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

Sandra Poikane[1], Richard K. Johnson[2], Jukka Arrovita[3], Jürgen Böhmer[4], Oliver Miler[5], Martin Pusch[5], Leonard Sandin[6], Ann Kristin Schartau[7], Angelo Solimni[8], Gorazd Urbanic [9,10]

[1] European Commission, Joint Research Centre, Institute for Environment and Sustainability, 21027 Ispra, Italy

[2] Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Box 7050, SE 750 07 Uppsala, Sweden

[3] Freshwater Centre, Finnish Environment Institute SYKE, 90014 Oulu, Finland

[4] Bioforum GmbH, Sudetenstr. 34, 73230 Kirchheim/Teck, Germany

[5] Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 301, 12587 Berlin, Germany

[6] Institute of Freshwater Research, Department of Aquatic Resources, Swedish University of Agricultural Sciences, Stångholmsvägen 2, SE-178 93 Drottningholm, Sweden

[7] Norwegian Institute for Nature Research, Gaustadalléen 21, 0349 Oslo, Norway

[8] Department of Public Health and Infectious Diseases, Sapienza University of Rome, 00185 Rome, Italy

[9] Institute for Water of the Republic of Slovenia, 1000 Ljubljana, Slovenia

[10] University of Ljubljana, Biotechnical Faculty, Department of Biology, 1000 Ljubljana, Slovenia

Keywords: Biological metrics; benthic macroinvertebrates; ecological assessment; lakes; Water Framework Directive; pressure-response relationships

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 recognised 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 that have recently been elaborated. 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 intercalibration groups, according to habitat sampled and pressure indicated. Boundaries of the 'good ecological status' were compared and harmonised 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). Therefore, there is 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 (Karr and Chu 2000, Bonada et al. 2006, 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 in the ecological assessment of inland and coastal waters (Hering et al. 2010, 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 are intended based on phytoplankton, macrophytes and phytobenthos, benthic invertebrates, and fish fauna. Biological assessment results have to be expressed as Ecological Quality Ratios (EQR), defined as the observed state / expected state. The EQR is divided into five status classes (high, good, moderate, poor, and bad), the most important distinction being that between good and moderate status, 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. 2014).

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 (Johnson et al. 1993, Resh and Jackson 1993, Birk et al. 2012). As macroinvertebrate communities may respond predictably to several human-induced stressors, their use is widespread, constituting the basis of many biomonitoring programs (e.g. Resh 2008, Birk et al. 2012), and fulfil many of the criteria for an ideal biomonitoring tool listed by Bonada et al. (2006). However, most studies advocating the use of macroinvertebrates in biomonitoring so far have focused on stream habitats (Resh and Jackson 1993, Hering et al. 2006, Birk et al. 2012). By contrast, fewer studies have addressed the efficacy of using lake macroinvertebrate assemblages in biomonitoring (see White and Irvine 2003, Johnson et al. 2004, 2007a, Brauns et al. 2007 a, b). A decade ago, the paucity of WFD compliant macroinvertebrate assessment tools was identified as one of the major gaps impeding the full assessment of the ecological quality of lakes (Solimini et al. 2006). Since then, stimulated by 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 hydromorphological alterations (cf. Young et al. 2005). Early lake assessment approaches using benthic invertebrates focused mainly on indicating eutrophication using profundal invertebrate communities (Thienemann 1918, Wiederholm 1980). Building on this earlier work, several WFD compliant eutrophication assessment metrics based on profundal invertebrate communities have been developed (Ruse 2010, Jyväsjärvi et al. 2010, 2012). Furthermore, littoral macroinvertebrate-based metrics have also been developed for assessing the impacts of acidification (Johnson et al. 2007a, Schartau et al. 2008, McFarland et al. 2010). Conversely, studies of the effects of hydromorphological alterations in lakes on littoral macroinvertebrates and their use as indicator organisms have remained scarce until recently. In recent years, knowledge in that field has increased markedly (Brauns et al. 2007a, 2011, McGoff and Irvine 2009, McGoff et al. 2013, Gabel et al. 2012, Porst et al. 2012, Czarnecka et al. 2014, Miler et al. 2015, Pilotto et al. 2015) which enabled developing approaches for the use of littoral benthic fauna to assess the ecological effects of morphological alterations (Urbanič et al. 2012, Urbanič 2014, Miler et al. 2013a, 2015).

When developing assessment tools for lakes based on benthic invertebrates, it has to be considered that macroinvertebrate assemblages in the eulittoral and sublittoral habitats (and to some extent even in the profundal habitat; Pilotto et al. 2012) are often affected by multiple pressures (Brauns et al. 2007a, 2007b, Jurca et al. 2012, Pilotto et al. 2015). For example, eulittoral macroinvertebrate assemblages respond not only to hydromorphological alteration, but also to eutrophication (Brauns et al. 2007b, Donohue et al. 2009, Jurca et al. 2012, McGoff and Sandin 2012, and Pilotto et al. 2012, 2015). Hence, some macroinvertebrate-based assessment methods likely indicate the combined effects of hydromorphological and eutrophication pressures, as well as acidification (Gabriels et al. 2010, Timm and Möls 2012, Šidagytė et al. 2013). In contrast, as water managers need to know what pressure(s) are causing impairment, pressure-specific biotic indication is preferred over the indication of 'general degradation' (Solimini et al. 2006, Hering et al. 2015).

A basic requirement for successful river basin management is the comparability of bioassessment approaches used in the area, as different data and indices can lead to inconsistent or conflicting assignment to ecological status classes (Cao and Hawkins 2011, Birk et al. 2013). In Europe, legislation stipulates that the values for the upper and lower "good" class boundaries should be harmonised through the intercalibration exercise. Therefore, the intercalibration was undertaken 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. 2014).

This task is particularly difficult for methods used in monitoring benthic invertebrate assemblages in lakes. 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 for assessing eutrophication (Thienemann 1918, Wiederholm 1980). Another difficulty is that the large biogeographical range among EU countries results in high natural variability (lake/habitat types) and in different types of impairment that need consideration. The response of the methods to certain human impacts is clearly influenced by the type and severity of impacts occurring in the respective country (Böhmer et al. 2014, Solimini et al. 2012). Densely populated central European countries, such as the Netherlands or Belgium, feature mostly degraded water bodies (Gabriels et al. 2010, Böhmer et al. 2014), whereas lakes in the northern part 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;
- To compare the lake assessment methods proposed by several countries and achieve a harmonisation of class boundaries;
- To 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 (IC): 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).

Table 1. Overview of lake benthic invertebrate assessment methods developed by various member states (MS) 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 Watertypes	**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 et al. (2007), Urbanič (2014)
United Kingdom	Chironomid Pupal Exuviae Technique (CPET)	**UK-CPET**	Whole lake, eutrophication	Ruse (2010)
United Kingdom	Lake Acidification Macroinvertebrate	**UK-LAMM**	Eulittoral, acidification	McFarland et al. (2010)

	Metric (LAMM)

Most of the methods (9 methods) 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 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 harmonisation.

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

(2) *Data collection and choosing the appropriate IC option* - Thirteen countries provided data from national monitoring or ongoing activities focused on developing 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). However, partitioning natural variability by lake type and regions still resulted in relatively large datasets: 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 (S2 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).

Two IC options 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. Buffagni et al. 2007, Bennett et al. 2011).

(3) *Development of pressure indices and biological common metrics*

The aim of pressure indices was to synthesize available information on morphological pressures into a single index value. In the Alpine region, five pressure variables were standardized to values from 1 to 5 (continuous values): (i) naturalness of shoreline at the sampling site; land use index within (ii) 15 m (LUS_{15}) and (iii) 100 m (LUS_{100}) from the sampling site; (iv) land use index within 100 m from the lake (LUL_{100}) and (v) % of altered shoreline around the lake (detailed description in Table S4). A pressure index (Morpho-index$_{ALP}$) for each sampling site was calculated using weighted averaging of standardized pressure variables as:

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

For the Central Baltic region, three pressure variables were standardized from 1 to 5 (continuous values): land use index within (i) 15 m (LUL_{15}) and (ii) 100 m (LUL_{100}) from the lake and (iii) % of altered shoreline around the lake. The pressure index was calculated as:

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

Additionally, an index comprising both morphological alterations and eutrophication (Morpho-TP index) was calculated in the Central-Baltic region based on the standardised values of Morpho-index$_{CB}$ and the annual mean concentration of total phosphorus (TP) as:

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

For a description of the calculation of land use indices see Table S4.

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 * site matrices. Many 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).

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). More details on these can be found in Table S9. 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 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 (see Table S9).

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 normalised 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 Table S10). 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 autocorrelation 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.

(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 a detailed description see EC 2010, Birk et al. 2013).

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 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 (R Core Team, 2012). Standardized ICM metric values were obtained by subtracting the offsets from the metric values.

(5) *Method comparison and harmonisation* – 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 view 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 only of one metric (see description of metrics in S5). 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) (LT and EE), Benthic Quality Index (Wiederholm 1980) (SE and FI), and Acidity Index (Henrikson and Medin 1986) (NO and EE), 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 three 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, UK: R^2 = 0.37 to 0.80) and anion neutralising capacity (ANC) (UK, NO: R^2 = 0.47 to 0.82) (for detailed information see Table S6).

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 2012, 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 = 0.32 to 0.69), some against morphological pressures (NL, DE, LT: R^2 = 0.33 to 0.67), and some assessed combinations of pressures (LT, DE: correlation for combined morphology and nutrients was slightly larger (0.22; 0.31) than for morphology alone (0.11; 0.25)). Finally, three methods addressed only the impacts of eutrophication. CPET scores were related (R^2=0.78, $P< 0.001$) to a compound pressure metric (total nitrogen x total phosphorus/mean depth). The BQI was significantly related with total phosphorus concentration (SE, FI: R^2 = 0.27-0.32, $P<0.001$), with stronger relationships observed in deep lakes (mean depth > 6m; Jyväsjärvi et al. 2012)

3.3. Intercalibration

Intercalibration groups and options

In total, four groups of methods were established according to the region, lake types, pressures and habitats (Table 2). 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.

The choice of intercalibration approach depended on how similar the assessment methods were among the countries participating in the exercise. In the Alpine and Central Baltic regions, methods differed in field sampling (sampling season, 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).

Table 2. Overview of the lake intercalibration groups (only finalized exercises).
EUTR: eutrophication, HM – hydromorphological modifications, ACID – acidification

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

Development of common metrics for intercalibration

Construction of intercalibration common metrics (ICM) by 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 (rstrategists/k-strategists), and (iv) % abundances of the feeding type collector-gatherers (% FG) calculated as: ALP-ICM = (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) % abundances with a preference for the lithal microhabitat (% HL) calculated as:

CB-ICM = (2*EPTCBO + ASPT + % ETO +% HL)/5.

Both ICMs correlated significantly with most of the pressure variables (Table 3). 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.

Table 3. Correlations between ICM and pressure variables (for explanations see Material and Methods above and S3).

	ALP-ICM		CB-ICM	
	Pearson's r	**P**	**Pearson's r**	**P**
Pressure variables:				
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

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. The 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, standardisation 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. 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 standardisation to normalize UK LAMM values.

In the Central Baltic and Alpine regions, sufficient data of reference sites were not available. Therefore, regression standardization (linear mixed models) was used to standardise 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 standardised single metrics into a common multimetric, all countries followed the common pressure response model.

Comparison of national metrics and ICMs

For all three regions, relationships between country metrics and ICMs were highly significant (Table 4), with slopes within the interval of 0.5 to 1.5. For the two countries in the Alpine region, DE and SI, 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 and SE BQI metrics for addressing eutrophication pressures was highly significant (r = 0.68, P < 0.001).

Table 4. Results of regression between national metrics and common Intercalibration (IC) metrics

MS	Pearson's r	Slope	P	Intercalibration approach / Intercalibration Common metrics (ICM)
Alpine region				
DE	0.76	0.98	< 0.001	Indirect comparison via ICM: 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-FL	0.56	0.99	< 0.001	Indirect comparison via ICM: Weighted average of normalised values of number of EPTCBO taxa, ASPT, % ETO, % Habitat preference lithal
DE	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 EQR - SE BQI	0.68	0.70	< 0.001	Direct comparison (regression of two methods)

a) Germany b) Slovenia

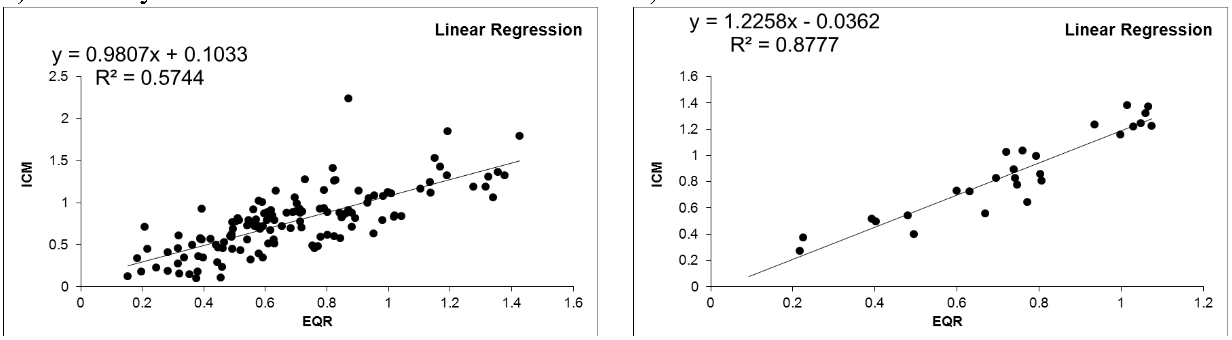


Fig 1. Linear regressions between national benthic invertebrate lake assessment methods and the intercalibration common metric (ICM) in Alpine lakes: a) Germany, b) Slovenia. For further regressions see Figure S7

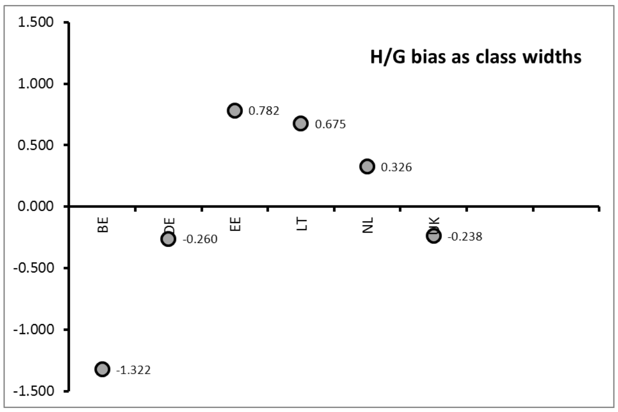
Harmonisation of class boundaries

The analysis of national boundaries for all three regions showed relatively good agreement with the global harmonization boundaries. For the Alpine region's 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 0.6 to 0.63, while the High/Good boundary value for the SE BQI was decreased from 0.9 to 0.84 and the Good/Moderate boundary from 0.7 to 0.67. In the Central Baltic regions, national boundaries from three assessment methods (BE, EE, LT) deviated by more than 0.25 class equivalents. The Belgian national 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 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 5.

Fig 2. 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 widths 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 Figure S8.

a) High-Good class boundary



b) Good-Moderate boundary

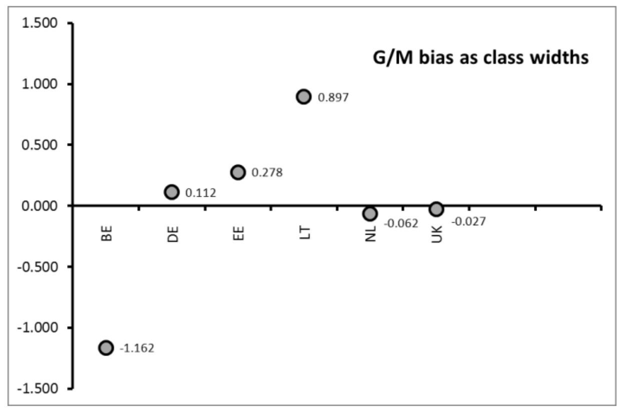


Table 5. Final H/G and G/M boundary EQR values for the national methods included in the EC Decision (EC 2013)

Region/ Member State	**Classification Method**	**Ecological Quality Ratios**	
		High-good boundary	**Good-moderate boundary**
ALP			
SI	Lake littoral benthic invertebrate index (LBI)	0.80	0.60
DE	German Macroinvertebrate Lake Assessment (AESHNA, part eulittoral of Alpine/Prealpine lakes)	0.80	0.60
CB			
BE-FL	Multimetric Macroinvertebrate Index Flanders (MMIF)	0.90	0.70
DE	German Macroinvertebrate Lake Assessment (AESHNA, part eulittoral of lowland lakes)	0.80	0.60
EE	Estimation of freshwater quality using macroinvertebrates	0.86	0.70
LT	Lithuanian Lake Macroinvertebrate Index (LLMI)	0.74	0.50
NL	WFD Metric for Natural Watertypes	0.80	0.60

UK	Chironomid Pupal Exuvial Technique (CPET)	0.77	0.64
NOR	Lake littoral acidification		
SE	Multimetric Invertebrate Lake Acidification index (MILA)	0.85	0.60
UK	Lake Acidification Macroinvertebrate Metric (LAMM)	0.86	0.70
NO	Multimetric Invertebrate Index for Clear Lakes (MultiClear)	0.95	0.74
NOR	Lake profundal eutrophication		
FI	Benthic Quality Index (BQI)	0.75	0.63
SE	Benthic Quality Index (BQI)	0.84	0.67

4. Discussion

Macroinvertebrates have traditionally been recognized as one of the most difficult biological groups for use in lake ecological assessment due to a number of reasons, such as their complex biotic structure, relatively high temporal variability and the high spatial heterogeneity (White and Irvin 2003, Brose et al. 2004; Solimini and Sandin, 2012). Accordingly, the use of macroinvertebrate communities in lake assessment programmes has been limited so far (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 (Karr and Chu 2000, Hering et al. 2006, 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 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 a 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 therefore 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, SE, and FI). 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, Irigoien et al. 2004), 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 (Resh 1979, Barbour et al. 1992, Resh and Jackson 1993, Thorne and Williams 1997, Johnson 1998). 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 (Hering et al. 2004, Böhmer 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); (2) incorrect assignment of taxa into functional groups (Karr 1999, Rawer-Jost et al. 2000, 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 (Moss et al. 2003, Schartau et al. 2008, Menetrey et al. 2005, 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 (Karr 1999, Dale and Beyeler 2001, Hering et al. 2006, Johnson et al. 2007b). For benthic invertebrates in lakes, several studies have shown weak or no pressure-response relationships, especially for littoral invertebrates and eutrophication pressure (Moss et al. 2003, Garcia-Criado et al. 2005, O'Toole et al. 2008, Timm and Möls 2012, Bazzanti et al. 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 (Johnson and Goedkoop 2002, Garcia-Criado et al. 2005, Brauns et al. 2007a, Timm and Möls 2012) may significantly superimpose 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 last two pressures (2). Factors that were likely important in isolating pressure-response relationships were:

- The 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.

- The appropriate description of a pressure gradient. This is 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.
- The 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. 2007a, Pilotto et al. 2012), 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 was developed to aid in the analysis of pressure-response relationships for this pressure combination (Šidagytė et al. 2013).
- The careful selection of assessment metrics. In theory, all metric types need to be included in the assessment methods (Karr and Chu 1999, Hering et al. 2006). 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).
- The 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. Gabriels et al. 2010, 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 harmonised to achieve comparable results (Cao and Hawkins 2011, Birk et al. 2013). In Europe, legislation mandates the comparison and harmonisation of assessment methods used by different countries, i.e. intercalibration (Poikane et al. 2014). Several examples of intercalibration have been described for rivers: benthic invertebrates (Buffagni et al. 2007, Bennett et al. 2011), diatoms (Kelly et al. 2009), macrophytes (Birk and Hering 2009), for lakes: phytoplankton (Poikane et al. 2010, 2014), macrophytes (G.-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 et al. 2006), (ii) the scarcity of reference sites and difficulties in defining comparable reference conditions (Birk and Hering 2009, Bennett et al. 2011, Kelly et al. 2014) and (iii) large biogeographical and methodological differences among the countries (Kelly et al. 2014) which 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 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/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, minimising 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 brings the question into focus of what is the most appropriate method when designing a monitoring programme (e.g. Salas et al. 2006, Borja et al. 2015). 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 a method's performance is its responsiveness to anthropogenic pressures (Lyche Solheim et al. 2013, Borja et al. 2015). Therefore, we have identified several best-performing methods for addressing diverse human pressures (Table 6) taking into consideration their strength and sensitivity, as well as data amount 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 7) 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).

Table 7. Common pressure and biological indices. LUL – Land use index regarding the lake, LUS- land use index regarding the site (explanations of calculation S4)

Type	Region	Pressure addressed	Abbreviation	Description
PRESSURE INDICES	Central-Baltic	Morphological alterations	Morpho index MI-CB	Weighted average of percentage of altered shoreline, LUL15 and LUL100
	Central-Baltic	Morphological alterations and eutrophication	Morpho-TP index TMI-CB	Weighted average of MI-CB and total phosphorus concentration
	Alpine	Morphological alterations	Morpho index MI-ALP	Weighted average of naturalness of shoreline, altered shoreline, LUS15, LUS100 and LUL100
BIOLOGCAL INDICES	Central-Baltic	Morphological alterations and combination of HM and eutrophication	Intercalibration Common Metrics ICM-CB	Weighted average of number of EPTCBO taxa, ASPT, % ETO, % habitat preference lithal
	Alpine	Morphological Alterations	Intercalibration Common Metrics ICM-ALP	Weighted average of Fauna index, taxa richness, reproduction strategy (r/k), % feeding type collector-gatherers

Conclusions

The efficacy of benthic invertebrates in assessing the 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 harmonisation 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.

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.

Barbour, M.T., Plafkin, J.L., Bradley, B.P., Graves, C.G. Wisseman, R.W., 1992. Evaluation of the EPA‟s Rapid Bioassessment benthic metrics: metric redundancy and variability among reference stream sites. Environ. Toxicol. Chem. 11, 437- 449.

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.

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

Birk, S., Hering, D., 2006. Direct comparison of assessment methods using benthic macroinvertebrates: a contribution to the EU Water Framework Directive intercalibration exercise. Hydrobiologia 56, 401-415.

Birk, S., Hering, D., 2009. A new procedure for comparing class boundaries of biological assessment methods: a case study from the Danube Basin. Ecol. Ind. 3, 528-539.

Birk, S., W. Bonne, A. Borja, S. Brucet, A. Courrat, 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. Ind. 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., K. Arbaciauskas, 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. Publications Office of the European Union, Luxembourg.
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., Rawer-Jost, C., Zenker, A., Meier, C., Feld, C.K., Biss, R., Hering, D., 2004. Assessing streams in Germany with benthic invertebrates: Development of a multimetric invertebrate based assessment system. Limnologica 34, 416-432.
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., 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.
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.

Brauns, M., Garcia, X., Pusch, M., Walz, N., 2007b. Eulittoral macroinvertebrate communities of lowland lakes: discrimination among trophic states. Freshwater Biol. 52, 1022-1032.

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

Brauns, M., Gücker, B., Wagner, C.,Walz, N., Pusch, M.T., 2011. Human lakeshore development alters the structure and trophic basis of littoral food webs. J. Appl. Ecol. 48(4), 916-925.

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. Freshwater 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. Freshwater 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. Policy 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.

Czarnecka, M., Pilotto, F., Pusch, M.T., 2014. Is coarse woody debris in lakes a refuge or a trap for benthic invertebrates exposed to fish predation? Freshwater Biol., 5, 2400-2412.

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

Davy-Bowker J., Murphy J.F., Rutt G.R., Steel J.E.C. and Furse M.T. (2005) The development and testing of a macroinvertebrate biotic index for detecting the impact of acidity on streams. Archiv fur Hydrobiologie, 163, 383– 403.

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.

Donohue, I., Jackson, A., Pusch, M.T., Irvine, K., 2009. Nutrient enrichment homogenizes lake benthic assemblages at local and regional scale. Ecol. 90, 3470–3477.

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.

Gabel, F., Garcia, X.F., Schnauder, I., Pusch, M.T., 2012. Effects of ship-induced waves on littoral benthic invertebrates. Freshwater Biol. 57: 2425-2435.

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.

García-Criado, F., Becares E., Fernandez-Aláez, C., Fernandez-Aláez, M., 2005. Plant-associated invertebrates and ecological quality in some Mediterranean shallow lakes: implications for the application of the EC Water Framework Directive. Aquatic Conserv. 15: 31-50.

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

Hering, D., Borja, A., Carstensen, J., Carvalho L., Elliott M., Feld C.K., Heiskanen A.S., Johnson, R.K., Moe, J., Pont, D., Lyche Solheim, A., van de Bund, W., 2010. The European Water Framework Directive at the age of 10: a critical review of the achievements with recommendations for the future. Sci. Total Environ. 408:4007-4019.

Hering, D., Carvalho, L., Argillier, C., Beklioglu, M., Borja, A., Cardoso, A. C., Duel, H., Ferreira, T., Globevnik, L., Hanganu, J., Hellsten, S., Jeppesen, E., Kodes, V., Lyche Solheim, A., Nõges, T., Ormerod, S., Panagopoulos, Y., Schmutz, S. Venohr, M., Birk, S. 2015. Managing aquatic ecosystems and water resources under multiple stress - an introduction to the MARS project. Sci. Total Environ. 503-504, 10-21.

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.

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

Irigoien, X., Huisman, J., Harris, R.P., 2004. Global biodiversity patterns of marine phytoplankton and zooplankton. Nature 429, 863-867.

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. Freshwater Biol. 45, 201-218.

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

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

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

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

Johnson, R.K., Furse, M.T., Hering, D., Sandin L., 2007b. Ecological relationships between stream communities and spatial scale: implications for designing catchment-level monitoring programs. Freshwater Biol. 52, 939-958.

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

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

Jurca, T., Donohue, L., Laketić, D., Radulović, S., Irvine, K., 2012. Importance of the shoreline diversity features for littoral macroinvertebrate assemblages. Fund. Appl. Limnol. 180, 175-184.

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

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.

Karr, J.R., 1999. Defining and measuring river health. Freshwater Biol. 41, 221-234.

Karr, J.R., Chu, E.W., 2000. Sustaining living rivers. Hydrobiologia 422/423, 1-14.

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.

McGoff, E., Aroviita, J., Pilotto, F., Miler, O., Solimini, A., Porst, G., Jurca, T., Donohue, L., Sandin, L., 2013. Assessing the relationship between the Lake Habitat Survey and littoral macroinvertebrate communities in European lakes. Ecol. Ind. 25, 205-214.

McGoff, E., Irvine, K., 2009. A test of the association between Lake Habitat Quality Assessment and macroinvertebrate community structure. Aquat. Conserv. 19, 520-533.

McGoff, E., Sandin, L., 2012. Catchment land-use effects on littoral macroinvertebrates in response to local habitat structure and trophic state. Fund. Appl. Limnol. 180,111-121.

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., 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. Fund. Appl. Limnol. 4, 353-369.

Miler, O., Porst, G., McGoff, E., Pilotto, F., Donohue, L., Jurca, T., Solimini, A.G., Sandin, L., Irvin, 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. Ind. 34, 398-410.

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.

Moss, B., Stephen, D., Alvarez, C., Becares, E., van de Bund, W., Collings, S.E., Van Donk, E., de Eyto, E., Feldmann, T., Fernandez-Alaez, C. et al., 2003. The determination of ecological status in shallow lakes - a tested system (ECOFRAME) for implementation of the European Water Framework Directive. Aquatic Conserv. 13, 507-549.

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.

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.

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

Pilotto, F., Bazzanti, M., Di Vito, V., Frosali, D., Livretti, F., Mastrantuono, L., Pusch, M.T., Sena, F., Solimini, A.G., 2015. Relative impacts of morphological alteration to shorelines and eutrophication on littoral macroinvertebrates in Mediterranean lakes. Freshwater Sci. 34, 410-422.

Pilotto, F., Free, G., Cardoso, A.C., Wolfram, G., Solimini, A.G., 2012. Spatial variance of profundal and sublittoral invertebrate benthic communities in response to eutrophication and morphological pressures. Fund. Appl. Limnol. 180, 101-110.

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. Manage. 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. Ind., 52, 533-544.

Porst, G., Bader, S., Münch, E., Pusch, M.T., 2012. Sampling approaches for the assessment of shoreline development based on littoral macroinvertebrates: the case of Lake Werbellin, Germany. Fund. Appl. Limnol. 180, 123-131.

Rawer-Jost, C., Böhmer, J., Blank, J., Rahmann, H., 2000. Macroinvertebrate functional feeding group methods in ecological assessment. Hydrobiologia 422, 225-232.

Resh, V.H., 1979. Sampling variability and life history features: basic considerations in the design of aquatic insect studies. J. Fish Res. Board Can. 36, 290-311.

Resh, V.H., 2008. Which group is best? Attributes of different biological assemblages used in freshwater biomonitoring programs. Environ. Monit. Assess. 138, 131-138.

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

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. Ind. 7, 412-429.

Ruse, L., 2010. Classification of nutrient impact on lakes using the chironomid pupal exuvial technique. Ecol. Ind. 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. Manage. 49, 308-331.

Sandin L, Solimini AG. 2012. Spatial variation in lake benthic macroinvertebrate ecological assessment: a synthesis of European case studies. *Fundamental and Applied Limnology* **180**: 185–191.

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. Luxembourg: Publications Office of the European Union.

Schartau, A.K., Moe, S.J., Sandin, L., McFarland, B., Raddum, G., 2008. Macroinvertebrate indicators of lake acidification: testing on data from UK, Norway and Sweden. Aquat. Ecol. 42, 293-305.

Šidagytė, E., Visinskiene, G., Arbaciauskas, K., 2013. Macroinvertebrate metrics and their integration for assessing the ecological status and biocontamination of Lithuanian lakes. Limnologica 43, 308-318.

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

Solimini, A.G., Sandin, L., 2012. The importance of spatial variation of benthic invertebrates for the ecological assessment of European lakes. Fund. Appl. Limnol. 180, 85-89.

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.

Thienemann, A. 1918. Untersuchungen über die Beziehungen zwischen dem Sauerstoffgehalt des Wassers und der Zusammensetzung der Fauna norddeutscher Seen. Arch. Hydrobiol. 12, 1-65.

Thorne, R.S.J., Williams, W.P., 1997. The response of benthic macroinvertebrates to pollution in developing countries: a multimetric system of bioassessment. Freshwater Biol. 37, 671-686.

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. Fund. Appl. Limnol. 180/2, 145-156.

G.-Tóth, L., 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. Fund. Appl. Limn. 180, 157-173.

Urbanič, G., Petkovska, V., Pavlin, M., 2007. Razvoj metodologije za vrednotenje hidromorfološke spremenjenosti jezer na podlagi bentoških nevretenčarjev v skladu z zahtevami Vodne direktive (Direktiva 2000/60/ES). Biotechnical Faculty, Department of Biology, University of Ljubljana. Ljubljana, 39 pp. (In Slovenian)

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.