Science of the Total Environment 543 (2016) 123–134

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Benthic macroinvertebrates in lake ecological assessment: A review of methods, intercalibration and practical recommendations

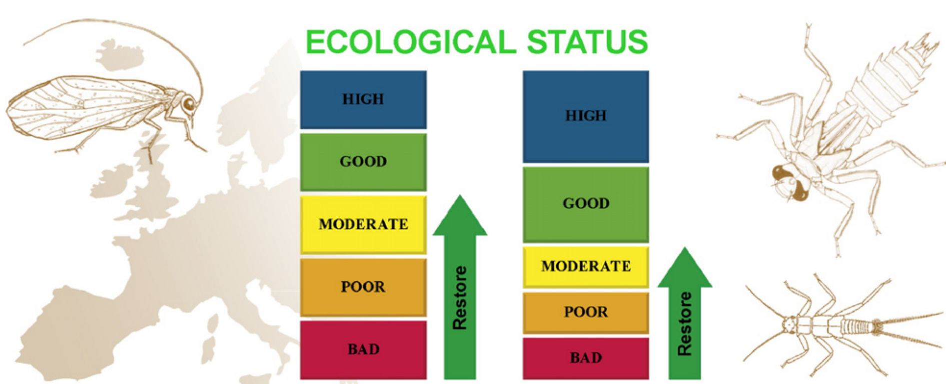
Sandra Poikane [a,*], Richard K. Johnson [b], Leonard Sandin [c], Ann Kristin Schartau [d], Angelo G. Solimini [e], Gorazd Urbanič [f], Kęstutis Arbačiauskas [g], Jukka Aroviita [h], Wim Gabriels [i], Oliver Miler [j], Martin T. Pusch [j], Henn Timm [k], Jürgen Böhmer [l]

[a] European Commission, Joint Research Centre, Institute for Environment and Sustainability, Water Resources Unit, 21027 Ispra, Italy
[b] Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Box 7050, SE-750 07 Uppsala, Sweden
[c] Institute of Freshwater Research, Department of Aquatic Resources, Swedish University of Agricultural Sciences, Stångholmsvägen 2, SE-178 93 Drottningholm, Sweden
[d] Norwegian Institute for Nature Research, Gaustadalléen 21, 0349 Oslo, Norway
[e] Department of Public Health and Infectious Diseases, Sapienza University of Rome, 00185 Rome, Italy
[f] Institute for Water of the Republic of Slovenia, 1000 Ljubljana, Slovenia
[g] Nature Research Centre, Akademijos Str. 2, LT-08412 Vilnius, Lithuania
[h] Freshwater Centre, Finnish Environment Institute SYKE, 90014 Oulu, Finland
[i] Flemish Environment Agency, Department Water Monitoring, Dr. De Moorstraat 24-26, B-9300 Aalst, Belgium
[j] Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 301, 12587 Berlin, Germany
[k] Centre for Limnology, Estonian University of Life Sciences, 61117 Rannu, Tartumaa, Estonia
[l] Bioforum GmbH, Sudetenstr. 34, 73230 Kirchheim/Teck, Germany

HIGHLIGHTS

- Ecological status of European surface waters is assessed using biological communities.
- We reviewed and intercalibrated 13 lake benthic invertebrate-based tools across Europe.
- These tools address acidification, eutrophication and morphological alterations.
- Two biological multimetric indices were developed for two large regions of Europe.
- We provide recommendations for the use of benthic invertebrates in lake assessment.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 17 September 2015
Received in revised form 4 November 2015
Accepted 4 November 2015
Available online 12 November 2015

Editor: D. Barcelo

ABSTRACT

Legislation in Europe has been adopted to determine and improve the ecological integrity of inland and coastal waters. Assessment is based on four biotic groups, including benthic macroinvertebrate communities. For lakes, benthic invertebrates have been recognized as one of the most difficult organism groups to use in ecological assessment, and hitherto their use in ecological assessment has been limited. In this study, we review and intercalibrate 13 benthic invertebrate-based tools across Europe. These assessment tools address different human impacts: acidification (3 methods), eutrophication (3 methods), morphological alterations (2 methods), and a combination of the last two (5 methods). For intercalibration, the methods were grouped into four

* Corresponding author.
E-mail address: sandra.poikane@jrc.ec.europa.eu (S. Poikane).

http://dx.doi.org/10.1016/j.scitotenv.2015.11.021

Keywords:
Biological metrics
Benthic invertebrates
Ecological assessment
Lakes
Water Framework Directive
Pressure–response relationships

intercalibration groups, according to the habitat sampled and putative pressure. Boundaries of the 'good ecological status' were compared and harmonized using direct or indirect comparison approaches. To enable indirect comparison of the methods, three common pressure indices and two common biological multimetric indices were developed for larger geographical areas. Additionally, we identified the best-performing methods based on their responsiveness to different human impacts. Based on these experiences, we provide practical recommendations for the development and harmonization of benthic invertebrate assessment methods in lakes and similar habitats.

1. Introduction

In recent years, much legislation has been developed in order to assess the ecological integrity of fresh waters worldwide (e.g. Clean Water Act in the USA, National Water Act in South-Africa, and Water Framework Directive in Europe). Furthermore, there is also growing interest in shifting the focus from assessment methods based on water chemistry and simple biotic metrics (e.g. saprobic index) towards more robust assessment methods based on indicators of degradation of ecological structure and function (Bonada et al., 2006; Karr, 1999; Stoddard et al., 2008). In Europe, since the adoption of the European Water Framework Directive (WFD) in 2000 (EC, 2000), much progress has been made regarding the ecological assessment of inland and coastal waters (Birk et al., 2012; Reyjol et al., 2014). A key concept of the European WFD is that a suite of biological assemblages is used to assess the ecological quality of surface waters. For lakes, assessment approaches based on phytoplankton, macrophytes and phytobenthos, benthic invertebrates, and fish fauna need to be implemented. Biological assessments, expressed as Ecological Quality Ratios (EQRs) – defined as the observed state/expected state – are divided into five status classes (high, good, moderate, poor and bad). For developing a programme of measures, the most important distinction is between good and moderate status (Birk et al., 2012) because, when the quality status is less than good, countries must take action to improve a water body until good status is achieved (Birk et al., 2013). Thus, the development of reliable assessment tools and the setting of ecological class boundaries have become two of the most critical and difficult tasks in implementing the WFD, with work still ongoing for several taxonomic groups (Birk et al., 2012; Brucet et al., 2013; Poikane et al., 2015).

Among the many taxonomic groups used in biomonitoring, from microbes to large metazoans such as fish and birds, macroinvertebrates are one of the most commonly used groups (Birk et al., 2012; Johnson et al., 1993; Resh and Jackson, 1993), fulfilling many of the criteria characterizing the ideal biomonitoring tool (Bonada et al., 2006). However, most studies advocating the use of macroinvertebrates in biomonitoring have focused on stream habitats (Hering et al., 2004; Resh and Jackson, 1993), with fewer studies addressing the efficacy of using lake macroinvertebrate assemblages (Brauns et al., 2007; Johnson et al., 2004, 2007). Indeed, a decade ago, the paucity of WFD-compliant macroinvertebrate assessment tools was identified as one of the major gaps impeding full assessment of the ecological quality of lakes (Solimini et al., 2006). Since then, stimulated by the WFD implementation, a multitude of biological metrics has been developed to assess the ecological quality of lakes (Brucet et al., 2013).

The main pressures affecting the integrity of lakes are eutrophication, acidification, and alterations of hydrology and geomorphology (cf. Young et al., 2005). Building on early assessment approaches (Wiederholm, 1980; Henrikson and Medin, 1986), several WFD compliant assessment metrics based on profundal (Jyväsjärvi et al., 2010, 2012) and littoral (Johnson et al., 2007; McFarland et al., 2010; Schartau et al., 2008) invertebrate communities have been developed to assess eutrophication and acidification. By contrast, quantifying the effects of hydromorphological alterations on littoral macroinvertebrates have only recently been developed (Brauns et al., 2007; Miler et al., 2015) and used for quantifying human-induced effects (Urbanič, 2014).

A basic requirement for successful river basin management is comparability of bioassessment approaches, as different data and indices can lead to inconsistent or conflicting assignment of ecological status (Birk et al., 2013; Cao and Hawkins, 2011). In Europe, legislation stipulates that values of the upper and lower "good" class boundaries must be harmonized (intercalibrated) to ensure that class boundaries are consistent with the normative definitions of the WFD and comparable between countries (Birk et al., 2013; Poikane et al., 2014b). For methods used in monitoring benthic invertebrate assemblages in lakes this task is particularly difficult. One reason is the diversity of methods currently used for addressing different pressures or combinations of pressures, often using different sampling methodologies and habitats (profundal, sublittoral or littoral). Another reason is that – compared to the use of phytoplankton in lakes and macroinvertebrates in streams – the use of benthic macroinvertebrates in lakes is relatively new, with the exception of profundal macroinvertebrates (Wiederholm, 1980). Furthermore, the large biogeographical range of EU countries results in high natural variability (lake/habitat types) and different types of impairment that need consideration. For example, densely populated central European countries, such as the Netherlands or Belgium, are comprised of mostly degraded water bodies (Gabriels et al., 2010), whereas lakes in the northern and eastern parts of the European Union, e.g. in Estonia, are often still in quite a natural state (Timm and Möls, 2012).

This paper describes the intercalibration exercise on benthic macroinvertebrate methods for assessing the ecological status of European lakes. The specific aims of this study are to:

- review the current status of macroinvertebrate methodologies proposed for European lakes, with particular attention to the metrics included and human impacts addressed;
- compare the lake assessment methods proposed by several countries and achieve a harmonization of class boundaries; and
- provide recommendations for the use of benthic invertebrates in the bioassessment of lakes.

2. Materials and methods

2.1. Assessment systems

Seventeen methods from 12 countries were considered as part of the intercalibration exercise: UK, Sweden and Germany each participated with several methods (addressing different pressures, different habitats or different lake types). From these methods, 13 methods from 10 countries were intercalibrated (see Table 1), while four methods — the German AESHNA sublittoral method (Miler et al., 2013b), the French macroinvertebrate index (Böhmer et al., 2014), the Italian BQI (Rossaro et al., 2007), and the Swedish ASPT (Johnson and Goedkoop, 2007) were excluded (see chapter on feasibility check).

Most of the methods ($n = 9$) were multimetric indices, while some (the Finnish and Swedish BQI, the UK CPET and LAMM) were single-metric methods. Metrics were grouped into four categories (sensitivity; richness/diversity; functional and taxonomic composition) based on classifications proposed by Hering et al. (2006); Stoddard et al. (2008) and Birk et al. (2012). Response of the methods to relevant pressures

Table 1
Overview of lake benthic invertebrate assessment methods developed by various member states participating in the intercalibration exercise (only intercalibrated methods).

Member state	Method	Acronym used further in the text	Habitat, pressure	Reference
Belgium	Multimetric Macroinvertebrate Index Flanders (MMIF)	BE	Eulittoral, eutrophication and morphological pressures	Gabriels et al. (2010)
Germany	German Macroinvertebrate Lake Assessment (AESHNA) for lowland lakes	DE-CB	Eulittoral, eutrophication and morphological pressures	Miler et al. (2013b)
Germany	German Macroinvertebrate Lake Assessment (AESHNA) for Alpine lakes	DE-ALP	Eulittoral, morphological pressures	Miler et al. (2013b)
Estonia	Estimation of freshwater quality using macroinvertebrates	EE	Eulittoral, eutrophication and morphological pressures	Timm and Möls (2012)
Finland	Benthic Quality Index (BQI)	FI-BQI	Profundal, eutrophication	Wiederholm (1980), Jyväsjärvi et al. (2010)
Lithuania	Lithuanian Lake Macroinvertebrate Index (LLMI)	LT	Eulittoral, eutrophication and morphological pressures	Šidagytė et al. (2013)
Netherlands	WFD — Metrics for Natural Water Types	NL	Eulittoral, eutrophication and morphological pressures	Böhmer et al. (2014)
Norway	Multimetric assessment method for acidification of clear lakes (MultiClear)	NO	Eulittoral, acidification	Sandin et al. (2014)
Sweden	Multimetric Index for Lake Acidity (MILA)	SE-MILA	Eulittoral, acidification	Johnson and Goedkoop (2007)
Sweden	Benthic Quality Index (BQI)	SE-BQI	Profundal, eutrophication	Wiederholm (1980), Johnson and Goedkoop (2007)
Slovenia	Slovenian Lake littoral benthic invertebrate index (LBI)	SI	Eulittoral, morphological pressures	Urbanič (2014)
United Kingdom	Chironomid Pupal Exuviae Technique (CPET)	UK-CPET	Whole lake, eutrophication	Ruse (2010)
United Kingdom	Lake Acidification Macroinvertebrate Metric (LAMM)	UK-LAMM	Eulittoral, acidification	McFarland et al. (2010)

was tested and evaluated using the coefficient of determination (R^2) and significance of linear regressions.

2.2. Intercalibration methodology

The intercalibration procedure involved five steps: (1) feasibility check; (2) data collection and choosing the appropriate IC option; (3) development of common metrics; (4) benchmark standardization and (5) method comparison and harmonization (for details see Birk et al., 2013; EC, 2011). A flowchart is provided in Fig. 1.

2.2.1. Feasibility check

An intercalibration feasibility check was performed with the aim to restrict the actual intercalibration analysis to methods that address the same common type(s) and anthropogenic pressure(s), and following a similar assessment concept. In this step, we grouped methods into intercalibration groups according to pressure type(s), habitat and geographical region. For example, the use of samples taken from profundal habitats to assess lake eutrophication, or littoral samples to assess acidification.

2.2.2. Data collection and choosing the appropriate IC option

Thirteen countries provided data from national monitoring or ongoing activities focused on developing the WFD compliant monitoring methods. Using a typology approach to reduce natural biological variation (cf. Poikane et al., 2010), data were collated for common lake types in each region (for type descriptions see Table S1 in Supplementary information). Partitioning natural variability by lake type and region resulted in: 214 samples from 19 lakes in the Alpine region, 931 samples from 216 lakes in the Central-Baltic region and 450 samples from 326 lakes in the Northern region (Table S2 in Supplementary information presents a more thorough description of datasets). Benthic macroinvertebrate samples were collected from the littoral zones of lakes using a hand net, while profundal samples were collected using an Ekman sampler (for more detailed information on field sampling and laboratory processing see Table S3 in Supplementary information).

Two intercalibration approaches were applied: (i) Direct comparison: when countries within the intercalibration group use similar field and laboratory protocols, national assessment methods were applied to the other countries' datasets and the average EQR value was calculated for each site. For example, Swedish assessment metrics were calculated using data taken from Swedish, Norwegian and UK sites. Afterwards, the Swedish assessment was compared with the average from other assessment systems (for more details, see Birk et al., 2013); (ii) Indirect comparison: when countries use different field and laboratory protocols, the national assessment metrics were converted into a comparable format of independent common metrics, and the national metrics were compared using these common metrics (e.g., Birk and Willby, 2010; Buffagni et al., 2007).

2.2.3. Development of pressure indices and biological common metrics

2.2.3.1. Development of common pressure indices. The aim of pressure indices was to synthesize information on relevant pressures into a single index value. We considered a set of eight stressor metrics: two metrics of land use at site level, two metrics of land use in near lake surroundings, one metric of land use in catchment, two metrics of shore alteration and total phosphorus concentration (Table 2). All pressure variables were standardized from 1 to 5 (continuous values). Individual metrics and combinations were tested for correlation with national methods. The multimetric which correlated best was chosen as the final stressor metric.

2.2.3.2. Development of biological common metrics. An Intercalibration Common Metric (ICM) is a biological metric widely applicable within a region or across regions which is used to convert national boundaries, via linear regression, to a common scale (Buffagni et al., 2007). ICMs were developed using biological data for comparing assessment methods used in the Alpine and Central-Baltic regions.

Using the Asterics software (version 3.1), 120 biological indices were calculated from species by site matrices. Many indices were excluded from further analyses as they were deemed to be numerically unsuitable, e.g. metrics having a narrow range of values or having many outliers and extreme values (Hering et al., 2006; Stoddard et al., 2008). More details on these 120 metrics can be found in Table S6 in Supplementary information. Subsequently, 71 metrics were correlated with selected anthropogenic pressures: morphological alterations, eutrophication and the combination of these two pressures (both for the whole dataset as well as for each country separately). To ensure a successful intercalibration, the metrics had to be well correlated with both the national assessment systems of all countries (i.e. with the national multimetric indices, normalized as EQR values (EQRs = Ecological

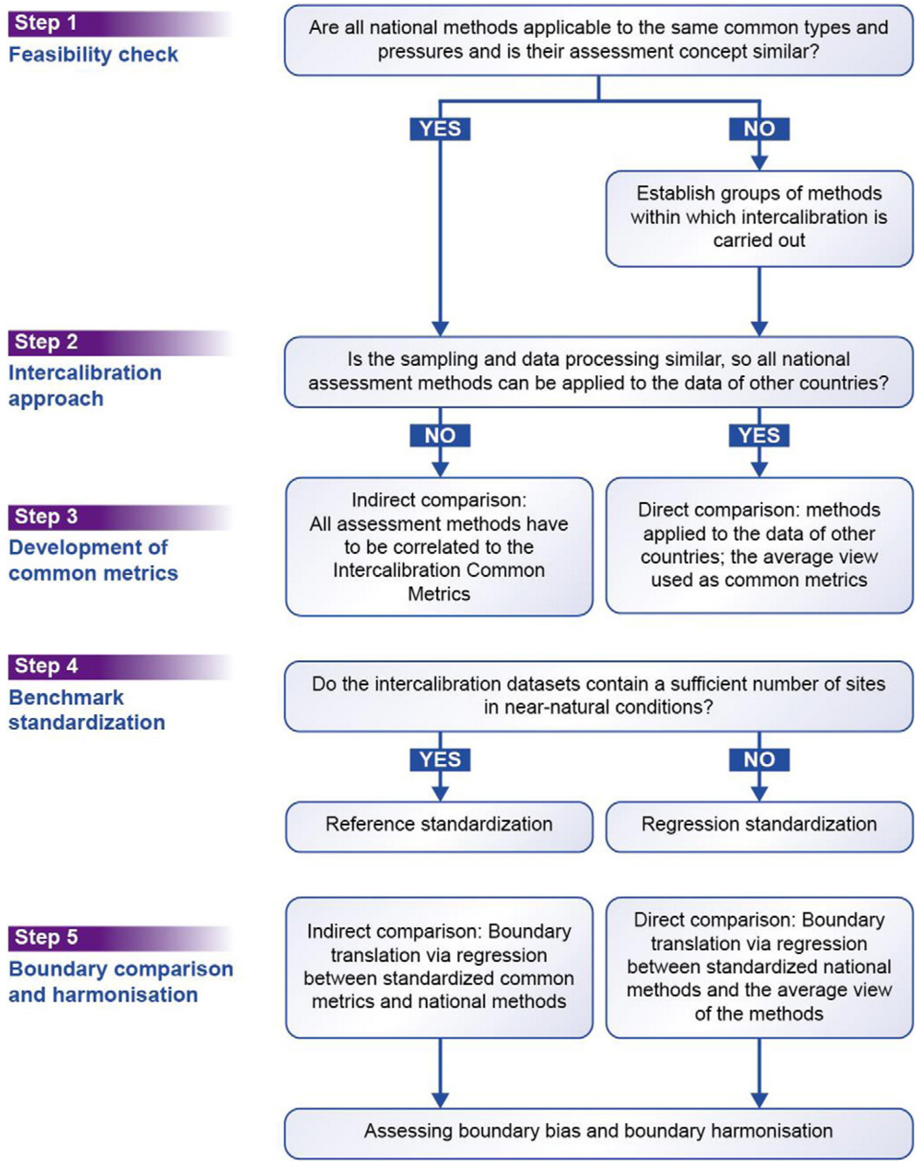


Fig. 1. Flowchart of the intercalibration process.

Quality Ratios from 0 to 1) and the selected pressures. Criteria for the selection of candidate metrics were, in descending order: (1) overall correlation strength with the national EQR values, (2) correlation strength with the national EQRs for each country separately, (3) overall correlation strength with the pressure variables, and (4) correlation strength with the pressure variables for each country separately. To judge the strength of these correlations Spearman's rank and Pearson's product–moment correlation coefficients were calculated between biological metrics and pressure metrics.

Based on the strength of these correlations, eight metrics were selected as candidates for calibrating multimetric indices for each of the two regions. Candidate metrics were normalized to a value between 0 and 1 (Ecological Quality Ratio) following a procedure described by Hering et al. (2006) and different multimetric combinations were correlated with the national methods and pressure variables (see Tables S7 and S8). These variants contained three to six metrics, with at least one metric belonging to each metric category (sensitivity/tolerance, taxonomic composition and functional groups, diversity). Also correlation among metrics was considered — the metric was considered redundant if correlated ($r > 0.8$) with other metrics. The multimetric indices that correlated best, both with the national methods and the pressure variables, were selected as the final ICM.

2.2.4. Benchmark standardization

Due to differences in biogeography and typology, as well as to differences in data acquisition, caution is advised when comparing biological data across broad spatial scales (Cao and Hawkins, 2011). Consequently, metric values were standardized in order to reduce intrinsic biogeographical and/or methodological differences between participating countries at the start of intercalibration. Two different approaches, described by Birk et al. (2013), were used: (i) "reference standardization" based on near-natural reference sites and (ii) "regression standardization" using pressure-response gradients (for detailed descriptions see Birk et al., 2013; EC, 2011).

For Northern regions, where many lakes are still in near-natural conditions, "reference standardization" was used (i.e. reference criteria were used to select reference sites). Each country calculated its national EQR using datasets from the other countries of the Northern region (e.g. the Norwegian EQR was calculated for reference sites situated in Norway, Sweden and the UK). ANOVA was used to compare values of

Table 2
Stressor variables for the development of pressure indices in the Alpine (ALP) and Central Baltic (CB) region.

Stressor variable	Explanation	Included in the pressure index
LUS15	Land-use index within 15 m of sampling site: % of non-natural[a] land uses (mainly urban and agricultural areas) directly adjacent to the site (15 m belt at 100 m shore length)	ALP
LUS100	Land-use index within 100 m of sampling site % of non-natural[a] land uses (mainly urban and agricultural areas) directly adjacent to the site (100 m belt at 100 m shore length)	ALP
LUL15	Land-use index from the % of land uses in the 15 m belt around the whole lake (4 × [% artificial] + 1.5 × [% agriculture])	CB
LUL100	Land-use index from the % of land uses in the 100 m belt around the whole lake (1 × % extensive agriculture + 2 × % intensive agriculture + 4 × % urban areas)	ALP, CB
LUL-catchment	Land-use index from the % of land uses in the lake catchment (1 × % extensive agriculture + 2 × % intensive agriculture + 4 × % urban areas)	
Naturalness of shoreline	National naturalness classification by expert judgement, based on morphology and land use of the shoreline and adjacent areas at the sampling sites (5 classes)	ALP
Altered shoreline	% altered shore length of total shore length	ALP, CB
TP	Total phosphorus concentrations (annual mean)	CB

[a] All anthropogenically altered areas, except woodlands, successional areas (e.g. scrublands) and natural marshes.

reference sites among all countries within the group. Among-country differences were then removed (factored out) prior to the intercalibration analysis.

For the Alpine and Central-Baltic regions, the "regression standardization" approach was used to standardize the ICM. Linear Mixed Models, with biological metrics as dependent variables, the pressure index as covariables and country as random factor were used to calculate offset values. Regression calculations were performed using the package 'lme4' in R software (Team RC, 2012). Standardized ICM metric values were obtained by subtracting the offsets from the metric values.

2.2.5. Method comparison and harmonization

Three steps were used to harmonize national classifications: (i) relationships between the national methods and the ICM were established (to be considered further, national metrics had to be significantly correlated with the ICM with r-values >0.5 and slopes between 0.5 and 1.5), (ii) national boundaries of high/good and good/moderate classifications were scaled to the ICMs using regression and compared with the global mean view of all countries, and (iii) national classification systems were adjusted so as not to exceed the agreed upon deviation from the boundary, i.e. the most that any national boundary could deviate from the global mean of all countries was ±0.25 classes and therefore the most widely divergent national methods could not differ from each other by more than 0.5 classes (Birk et al., 2013).

3. Results

3.1. Assessment systems: metrics included

Thirteen macroinvertebrate assessment methods were intercalibrated comprising in total 44 metrics. Nine of the assessment methods are multimetric methods consisting of up to five metrics, whereas four methods consist of only one metric (see description of metrics in Table S4 in Supplementary information). Almost half (43%) of the 44 metrics belonged to sensitivity/tolerance metrics, and were included in all assessment methods. Some countries used traditional indices such as the ASPT index (Armitage et al., 1983) (Lithuania and Estonia), Benthic Quality Index (Wiederholm, 1980) (Sweden and Finland), and Acidity Index (Henrikson and Medin, 1986) (Norway and Estonia), whereas most countries developed new sensitivity indices such as the Fauna Index (Miler et al., 2013b), Littoral Fauna Index (Urbanič, 2014), Mean Tolerance Score (Gabriels et al., 2010), chironomid pupal exuvial technique (CPET) index (Ruse, 2010) and Lake Acidification Macroinvertebrate Metric (LAMM) (McFarland et al., 2010).

Most methods also included some measure of taxon richness and diversity (37% of all metrics), such as total taxon richness, Shannon-Wiener diversity, number of EPT taxa, number of Ephemeroptera taxa, or number of Gastropoda taxa. Only three methods included functional metrics (9%), and four included composition/abundance metrics (11%).

3.2. Pressure–response relationships

Three assessment methods were calibrated to assess acidification pressure, with strong relationships with pH (NO, SE-MILA, UK-LAMM: $R^2 = 0.37$ to 0.80) and acid neutralizing capacity (ANC) (NO, UK-LAMM: $R^2 = 0.47$ to 0.82) (for detailed information see Table S5 in Supplementary information).

Two methods (DE-ALP and SI) were developed to assess the effects of hydromorphological alterations on benthic invertebrate assemblages. Relationships were tested using the Lakeshore Modification Index (Peterlin and Urbanič, 2013, Slovenia, $R^2 = 0.80$) and Morpho-Index (Germany, $R^2 = 0.23$ to 0.45).

Four methods addressed both the effects of elevated nutrients and hydromorphological alterations. Some methods were tested against eutrophication variables (EE, LT: $R^2 = 0.10$ to 0.4848), some against morphological pressures (NL, DE-CB, LT: $R^2 = 0.11$ to 0.6161), and some assessed combinations of pressures (LT, DE-CB: R^2 for combined morphology and nutrients was slightly larger (0.22 and 0.31 respectively for LT and DE-CB) than for morphology alone (0.11 and 0.25). Finally, three methods addressed only the impacts of eutrophication. UK-CPET scores were related ($R^2 = 0.78$, $P < 0.001$) to a compound pressure metric (total nitrogen x total phosphorus/mean depth). The FI-BQI was significantly related to total phosphorus concentration (SE, FI: $R^2 = 0.27$–0.32, $P < 0.001$), with stronger relationships observed in deep lakes (mean depth > 6 m; Jyväsjärvi et al., 2012).

3.3. Intercalibration

3.3.1. Intercalibration groups and options

In total, four groups of methods were established according to the region, lake types, pressures and habitats (Table 3). In the Alpine region, assessment methods focused on the effects of hydromorphological alterations on eulittoral habitats, while in the Central-Baltic region, the effects of combined pressures on assemblages in eulittoral habitats were evaluated. For the Northern region, two groups were formed: one addressing the effects of eutrophication on profundal assemblages, and the other addressing the effects of acidification on littoral assemblages.

In the Alpine and Central-Baltic regions, methods differed in field sampling (sampling time, habitats sampled) and laboratory procedures (taxonomic resolution); consequently, an indirect comparison with independent common metrics was used. By contrast, assessment methods used in the Northern region were similar, allowing for direct comparisons between assessment methods (i.e. each national method was applied to datasets from the other countries and assessment results were compared).

3.3.2. Development of common pressure indices

Two pressure indices were constructed describing morphological alterations in the Alpine (Morpho-index$_{ALP}$) and Central-Baltic region (Morpho-index$_{CB}$). Additionally, an index comprising both morphological

Table 3
Overview of the lake intercalibration groups, pressures and habitats addressed.

Region	Pressure addressed	Habitat	Methods intercalibrated	Intercalibration option
Alpine	HYMO[a]	Eulittoral	DE-ALP, SI	Comparison via ICM_{ALP}
Central Baltic	HYMO and EUTR[b]	Eulittoral	BE, DE-CB, EE, LT, NL, UK-CPET	Comparison via ICM_{CB}
Northern	EUTR	Profundal	FI, SE-BQI	Direct comparison
Northern	ACID[c]	Eulittoral	NO, SE-MILA, UK-LAMM	Direct comparison

[a] HYMO — morphological alterations.
[b] EUTR — eutrophication.
[c] ACID — acidification.

alterations and eutrophication (Morpho-TP index) was calculated for the Central-Baltic region.

For Alpine region, a pressure index Morpho-index$_{ALP}$ for each sampling site was calculated using weighted averaging of standardized pressure variables (Table 2) as:

$$\text{Morpho-index}_{ALP} = (2 \times \text{naturalness of shoreline} + LUS_{15} + LUS_{100} + LUL_{100} + \%\text{altered shoreline})/6.$$

For the Central-Baltic region, the pressure index Morpho-index$_{CB}$ was calculated as:

$$\text{Morpho-index}_{CB} = (2 \times LUL_{15} + LUL_{100} + \%\text{altered shoreline})/4.$$

Combined pressure index Morpho-TP index was calculated for the Central-Baltic region based on the standardized values of Morpho-index$_{CB}$ and the annual mean concentration of total phosphorus (TP) as:

$$\text{Morpho-TP index} = (2 \times \text{Morpho-index}_{CB} + TP)/3.$$

3.3.3. Development of common biological metrics for intercalibration

Construction of intercalibration common metrics (ICMs) for the Alpine and Central-Baltic regions resulted in two multimetric ICMs. The ICM constructed for the Alpine region comprised four metrics: (i) Fauna index (FI), (ii) number of taxa (NoT), (iii) reproduction strategy (r-strategists/k-strategists), and (iv) % abundance classes of the feeding type collector-gatherers (% FG) calculated as:

$$ICM_{ALP} = 2FI + NoT + r/k + \%FG/5.$$

The ICM for the Central-Baltic region consisted of four metrics: (i) number of Ephemeroptera, Plecoptera, Trichoptera, Coleoptera, Bivalvia, Odonata taxa (EPTCBO), (ii) ASPT index, (iii) % abundance classes of Ephemeroptera, Trichoptera, Odonata taxa (% ETO), and (iv) % abundance classes with a preference for the lithal microhabitat (% HL) calculated as:

$$ICM_{CB} = (2 * EPTCBO + ASPT + \%ETO + \%HL)/5.$$

Both ICMs correlated significantly with most of the pressure variables (Table 4). The strongest relationships between ICMs and pressure variables were generally found in the Central-Baltic region ($r = -0.47$ to -0.62). The Morpho-TP Index showed the strongest correlation ($r = -0.62$) compared to morphology ($r = -0.57$) and TP ($r = -0.47$) alone. This is comparable to other studies which have reported stronger correlations for common metrics based on phytoplankton ($r = 0.39$–0.79), macrophytes (0.56–0.74) and phytobenthos (0.57–0.75), but lower for fish (0.36–0.68) and benthic invertebrates (0.33–0.69) (Carvalho et al., 2013; Kelly et al., 2014; Lyche Solheim et al., 2013).

Table 4
Correlations between Intercalibration Common Metrics (ICMs) and pressure variables (for explanations see Table 2 and Table 8).

Pressure variables	Alpine ICM		Central Baltic ICM	
	Pearson's r	P	Pearson's r	P
Naturalness of site	−0.49	<0.001		
Morpho index	−0.42	<0.001	−0.57	<0.001
Morpho-TP index			−0.62	<0.001
Total phosphorus			−0.47	<0.001

3.3.4. Benchmark standardization

In the Northern region, 78 near-natural reference lakes assessing lake eutrophication based on profundal macroinvertebrates were selected using a priori reference criteria. Analysis of profundal macroinvertebrate assemblages at reference sites showed no differences when the SE-BQI was tested between SE and FI reference sites (*t*-test, $P > 0.05$), whereas the FI-BQI differed between SE and FI reference conditions (*t*-test, $P < 0.0005$). Consequently, standardization was used in the analysis of the FI-BQI (i.e. the EQRs were divided by the corresponding median EQR at benchmark sites).

For assessing lake acidification based on littoral assemblages in the Northern region, 26 reference sites were selected according to reference criteria. In these lakes pH and/or ANC (Acid Neutralizing Capacity) shows only non-significant deviation from calculated site-specific or type-specific pre-acidification levels using MAGIC modelling (Sandin et al., 2014). We compared variability among reference sites in SE, the UK, and NO using three metrics. Neither the Swedish MILA metric nor the Norwegian MultiClear metric differed when reference sites from different countries were compared (t-test, $P > 0.05$). However for the UK-LAMM metric, values for the UK were higher than SE and NO reference data (t-test, $P < 0.005$). Therefore, we used benchmark standardization to normalize UK-LAMM values.

In the Central-Baltic and Alpine regions, sufficient data from reference sites were not available. Therefore, regression standardization (linear mixed models) was used to standardize all single metrics within the ICM. To obtain the standardized ICM metrics the offsets given by the model were subtracted from the metric values. After combination of standardized single metrics into a common multimetric, all countries followed the common pressure response model.

3.3.5. Comparison of national metrics and ICMs

For all three regions, relationships between country metrics and ICMs were highly significant (Table 5), with slopes within the interval of 0.5 to 1.5. For the two countries in the Alpine region, DE and SI (see Fig. 2), metrics were strongly related to the ICM (DE, $r = 0.76$, $P < 0.001$; SI, $r = 0.94$, $P < 0.001$). For lakes of the Central-Baltic region, correlations were higher for countries with broad environmental gradients (e.g. NL and DE, r-values of 0.70 and 0.63, respectively) than countries with relatively short gradients (e.g. LT, $r = 0.36$, $P = 0.007$). Correlation between the UK-CPET metric and ICM was higher when ICM values were aggregated by lake ($r = 0.66$, $P < 0.001$) as only one CPET assessment value was available for each lake compared to many site-specific ICM values per lake. The correlation between FI-BQI and SE-BQI metrics for addressing eutrophication pressures was highly significant ($r = 0.68$, $P < 0.001$).

3.3.6. Harmonization of class boundaries

Analysis of national boundaries for all three regions showed relatively good agreement with global harmonization boundaries. For Alpine

Table 5
Results of regression analysis between national assessment methods and intercalibration common metrics (ICM).

National method	Pearson's r	Slope	P	Intercalibration approach and ICM
Alpine region				
DE	0.76	0.98	<0.001	Indirect comparison via ICM_{ALP}: weighted average of Fauna index, taxa richness, reproduction strategy (r/k), % feeding type collector-gatherers
SI	0.94	1.23	<0.001	
Central Baltic region				
BE	0.56	0.99	<0.001	Indirect comparison via ICM_{CB}: weighted average of normalized values of number of EPTCBO taxa, ASPT, % ETO, % habitat preference lithal
DE-CB	0.63	0.62	<0.001	
EE	0.63	0.96	0.009	
LT	0.36	0.69	0.007	
NL	0.70	1.39	<0.001	
UK-CPET	0.66	1.09	<0.001	
Northern region — acidification				
SE-MILA	0.45	0.53	<0.001	Direct comparison (the average value of all methods used for comparison)
UK-LAMM	0.66	0.66	<0.001	
NO	0.76	0.44	<0.001	
Northern region — eutrophication				
FI BQI–SE BQI	0.68	0.70	<0.001	Direct comparison (regression of two methods)

and Northern region's acidification metrics no boundary adjustments were necessary (<0.25 class difference). For Northern region eutrophication metrics, the Good/Moderate boundary value for the FI BQI was increased from 0.60 to 0.63, while the High/Good boundary value for the SE BQI was decreased from 0.90 to 0.84 and the Good/Moderate boundary from 0.7 to 0.67. In the Central-Baltic region, national boundaries from three assessment methods (BE, EE, LT) deviated by more than 0.25 class equivalents (Fig. 3). The Belgian metric MMIF was not sufficiently stringent (deviation of −1.32 class equivalents), while the Estonian metric was deemed to be too stringent (+0.78). The Belgian metric MMIF was adjusted by revising the reference values, after which MMIF deviated by −0.125 from the global Good/Moderate boundary and by −0.033 from the High/Good boundary. Two countries with stringent class boundaries (LT, EE) lowered the values for the High/Good boundary to slightly above the global harmonization band. Final intercalibration results are given in Table 6.

4. Discussion

Macroinvertebrates have traditionally been recognized as one of the most difficult biological groups for use in lake ecological assessment due to several reasons, such as their complex biotic structure, relatively high temporal variability and the high spatial heterogeneity (Brose et al., 2004; Solimini and Sandin, 2012; White and Irvine, 2003). Accordingly, the use of macroinvertebrate communities in lake assessment programmes has been relatively limited (Solimini et al., 2006). However, in this study we reviewed and intercalibrated 13 benthic invertebrate assessment tools across Europe and summarized findings that may be of use when considering using benthic invertebrates in lake assessment in other countries.

4.1. Assessment tools: metrics included

There is a broad consensus that multimetric indices have to contain at least one metric from each metric type (e.g. richness/diversity, sensitivity/tolerance, composition and functional metrics) in order to reflect the complexity of biological communities (Hering et al., 2006; Karr, 1999; Stoddard et al., 2008). According to the EU WFD, macroinvertebrate-based assessment methods are required to reflect changes in diversity, in the ratio of disturbance sensitive to insensitive taxa, and in the abundance and taxonomic composition of benthic communities in rivers, lakes, transitional and coastal waters (EC, 2000). Nevertheless, four out of 13 assessment methods studied here consisted of single indices. Metrics of sensitivity/tolerance (43%) and richness/diversity (37%) were the most widely used, while measures of taxonomical composition and function (the latter optional according to the WFD) were included in only a few assessment systems. Furthermore, abundance was not used in any assessment method (except relative abundance). To be included, metrics should be responsive to anthropogenic pressures, have low natural variability and be ecologically meaningful and interpretable (Hering et al., 2006). Since not all macroinvertebrate metrics correspond equally well to these criteria, those that did not were excluded from the assessment method development.

Sensitivity metrics are widely used in bioassessment methods as they respond predictably to different environmental gradients (Johnson, 1998). In several cases traditional indices (e.g. ASPT index) were used in national monitoring programmes. However, in conjunction with the implementation of the WFD, new indices were developed indicating acidification (McFarland et al., 2010), eutrophication (Ruse, 2010), and lakeshore modification (Miler et al., 2013a; Urbanič, 2014). Metrics of richness and diversity are also frequently used based on the well documented loss of richness and diversity to human-generated disturbances (McFarland et al., 2010; Šidagytė et al., 2013). Nevertheless, richness was not included in all assessment approaches (e.g. UK-CPET, UK-LAMM, SE-BQI, and FI-BQI). Likely, one of the reasons for not including taxon richness is the unimodal relationship often found between richness and trophic gradients (Dodson et al., 2000; Jeppesen et al., 2000; Mittelbach et al., 2001), indicating that intermediate disturbance enhances species richness (Townsend et al., 1997).

In contrast, absolute macroinvertebrate abundances were not used in any of the assessment systems, since this parameter is known to be highly variable in aquatic invertebrate communities (Johnson, 1998;

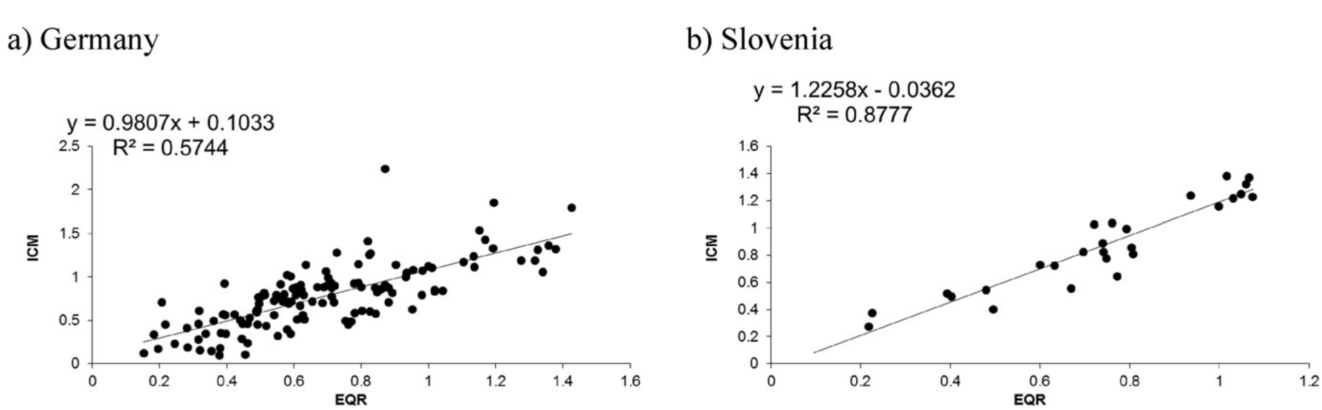


Fig. 2. Linear regressions between national benthic invertebrate lake assessment methods and the intercalibration common metric (ICM) in Alpine lakes: a) Germany, b) Slovenia. For further regressions see Fig. S1 in Supplementary data. ICM — Intercalibration Common Metric, EQR — Ecological Quality Ratio (final result of national assessment system).

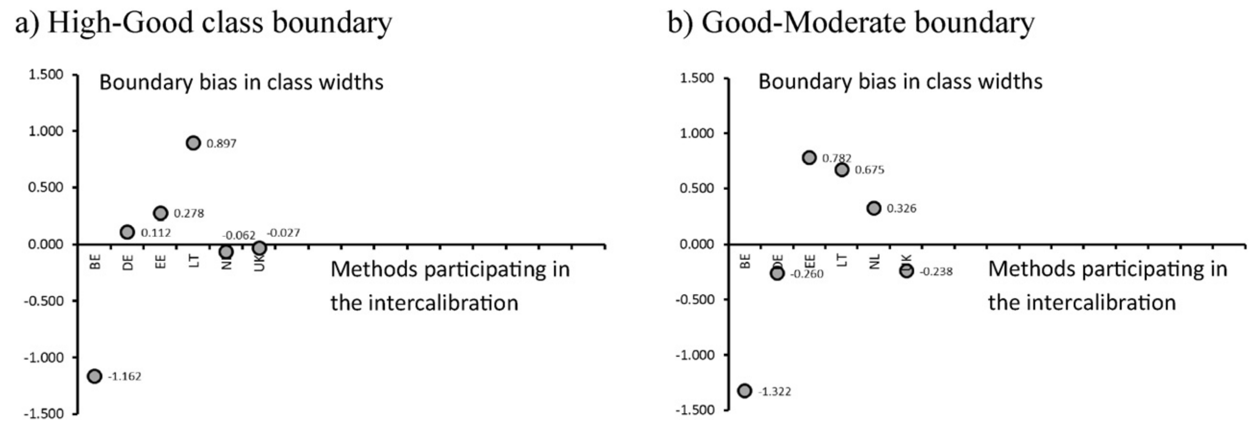


Fig. 3. Comparison of lake benthic invertebrate methods within the Central Baltic region. Bias of the boundaries of national methods participating in the intercalibration exercise is expressed in class width deviation from the mean view. All national boundaries should deviate less than ±0.25 classes from the mean view (zero bias). BE — Belgium, DE — Germany, EE — Estonia, LT — Lithuania, NL — The Netherlands, UK — United Kingdom. For other regions see Fig. S2 in Supplementary data.

Resh and Jackson, 1993). Osenberg et al. (1994) also argued that absolute abundances of invertebrates are rarely, if ever, used in ecological assessment due to the difficulties associated with detecting anthropogenic change with any degree of confidence. For example, Sandin and Johnson (2000) showed that invertebrate abundance was the least informative of 10 metrics tested, with the lowest effect size (a measure of the magnitude of impact) and the highest spatial, temporal and sample variability. Indeed, high spatial (due to habitat heterogeneity) and temporal (seasonal) variability are often two factors confounding estimates and use of invertebrate densities in bioassessment.

Functional metrics are widely used in stream (Böhmer et al., 2004; Hering et al., 2004) and coastal (Salas et al., 2006) assessments, although to a far lesser extent in lake assessment methods (but see Miler et al., 2013a; b). The main obstacles for using functional metrics can be summarized as: (1) lack of knowledge of biological traits of lake benthic invertebrates and how different functional groups/biological traits respond to different pressures (Solimini et al., 2006) and (2) incorrect assignment of taxa into functional groups (Karr, 1999; Trigal et al., 2009) due to omnivory, ontogeny, insufficient taxonomic identification, or lack of reliable ecological background information. Several studies have failed to show a relationship between functional metrics/groups of benthic invertebrate assemblages and anthropogenic pressures (Trigal et al., 2009; Urbanič et al., 2012). Hence further research is needed to determine the efficacy of using functional metrics in lake assessment.

4.2. Assessment methods: pressures addressed

Establishing reliable empirical relationships between anthropogenic impacts and biological responses is often a critical step in designing robust monitoring programmes (Dale and Beyeler, 2001; Hering et al., 2006; Karr, 1999). For benthic invertebrates in lakes, several studies have shown weak or no pressure-response relationships, especially for littoral invertebrates and eutrophication pressure (Bazzanti et al., 2012; O'Toole et al., 2008; Timm and Möls, 2012). Many studies show that natural factors, particularly lake area (Timm and Möls, 2012), alkalinity (O'Toole et al., 2008), depth (Brodersen et al., 1998), wind exposure (Brodersen, 1995) and, most important, habitat type (Brauns et al., 2007; Johnson and Goedkoop, 2002) may smother the effects of anthropogenic impact on local littoral benthic invertebrate assemblages.

However, our study of 13 benthic invertebrate assessment systems revealed significant relationships with acidification (3 methods), eutrophication (5), morphological alterations (5) and the combination of the

Table 6
Final high-good and good-moderate class boundary Ecological Quality Ratio (EQR) values for the national assessment methods included in the European Commission Intercalibration Decision (EC, 2013).

Region/member state	Ecological assessment method	Ecological Quality Ratios	
		High-good boundary	Good-moderate boundary
Alpine region			
Germany	German Macroinvertebrate Lake Assessment (AESHNA, part eulittoral of Alpine/Prealpine lakes)	0.80	0.60
Slovenia	Lake littoral benthic invertebrate index (LBI)	0.80	0.60
Central Baltic region			
Belgium	Multimetric Macroinvertebrate Index Flanders (MMIF)	0.90	0.70
Estonia	Estimation of freshwater quality using macroinvertebrates	0.86	0.70
Germany	German Macroinvertebrate Lake Assessment (AESHNA, part eulittoral of lowland lakes)	0.80	0.60
Lithuania	Lithuanian Lake Macroinvertebrate Index (LLMI)	0.74	0.50
Netherlands	WFD Metric for Natural Water types	0.80	0.60
United Kingdom	Chironomid Pupal Exuvial Technique (CPET)	0.77	0.64
Northern region — littoral acidification			
Norway	Multimetric Invertebrate Index for Clear Lakes (MultiClear)	0.95	0.74
Sweden	Multimetric Invertebrate Lake Acidification index (MILA)	0.85	0.60
United Kingdom	Lake Acidification Macroinvertebrate Metric (LAMM)	0.86	0.70
Northern region — profundal eutrophication			
FI	Benthic Quality Index (BQI)	0.75	0.63
SE	Benthic Quality Index (BQI)	0.84	0.67

last two pressures (2). Factors that were likely important in isolating pressure-response relationships were:

- Use of habitat-specific invertebrate assemblages to assess selected pressures, considering the vertical zonation of benthic invertebrates with lake depth. Profundal assemblages are strongly affected by eutrophication (oxygen deficiency) in many lake types, while littoral assemblages are better indicators of acidification and morphological pressures.
- Appropriate choice of pressure descriptors to build pressure-response relationships. This is relatively easy for certain pressures such as acidification (pH, ANC) and eutrophication (TP, trophic metrics), but difficult for other pressures such as morphological alterations. Here, pressure-specific indices, like the Lakeshore Modification Index developed for Slovenia (Peterlin and Urbanič, 2013) or the Morpho-Index developed for the Alpine and Central-Baltic regions (this paper) constitute fruitful approaches.
- Conceptual models of how multiple pressures, which may affect lake invertebrates, are useful when analysing pressure-response relationships. For example, eulittoral assemblages respond to both eutrophication and hydromorphological pressures (Brauns et al., 2007), and thus determining cause and effect can be difficult in densely populated areas like those of Central Europe where eutrophication is widespread and often co-occurs with other pressures. Therefore, a combined Morpho-TP index (this paper) was developed to aid in the analysis of pressure-response relationships for this pressure combination.
- Careful selection of assessment metrics. In theory, all metric types need to be included in the assessment methods (Hering et al., 2006; Karr, 1999). Our study showed, however, that in many cases only one or two metric types were included, as other metrics did not respond predictably across the pressure gradient. Sensitivity indices were the most reliable metric category, followed by richness and diversity metrics, while functional metrics were not included as their response was comparatively weaker (Schartau et al., 2008; Urbanič et al., 2012).
- Development of new metrics and assessment methods. In several cases, traditional indices such as EPT taxa richness or the AWIC index did not respond as predicted to the tested pressures (McFarland et al., 2010; Šidagytė et al., 2013). For morphological alterations, no methods were established at the start of the intercalibration exercise (Urbanič, 2014). Therefore, new metrics and methods were being developed (cf. McFarland et al., 2010; Šidagytė et al., 2013; Urbanič, 2014).

4.3. Intercalibration

If different assessment methods are used over a broad range of geographical conditions, they have to be harmonized to achieve comparable results (Birk et al., 2013; Cao and Hawkins, 2011). In Europe, legislation mandates the comparison and harmonization of assessment methods used by different countries, i.e. intercalibration (Poikane et al., 2014b). Several examples of intercalibration have been described for rivers: benthic invertebrates (Buffagni et al., 2007), diatoms (Kelly et al., 2009), macrophytes (Birk and Willby, 2010), for lakes: phytoplankton (Poikane et al., 2010, 2014a), macrophytes (Tóth et al., 2008), diatoms (Kelly et al., 2014), and for coastal areas: benthic invertebrates (Borja et al., 2007). These intercalibration exercises were confronted with a number of challenges: (i) differences in assessment concepts (Birk and Willby, 2010; Hering et al., 2004), (ii) the scarcity of reference sites and difficulties in defining comparable reference conditions (Birk and Hering, 2009) and (iii) large biogeographical and methodological differences among the countries (Kelly et al., 2014) which may render the comparison unreliable.

Despite these difficulties, our study demonstrates successful comparison and intercalibration of 13 benthic invertebrate methods across Europe. Many of the aforementioned difficulties were overcome by adopting the following procedures:

- Grouping the assessment methods into the relevant intercalibration groups according to the pressure addressed and habitat sampled (e.g. littoral acidification and profundal eutrophication groups);
- Choosing the appropriate intercalibration approach. Although direct comparison is the preferred option, as it allows for a straightforward comparison of methods, it can only be used when it is possible to apply each method to another country's data. This was the case in the Northern region.
- Development of common pressure and biological metrics. When national methods differed significantly, intercalibration common

Table 7
Selection of available methods addressing selected human pressures.

Region/pressure	Country/method	Pressure proxy	Variation explained	Pressure range	Remarks
Acidification					
Northern	UK — LAMM	ANC[a]	82% ($n = 106$, $P < 0.001$)	ANC 0–130 μeq L^{-1}	Littoral, stony substrates; more adapted to humic lakes (mean DOC >5 mg L^{-1})
Northern	Sweden — MILA	pH	70% ($n = 70$, $P < 0.0001$)	pH 4.5–7.5	Littoral, stony substrates; both for humic and clear lakes
Eutrophication					
Central-Baltic	UK — CPET	TP*TN[b]/mean depth	79% ($n = 166$, $P < 0.001$)	TP 0.002–0.99 mg L^{-1} TN 0.03–11.9 mg L^{-1}	Whole-lake assessment by collecting chironomid pupal exuviae
Central-Baltic	Lithuania — LLMI	TP	48% ($n = 66$, $P < 0.001$)	TP 0.005–0.056 mg L^{-1}	Littoral samples; stony substrates and submerged macrophytes samples
Northern	Finland — BQI	TP	26–32% ($n = 60$, $P < 0.001$)	TP 0.005–0.035 mg L^{-1}	Profundal samples; Better performance in deep lakes (mean depth > 6 m)
Morphological alterations					
Alpine	Slovenia — LBI	LMI[c]	80% ($n = 30$, $P < 0.001$)	LMI 10–35	Littoral zone, multihabitat sampling
Alpine	Germany — AESHNA for Alpine lakes	Morpho-index	35–45% ($n = 131$, $P < 0.001$)	Morpho-index$_{ALP}$ 1–5	Littoral zone, multihabitat sampling; Better performance for lakes with depth < 15 m
Combination of eutrophication and morphological alterations					
Central-Baltic	Germany — AESHNA for lowland lakes	Morpho-TP index	31% ($n = 491$, $P < 0.001$)	Morpho-TP index 1.5–2.5	Littoral zone, multihabitat sampling;

[a] Acid-neutralizing capacity.
[b] TP — total phosphorus concentration, TN — total nitrogen concentration.
[c] Lakeshore modification index (Peterli and Urbanič, 2013).

Table 8
Description of common pressure and biological metrics.

Type	Region	Pressure addressed	Abbreviation	Description
Pressure indices	Central-Baltic	Morphological alterations	Morpho-index$_{CB}$	Weighted average of percentage of altered shoreline, LUL15[a] and LUL100[b]
	Central-Baltic	Morphological alterations and eutrophication	Morpho-TP index	Weighted average of Morpho-index$_{CB}$ and total phosphorus concentration
	Alpine	Morphological alterations	Morpho-index$_{ALP}$	Weighted average of naturalness of shoreline, altered shoreline, LUS15, LUS100 and LUL100
Biological indices	Central-Baltic	Morphological alterations and eutrophication	ICM$_{CB}$	Weighted average of number of EPTCBO taxa, ASPT, % ETO, % habitat preference lithal
	Alpine	Morphological alterations	ICM$_{ALP}$	Weighted average of Fauna index, taxa richness, reproduction strategy (r/k), % feeding type collector-gatherers

[a] LUL — land use index regarding the lake.
[b] LUS — land use index regarding the site (explanations in Table 2).

metrics (ICMs) were calibrated in order to compare national definitions of good status. The main criteria for the selection of metrics to be included in a multimetric index (Buffagni et al., 2007) were: (1) inclusion of the main aspects outlined for aquatic invertebrates in the WFD (sensitivity, richness and diversity, taxonomic composition), (2) the ability to describe degradation gradients, and (3) the capacity to relate to the national methods in the region.
- Standardization of national classifications using reference sites or, when reference sites are too few or lacking, use of regression to establish pressure-response relationships. This approach, albeit statistically complex, efficiently handles differences among biological datasets, minimizing biogeographical and methodological variations.

4.4. Practical recommendations

In Europe, legislation requires the Member States to develop and intercalibrate benthic invertebrate-based assessment tools for freshwaters and coastal waters. At present, only 10 out of 28 member states have intercalibrated assessment methods for lakes, while in many other member states methods are still largely under development (Poikane et al., 2015). The development of methods is especially important for countries that may join the European Union in the coming years, and for countries on other continents having similar environmental legislation.

This raises the question what is the most appropriate method when designing a monitoring programme (e.g., Borja et al., 2015; Salas et al., 2006). It is widely acknowledged that: (1) greater emphasis should be placed on evaluating the suitability of existing indices prior to developing new ones (Borja et al., 2015) and (2) the most important factor to evaluate method performance is its responsiveness to anthropogenic pressures (Borja et al., 2015; Lyche Solheim et al., 2013). Therefore, we have identified several best-performing methods for addressing diverse human pressures (Table 7) taking into consideration their strength and sensitivity, as well as the amount data used in their development. We have included the % of explained variance, pressure range and habitats assessed for each method that may be used as guidance for selecting the most suitable method.

Additionally, we have developed three pressure metrics and two biological multimetrics (Table 8) for addressing morphological alterations (Alpine region) and combination of morphological alterations and eutrophication (Central-Baltic region). Hence, countries that still develop assessment methods should consider including these methods in their evaluations, although bearing in mind that adaptation of the metrics may be needed to account for region- or type-specific conditions before adoption into national classification systems (Lyche Solheim et al., 2013).

5. Conclusions

The efficacy of benthic invertebrates for assessing anthropogenic effects on lakes has been a topic of debate in the last few decades. Our study shows that benthic invertebrates can be used in lake assessment:

- Thirteen benthic invertebrate-based assessment methods were developed and intercalibrated across Europe, covering different geographical zones and water body types (Belgium, Estonia, Finland, Germany, Lithuania, the Netherlands, Norway, Slovenia, Sweden, and the United Kingdom);
- The benthic invertebrate assessment methods were shown to adequately address several pressures and pressure combinations, i.e. acidification (3 methods), eutrophication (3), hydromorphological alterations (2) and their combinations (5);
- Effective comparison and harmonization of classification boundaries is possible, if: (i) methods are grouped according to pressures and habitats assessed and (ii) appropriate options (direct or indirect comparison) are chosen;
- Furthermore, we identified several best-performing methods addressing three commonly occurring human pressures – acidification, eutrophication, morphological alterations – and a combination of the last two. Moreover, two biological common metrics were developed addressing hydromorphological alterations (Alpine region) and combination of morphological alterations and eutrophication (Central-Baltic region) which can be adopted by countries that have not yet developed benthic assessment tools.

Acknowledgments

This work is the part of the WFD Common Implementation Strategy working group ECOSTAT work programme. All participants in this exercise acknowledge support from national governments and regulatory agencies. We greatly acknowledge the contributions of the national experts: Christine Argillier (France), Rachel Benstead (UK), Angela Boggero (Italy), Fiona Carse (UK), David Colvill (UK), Ian Fozzard (UK), Muriel Gevrey (France), Willem Goedkoop (Sweden), Emma Göthe (Sweden), Ruth Little (UK), Ben McFarland (UK), Heikki Mykrä (Finland), Gwendolin Porst (Germany), Bart Reeze (the Netherlands), and Georg Wolfram (Austria).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.scitotenv.2015.11.021.

References

Armitage, P.D., Moss, D., Wright, J.F., Furse, M.T., 1983. The performance of a new biological water quality score system based on a wide range of unpolluted running-water sites. Water Res. 17, 333–347.
Bazzanti, M., Mastrantuono, L., Solimini, A.G., 2012. Selecting macroinvertebrate taxa and metrics to assess eutrophication in different depth zones of Mediterranean lakes. Fundam. Appl. Limnol. 180, 133–143.
Birk, S., Hering, D., 2009. A new procedure for comparing class boundaries of biological assessment methods: a case study from the Danube Basin. Ecol. Indic. 3, 528–539.
Birk, S., Willby, N., 2010. Towards harmonization of ecological quality classification: establishing common grounds in European macrophyte assessment for rivers. Hydrobiologia 652, 149–163.

Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., Solimini, A., van de Bund, W., Zampoukas, N., Hering, D., 2012. Three hundred ways to assess Europe's surface waters: an almost complete overview of biological methods to implement the Water Framework Directive. Ecol. Indic. 18, 31–41.

Birk, S., Willby, N.J., Kelly, M.G., Bonne, W., Borja, A., Poikane, S., van de Bund, W., 2013. Intercalibrating classifications of ecological status: Europe's quest for common management objectives for aquatic ecosystems. Sci. Total Environ. 454-455, 490–499.

Böhmer, J., Rawer-Jost, C., Zenker, A., 2004. 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., Arbaciauskas, K., Benstead, R., Gabriels, W., Porst, G., Reeze, B., Timm, H., 2014. Water Framework Directive intercalibration technical report: Central-Baltic lake benthic invertebrate ecological assessment methods. EUR 26511 EN. Publications Office of the, European Union, Luxembourg.

Bonada, N., Prat, N., Resh, V.H., Statzner, B., 2006. Developments in aquatic insect biomonitoring: a comparative analysis of recent approaches. Annu. Rev. Entomol. 51, 495–523.

Borja, A., Josefson, A.B., Miles, A., Muxika, I., Olsgard, F., Phillips, G., Rodríguez, J.G., Rygg, B., 2007. An approach to the intercalibration of benthic ecological status assessment in the North Atlantic ecoregion, according to the European Water Framework Directive. Mar. Pollut. Bull. 55, 42–52.

Borja, A., Marin, S., Muxica, I., Pino, L., Rodriguez, J.G., 2015. Is there a possibility of ranking benthic quality assessment indices to select the most responsive to different human pressures? Mar. Pollut. Bull. 97, 85–94.

Brauns, M., Garcia, X.F., Walz, N., Pusch, M.T., 2007. Effects of human shoreline development on littoral macroinvertebrates in lowland lakes. J. Appl. Ecol. 44, 1138–1144.

Brodersen, K.P., 1995. The effect of wind exposure and filamentous algae on the distribution of surf zone macroinvertebrates in Lake Esrom, Denmark. Hydrobiologia 297, 131–148.

Brodersen, K.P., Dall, P.C., Lindegaard, C., 1998. The invertebrate fauna in the upper stony littoral of Danish lakes: macroinvertebrates as trophic indicators. Freshw. Biol. 39, 577–592.

Brose, U., Ostling, A., Harrison, K., Martinez, N.D., 2004. Unified spatial scaling of species and their trophic interactions. Nature 428, 167–171.

Brucet, S., Poikane, S., Lyche Solheim, A., Birk, S., 2013. Biological assessment of European lakes: ecological rationale and human impacts. Freshw. Biol. 58, 1106–1115.

Buffagni, A., Erba, S., Furse, M.T., 2007. A simple procedure to harmonize class boundaries of assessment systems at the pan-European scale. Environ. Sci. Pol. 10, 709–724.

Cao, Y., Hawkins, C.P., 2011. The comparability of bioassessments: a review of conceptual and methodological issues. J. North Am. Benthol. Soc. 30, 680–701.

Carvalho, L., Poikane, S., Lyche Solheim, A., Phillips, G., Borics, G., Catalan, J., De Hoyos, C., Drakare, S., Dudley, B.J., Järvinen, M., Laplace-Treyture, C., Maileht, K., McDonald, C., Mischke, U., Moe, J., Morabito, G., Nõges, P., Nõges, T., Ott, I., Pasztaleniec, A., Skjelbred, B., Thackeray, S.J., 2013. Strength and uncertainty of phytoplankton metrics for assessing eutrophication impacts in lakes. Hydrobiologia 704, 127–140.

Dale, V.H., Beyeler, S.C., 2001. Challenges in the development and use of ecological indicators. Ecol. Indic. 1, 3–10.

Dodson, S.I., Arnott, S.E., Cottingham, K.L., 2000. The relationship in lake communities between primary productivity and species richness. Ecology 81, 2662–2679.

European Commission [EC], 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23rd October 2000 establishing a framework for community action in the field of water policy. Official Journal of the European Communities, L327/1. European Commission, Brussels.

European Commission [EC], 2011. Common Implementation strategy for the Water Framework Directive (2000/60/EC). Guidance Document 14. Guidance on the Intercalibration process 2008–2011. Luxembourg, Office for Official publications of the European Communities.

European Commission [EC], 2013. Commission Decision of 20 September 2013 establishing, pursuant to Directive 2000/60/EC of the European parliament and of the council, the values of the member state monitoring system classifications as a result of the intercalibration exercise. Official Journal of the European Communities L266/1. European Commission, Brussels.

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.

Henrikson, L., Medin, M., 1986. Biologisk bedömning av försurningspåverkan på Lelångens tillflöden och grundområden 1986. Aquaekologerna, Älvsborgs län (in Swedish).

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

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, 311–324.

Jeppesen, E., Jensen, J.P., Søndergaard, M., Lauridsen, T., Landkildehus, F., 2000. Trophic structure, species richness and biodiversity in Danish lakes: changes along a phosphorus gradient. Freshw. Biol. 45, 201–218.

Johnson, R.K., 1998. Spatio-temporal variability of temperate lake macroinvertebrate communities: detection of impact. Ecol. Appl. 8, 61–70.

Johnson, R.K., Goedkoop, W., 2002. Littoral macroinvertebrate communities: spatial scale and ecological relationships. Freshw. Biol. 47, 1840–1854.

Johnson, R.K., Goedkoop, W., 2007. Bedömningsgrunder för bottenfauna i sjöar och vattendrag: Anväändarmanual och bakgrundsdokument. Swedish University of Agricultural Sciences, Report 4 84 p.

Johnson, R.K., Wiederholm, T., Rosenberg, D.M., 1993. Freshwater biomonitoring using individual organisms, populations, and species assemblages of benthic macroinvertebrates. Freshwater Biomonitoring and Benthic Invertebrates. Chapman and Hall, pp. 40–158.

Johnson, R.K., Goedkoop, W., Sandin, L., 2004. Spatial scale and ecological relationships between the macroinvertebrate communities of stony habitats of streams and lakes. Freshw. Biol. 49, 1179–1194.

Johnson, R.K., Goedkoop, W., Fölster, J., Wilander, A., 2007. Relationships between macroinvertebrate assemblages of stony littoral habitats and water chemistry variables indicative of acid-stress. Water Air Soil Pollut. 7, 323–330.

Jyväsjärvi, J., Nyblom, J., Hämäläinen, H., 2010. Palaeolimnological validation of estimated reference values for a lake profundal macroinvertebrate metric (Benthic Quality Index). J. Paleolimnol. 44, 253–264.

Jyväsjärvi, J., Aroviita, J., Hämäläinen, H., 2012. Performance of profundal macroinvertebrate assessment in boreal lakes depends on lake depth. Fundam. Appl. Limnol. 180, 91–100.

Karr, J.R., 1999. Defining and measuring river health. Freshw. Biol. 41, 221–234.

Kelly, M., Bennett, C., Coste, M., Delgado, C., Delmas, F., Denys, L., Ector, L., Fauville, C., Ferreol, M., Golub, M., Jarlman, A., Kahlert, M., Lucey, J., Ní Chatháin, B., Pardo, I., Pfister, P., Picinska-Faltynowicz, J., Rosebery, J., Schranz, C., Schaumburg, J., van Dam, H., Vilbaste, S., 2009. A comparison of national approaches to setting ecological status boundaries in phytobenthos assessment for the European Water Framework Directive: results of an intercalibration exercise. Hydrobiologia 621, 169–182.

Kelly, M., Urbanic, G., Acs, E., Bertrin, V., Burgess, A., Denys, L., Gottschalk, S., Kahlert, M., Karjalainen, S.M., Kennedy, B., Kosi, G., Marchetto, A., Morin, S., Picinska-Fałtynowicz, J., Poikane, S., Rosebery, J., Schoenfelder, J., Schoenfelder, I., Varbiro, G., 2014. Comparing aspirations: intercalibration of ecological status concepts across European lakes using littoral diatoms. Hydrobiologia 734, 125–141.

Lyche Solheim, A., Feld, C.K., Birk, S., Phillips, G., Carvalho, L., Morabito, G., Mischke, U., Willby, N., Søndergaard, M., Hellsten, S., 2013. Ecological status assessment of European lakes: a comparison of metrics for phytoplankton, macrophytes, benthic invertebrates and fish. Hydrobiologia 704, 57–74.

McFarland, B., Carse, F., Sandin, L., 2010. Littoral macroinvertebrates as indicators of lake acidification within the UK. Aquat. Conserv. 20, 105–116.

Miler, O., Brauns, M., Böhmer, J., Pusch, M., 2013a. Feinabstimmung des Bewertungsverfahrens von Seen mittels Makrozoobenthos. Project report Länderfinanzierungsprogramm 'Wasser, Boden und Abfall' (project no. O 5.10/2011).

Miler, O., Porst, G., McGoff, E., Pilotto, F., Donohue, L., Jurca, T., Solimini, A.G., Sandin, L., Irvine, K., Aroviita, J., Clark, R., Pusch, M.T., 2013b. Morphological alterations of lake shores in Europe: a multimetric ecological assessment approach using benthic macroinvertebrates. Ecol. Indic. 34, 398–410.

Miler, O., Ostendorp, W., Brauns, M., Porst, G., Pusch, M.T., 2015. Ecological assessment of morphological shore degradation at whole lake level aided by aerial photo analysis. Fundam. Appl. Limnol. 4, 353–369.

Mittelbach, G.G., Steiner, C.F., Scheiner, S.M., Gross, K.L., Reynolds, H.L., Waide, R.B., Willig, M.R., Dodson, S.I., Gough, L., 2001. What is the observed relationship between productivity and species richness? Ecology 82, 2381–2396.

Osenberg, C.W., Schmitt, R.J., Holbrook, S.J., Abu-Saba, K.E., Flegal, A.R., 1994. Detection of environmental impacts: natural variability, effect size and power analysis. Ecol. Appl. 4, 16–30.

O'Toole, C., Donohue, I., Moe, S.J., Irvine, K., 2008. Nutrient optima and tolerances of benthic invertebrates, the effects of taxonomic resolution and testing of selected metrics in lakes using an extensive European data base. Aquat. Ecol. 42, 277–291.

Peterlin, M., Urbanič, G., 2013. A Lakeshore Modification Index and its association with benthic invertebrates in alpine lakes. Ecohydrol. 6, 297–311.

Poikane, S., Alves, M., Argillier, C., van den Berg, M., Buzzi, F., Hoehn, E., de Hoyos, C., Karottki, I., Laplace-Treyture, C., Lyche Solheim, A., et al., 2010. Defining chlorophyll-a reference conditions in European lakes. Environ. Manag. 45, 1286–1298.

Poikane, S., Portielje, R., van den Berg, M., Phillips, G., Brucet, S., Carvalho, L., Mischke, U., Ott, I., Soszka, H., Van Wichelen, J., 2014a. Defining ecologically relevant water quality targets for lakes in Europe. J. Appl. Ecol. 51, 592–602.

Poikane, S., Zampoukas, N., Borja, A., Davies, S., van de Bund, W., Birk, S., 2014b. Intercalibration of aquatic ecological assessment methods in the EU: lessons learned and way forward. Environ. Sci. Pol. 44, 237–246.

Poikane, S., Birk, S., Carvalho, L., Böhmer, J., de Hoyos, C., Gassner, H., Hellsten, S., Kelly, M., Lyche Solheim, A., Olin, M., Pall, K., Phillips, G., Portielje, R., Ritterbusch, D., Sandin, L., Schartau, A.K., Solimini, A.G., van den Berg, M., Wolfram, G., van de Bund, W., 2015. A hitchhiker's guide to European lake ecological assessment and intercalibration. Ecol. Indic. 52, 533–544.

Resh, V.H., Jackson, J.K., 1993. Rapid assessment approaches to biomonitoring using benthic macroinvertebrates. In: Rosenberg, D.M., Resh, V.H. (Eds.), Freshwater Biomonitoring and Benthic Macroinvertebrates. Chapman and Hall, New York, pp. 195–223.

Reyjol, Y., Argillier, C., Bonne, W., Borja, A., Cardoso, A.C., Daufresne, M., Kernan, M., Ferreira, M.T., Martini, F., Poikane, S., Prat, N., Lyche Solheim, A., Stroffek, S., Usseglio-Polatera, P., Villeneuve, B., van de Bund, W., 2014. Assessing the ecological status in the context of the European Water Framework Directive: where do we go now? Sci. Total Environ. 497-498, 332–344.

Rossaro, B., Marziali, L., Cardoso, A.C., Solimini, A., Free, G., Giacchini, R., 2007. A biotic index using benthic macroinvertebrates for Italian lakes. Ecol. Indic. 7, 412–429.

Ruse, L., 2010. Classification of nutrient impact on lakes using the chironomid pupal exuvial technique. Ecol. Indic. 10, 594–601.

Salas, F., Marcos, C., Neto, J.M., Patrıcio, J., Perez-Ruzafa, A., Marques, J.C., 2006. User-friendly guide for using benthic ecological indicators in coastal and marine quality assessment. Ocean Coast. Manag. 49, 308–331.

Sandin, L., Johnson, R.K., 2000. The statistical power of selected indicator metrics using macroinvertebrates for assessing acidification and eutrophication of running waters. Hydrobiologia 422, 233–243.

Sandin, L., Schartau, A.K., Aroviita, J., Carse, F., Colvill, D., Fozzard, I., Goedkoop, W., Göthe, E., Little, R., McFarland, B., Mykrä, H., 2014. Water Framework Directive intercalibration technical report: northern lake benthic invertebrate ecological assessment methods. Publications Office of the European Union, Luxembourg.
Schartau, A.K., Moe, S.J., Sandin, L., McFarland, B., Raddum, G.G., 2008. Macroinvertebrate indicators of lake acidification: testing on data from UK, Norway and Sweden. Aquat. Ecol. 42, 293–305.
Šidagytė, E., Višinskiene, G., Arbačiauskas, K., 2013. Macroinvertebrate metrics and their integration for assessing the ecological status and biocontamination of Lithuanian lakes. Limnologica 43, 308–318.
Solimini, A.G., Sandin, L., 2012. The importance of spatial variation of benthic invertebrates for the ecological assessment of European lakes. Fundam. Appl. Limnol. 180, 85–89.
Solimini, A.G., Free, G., Donohue, I., Irvine, K., Pusch, M., Rossaro, B., Sandin, L., Cardoso, A.C., 2006. Using benthic macroinvertebrates to assess ecological status of lakes. EUR 22347 EN. Luxembourg: Office for Official Publications of the European Communities.
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. J. North Am. Benthol. Soc. 27, 878–891.
Team RC, 2012. R: a language and environment for statistical computing. http://www.r-project.org/.
Timm, H., Möls, T., 2012. Littoral macroinvertebrates in Estonian lowland lakes: the effects of habitat, season, eutrophication and land use on some metrics of biological quality. Fundam. Appl. Limnol. 180/2, 145–156.
Tóth, L.G., Poikane, S., Penning, E., Free, G., Mäemets, H., Kolada, A., Hanganu, J., 2008. First steps in the Central-Baltic intercalibration exercise on lake macrophytes: where do we start? Aquat. Ecol. 42, 265–275.
Townsend, C.R., Scarsbrook, M.R., Dolédec, S., 1997. The intermediate disturbance hypothesis, refugia, and biodiversity in streams. Limnol. Oceanogr. 42, 938–949.
Trigal, C., García-Criado, F., Fernández-Aláez, C., 2009. Towards a multimetric index for ecological assessment of Mediterranean flatland ponds: the use of macroinvertebrates as bioindicators. Hydrobiologia 618, 109–123.
Urbanič, G., 2014. A Littoral Fauna Index for assessing the impact of lakeshore alterations in Alpine lakes. Ecohydrol. 7 (2), 703–716.
Urbanič, G., Petkovska, V., Pavlin, M., 2012. The relationship between littoral benthic invertebrates and lakeshore modification. Fundam. Appl. Limnol. 180, 157–173.
White, J., Irvine, K., 2003. The use of littoral mesohabitats and their macroinvertebrate assemblages in the ecological assessment of lakes. Aquat. Conserv. 13, 331–351.
Wiederholm, T., 1980. Use of zoobenthos in lake monitoring. J. Water Poll. Cont. Fed. 52, 537–547.
Young, J., Watt, A., Nowicki, P., Alard, D., Clitherow, J., Henel, K., Johnson, R., Laczko, E., McCracken, D., Matouch, S., Niemela, J., Richards, C., 2005. Towards sustainable land use: identifying and managing the conflicts between human activities and biodiversity conservation in Europe. Conserv. Biol. 14, 1641–1661.