. Lewis Publishers, Boca Raton, London, Tokyo.

Ellenberg, H., H. E. Weber, R. Düll, V. Wirth, W. Werner & D. Paulißen, 1992. Zeigerwerte von Pflanzen in Mitteleuropa, Scripta Geobotanica 18.

Eurolimpacs, 2004. Integrated project to evaluate the impacts of global change on European freshwater ecosystems. Macro-invertebrate Taxa and autecology database. www.freshwaterecology.info.

Fame Consortium, 2004. Manual for the application of the European Fish Index – EFI. A fish-based method to assess the ecological status of European rivers in support of the Water Framework Directive. Version 1.1, January 2005.

Feld, C. K., 2004. Identification and measure of hydromorphological degradation in Central European lowland streams. Hydrobiologia 516: 69–90.

Feld, C. K., 2005. Assessing hydromorphological degradation of sand-bottom lowland rivers in Central Europe using benthic macroinvertebrates. Dissertation, University of Duisburg-Essen, Essen, 127 pp + App.

Furse, M., D. Hering, O. Moog, P. Verdonschot, R. K. Johnson, K. Brabec, K. Gritzalis, A. Buffagni, P. Pinto, N. Friberg, J. Murray-Bligh, J. Kokes, R. Alber, P. Usseglio-Polatera, P. Haase, R. Sweeting, B. Bis, K. Szoszkiewicz, H. Soszka, G. Springe, F. Sporka & I. Krno, 2006. The STAR project: context, objectives and approaches. Hydrobiologia 566: 3–29.

Henrikson, L. & M. Medin, 1986. Biologisk bedömning av försurningspåverkan på Lelångens tillflöden och grundområden 1986. Aquaekologerna, Rapport till länsstyrelsen i Älvsborgs län.

Hering, D., J. Böhmer, C. Meier, R. Biss, C. Rawer-Jost, C. K. Feld & A. Zenker, 2004a. Development of a multimetric invertebrate based assessment system for German rivers. Limnologica 34(4): 398–415.

Hering, D., O. Moog, L. Sandin & P. F. M. Verdonschot, 2004b. Overview and application of the AQEM assessment system. Hydrobiologia 516: 1–20.

Hering, D., R. K. Johnson, S. Kramm, S. Schmutz, K. Szoszkiewicz & P. F. M. Verdonschot, submitted. Assessment of European rivers with diatoms, macrophytes, invertebrates and fish: A comparative metric-based analysis of organism response to stress.

Heiskanen, A. S., W. van de Bund, A. C. Cardoso & P. Noges, 2004. Towards good ecological status of surface waters in

Europe – interpretation and harmonisation of the concept. Water Science and Technology 49: 169–177.

Holland, A. F. (ed), 1990. Near Coastal Program Plan for 1990–Estuaries. U.S. Environmental Protection Agency Environmental Research Laboratory, Office of Research and Development, Washington, DC 259 pp, EPA-600/9-90-033.

Holmes, N. T. H., J. R. Newman, S. Chadd, K. J. Rouen, L. Saint & F. H. Dawson, 1999. Mean Trophic Rank: A Users Manual. R&D Technical Report No. E38. Environment Agency, Bristol, UK.

Hughes, R. M., P. R. Kaufmann, A. T. Herlihy, T. M. Kincaid, L. Reynolds & D. P. Larsen, 1998. A process for developing and evaluating indices of fish assemblage integrity. Canadian Journal of Fisheries and Aquatic Sciences 55: 1618–1631.

Hughes, R. M. & T. Oberdorff, 1999. Applications of IBI concepts and metrics to waters outside the United States. In T. P. Simon (ed.), Assessing the Sustainability and Biological Integrity of Water Resource Quality Using Fish Communities. CRC Press, Boca Raton, Florida: 79–96.

Johnson, R. K., D. Hering, M. T. Furse & R. T. Clarke, 2006. Detection of ecological change using multiple organism groups: metrics and uncertainty. Hydrobiologia 566: 115–137.

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

Karr, J. R. & E. W. Chu, 1999. Restoring Life in Running Waters: Better Biological Monitoring. Island Press, Washington, DC 200 pp.

Karr, J. R. & B. L. Kerans, 1992. Components of biological integrity – Their definition and use in development of an invertebrate IBI, in Midwest Pollution Control Biologists Meeting, Chicago, Ill., 1991, Proceedings: U.S. Environmental Protection Agency, Region V, EPA-905/R-92-003, pp. 1–16.

Kelly, M. G. & B. A. Whitton, 1995. The Trophic Diatom Index: a new index for monitoring eutrophication in rivers. Journal of Applied Phycology 7: 433–444.

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

Lorenz, A., D. Hering, C. Feld & P. Rolauffs, 2004. A new method for assessing the impact of hydromorphological degradation on the macroinvertebrate fauna of five German stream types. Hydrobiologia 516: 107–127.

Merritt, R. W. & K. W. Cummins, 1996. An Introduction to the Aquatic Insects of North America. Dubuque, Kendall/Hunt 862 pp.

Mischke, U. & H. Behrendt, 2005. Vorschlag zur Bewertung ausgewählter Fließgewässertypen anhand des Phytoplanktons. In C. K. Feld, S. Rödiger, M. Sommerhäuser & G. Friedrich (eds), Typologie, Bewertung, Management von Oberflächengewässern. E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart: 46–62.

Moog, O. (ed), 1995. Fauna Aquatica Austriaca. Bundesministerium für Land- und Forstwirtschaft. Wasserwirtschaftskataster Wien.

Moog, O., A. Schmidt-Kloiber, R. Vogl & V. Koller-Kreimel, 2001. ECOPROF-Software. Wasserwirtschaftskataster, Bundesministeriums für Land- & Forstwirtschaft. Umwelt & Wasserwirtschaft, Wien.

Ofenböck, T., O. Moog, J. Gerritsen & M. T. Barbour, 2004. A stressor specific multimetric approach for monitoring running waters in Austria using benthic macro-invertebrates. Hydrobiologia 516: 253–270.

Plafkin, J. L., M. T. Barbour, K. D. Porter, S. K. Gross & R. M. Hughes, 1989. Rapid Bioassessment Protocols for use in Streams and Rivers: Benthic Macroinvertebrates and Fish. U.S. Environmental Protection Agency, Office of Water Regulations and Standards, Washington, DCEPA 440-4-89-001.

Resh, V. B., D. M. Rosenberg & T. B. Reynoldson, 2000. Selection of benthic macroinvertebrate metrics for monitoring water quality of the Fraser River, British Columbia: implications for both multimetric approaches and multivariate models. In J. F. Wright, D. W. Sutcliffe & M. T. Furse(eds), RIVPACS and Similar Techniques for Assessing the Biological Quality of Freshwaters. Freshwater Biological Association and Environment Agency, UK, Ableside, Cumbria, UK.

Rott, E., P. Pfister, H. Dam, E. Pipp, K. Pall, N. Binder & K. Ortler, 1999. Indikationslisten für Aufwuchsalgen. Bundesministerium für Land – und Forstwirtschaft, Wien 248 pp.

Schaumburg, J., C. Schranz, J. Foerster, A. Gutowski, G. Hofmann, P. Meilinger, S. Schneider & U. Schmedtje, 2004. Ecological classification of macrophytes and phytobenthos for rivers in Germany according to the Water Framework Directive. Limnologica 34: 283–301.

Schmedtje, U. & M. Colling, 1996. Ökologische Typisierung der aquatischen Makrofauna. Informationsberichte des Bayerischen Landesamtes für Wasserwirtschaft 4/96.

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

Schmidt-Kloiber, A., W. Graf, A. Lorenz & O. Moog, 2006. The AQEM/STAR taxalist – a pan-European macro-invertebrate ecological database and taxa inventory. Hydrobiologia 566: 325–342.

Schweder, H., 1992. Neue Indices für die Bewertung des ökologischen Zustandes von Fließgewässern, abgeleitet aus der Makroinvertebraten-Ernährungstypologie. Limnologie Aktuell 3: 353–377.

Tachet, H., P. Richoux, M. Bournaud & P. Usseglio-Polatera, 2002. Invertébrés d'eau douce: Systematique, Biologie, Ecologie. CNRS Editiona, p. 587.

Vlek, H. E., P. F. M. Verdonschot & R. C. Nijboer, 2004. Towards a multimetric index for the assessment of Dutch streams using benthic macroinvertebrates. Hydrobiologia 516: 175–191.