Water Framework Directive Intercalibration Technical Report|: Central Baltic Lake Benthic invertebrate ecological assessment methods

BOOK · JANUARY 2014

DOI: 10.2788/71003

8 AUTHORS, INCLUDING:

Kęstutis Arbačiauskas
Nature Research Centre
51 PUBLICATIONS **324** CITATIONS
SEE PROFILE

Wim Gabriels
Flanders Environment Agency
22 PUBLICATIONS **393** CITATIONS
SEE PROFILE

Gwendolin Porst
Technische Universität Berlin
14 PUBLICATIONS **92** CITATIONS
SEE PROFILE

Sandra Poikane
European Commission
49 PUBLICATIONS **760** CITATIONS
SEE PROFILE

J R C T E C H N I C A L R E P O R T S

Water Framework Directive Intercalibration Technical Report

Central Baltic Lake Benthic invertebrate ecological assessment methods

Jürgen Böhmer, Kestutis Arbaciauskas
Rachel Benstead, Wim Gabriels,
Gwendolin Porst, Bart Reeze, Henn Timm

Edited by Sandra Poikane

2014

Report EUR 26504 EN

Joint Research Centre

European Commission
Joint Research Centre
Institute for Environment and Sustainability

Contact information
Sandra Poikane
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 46, 21027 Ispra (VA), Italy
E-mail: sandra.poikane@ec.europa.eu
Tel.: +39 0332 78 9720
Fax: +39 0332 78 9352

http://ies.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

This publication is a Technical Report by the Joint Research Centre of the European Commission.

JRC88273

EUR 26504 EN

ISBN 978-92-79-35457-1

ISSN 1831-9424

doi: 10.2788/7103

Cover photo: Sandra Poikane

Luxembourg: Publications Office of the European Union, 2014

Printed in Ispra, Italy

Introduction

The European Water Framework Directive (WFD) requires the national classifications of good ecological status to be harmonised through an intercalibration exercise. In this exercise, significant differences in status classification among Member States are harmonized by comparing and, if necessary, adjusting the good status boundaries of the national assessment methods.

Intercalibration is performed for rivers, lakes, coastal and transitional waters, focusing on selected types of water bodies (intercalibration types), anthropogenic pressures and Biological Quality Elements. Intercalibration exercises were carried out in Geographical Intercalibration Groups - larger geographical units including Member States with similar water body types - and followed the procedure described in the WFD Common Implementation Strategy Guidance document on the intercalibration process (European Commission, 2011).

In a first phase, the intercalibration exercise started in 2003 and extended until 2008. The results from this exercise were agreed on by Member States and then published in a Commission Decision, consequently becoming legally binding (EC, 2008). A second intercalibration phase extended from 2009 to 2012, and the results from this exercise were agreed on by Member States and laid down in a new Commission Decision (EC, 2013) repealing the previous decision. Member States should apply the results of the intercalibration exercise to their national classification systems in order to set the boundaries between high and good status and between good and moderate status for all their national types.

Annex 1 to this Decision sets out the results of the intercalibration exercise for which intercalibration is successfully achieved, within the limits of what is technically feasible at this point in time. The Technical report on the Water Framework Directive intercalibration describes in detail how the intercalibration exercise has been carried out for the water categories and biological quality elements included in that Annex.

The Technical report is organized in volumes according to the water category (rivers, lakes, coastal and transitional waters), Biological Quality Element and Geographical Intercalibration group. This volume addresses the intercalibration of the Central Baltic Benthic invertebrate ecological assessment methods.

Contents

Annexes

1. Introduction

In the Central-Baltic Lake Benthic invertebrates GIG:

- Six Member States (Belgium-Flanders, Estonia, Germany, Lithuania, the Netherlands and UK) submitted their benthic invertebrates assessment methods;
- After evaluation of WFD compliance and the IC feasibility, all methods were included in the IC exercise as all address eutrophication and hydromorphological alterations (except UK - only eutrophication) and follow the same assessment concept focusing on eulittoral (except UK - the whole lake samples);
- Intercalibration "Option 2" was used - indirect comparison of assessment methods using a common metric;
- IC common metric was developed specifically for this IC exercise comprising 4 metrics, it was benchmark-standardized using "continuous benchmarking" approach;
- The comparability analysis showed considerable boundary disagreement, so the method modification was needed for BE and boundary adjustment was needed for LT and EE which brough all boundaries in the harmonization band;
- The final results include EQRs of BE-FL, DE, EE, LT, NL and UK lake benthic invertebrates' assessment systems for 2 common types: L-CB1 and L-CB2.

2. Description of national assessment methods

In the Central-Baltic Benthic invertebrates GIG, six countries participated in the intercalibration (Table 2.1, more details in Annex A).

Table 2.1 Overview of the national lake benthic invertebrate assessment methods in the Central-Baltic GIG.

Member State	Method	Status
BE/FL	Multimetric Macroinvertebrate Index Flanders (MMIF)	Finalized agreed national method
DE	German Macroinvertebrate Lake Assessment (AESHNA)	Intercalibratable finalized method
EE	Estimation of Freshwater Quality Using Macroinvertebrates	Finalized agreed national method
LT	Lithuanian Lake Macroinvertebrate Index	Intercalibratable finalized method
NL	WFDi - Metric for Natural Watertypes	Finalized agreed national method
UK	Chironomid Pupal Exuvial Technique (CPET)	Finalized agreed national method

2.1. Methods and required BQE parameters

All MS have developed full BQE methods (see Table 2.2, for scientific literature and computation details see Annex A).

Table 2.2 Overview of the metrics included in the national benthic invertebrates assessment methods (RA – relative abundance)

MS	Full BQE	Taxonomic composition	Abundance	Disturbance sensitive taxa	Diversity	Combination rule of metrics
BE/ FL	yes	Number of EPT taxa; Number of other sensitive taxa	RA included	Number of EPT taxa; Number of other sensitive taxa; mean tolerance score	Shannon-Wiener diversity index; Total number of present taxa; Number of ept taxa; Number of other sensitive taxa	Average metric scores
EE	yes	EPT taxa richness	RA included	EPT taxa richness; ASPT index; Swedish Acidity index	Shannon diversity; Total taxa richness	Average metric scores
DE	yes	RA of Odonata; RA of habitat type Lithal (% of abundance classes); RA of Chironomidae (% of abundance classes) for riverine lakes	Rel. abundance included	Fauna index	Number of ETO-Taxa; Margalef-Diversität for riverine lakes	Average metric scores
LT	yes	Number of EPTCBO-taxa RA of EPC-taxa	RA included	ASPT index	Shannon diversity	Average metric scores
NL	yes	KM% = rel. number of typical (for watertype) species in a sample %DN = RA of dominant negative species; %DP = RA of dominant positive species	RA included	See tax. comp.	KM% = rel. number of typical (for watertype) species in a sample. This metric is highly correlated with numer of taxa	Weighted averaging of metrics
UK	yes	Chironomid Pupal Exuviae Technique (CPET)	No (see footnote*)	CPET metrics	No (see footnote*)	Not applicable

*UK: CPET has a strong relationship to the pressure gradient for N and P (r^2 = 0.78), this often being greater than that of other benthic invertebrate metrics that include abundance and diversity indices. The inclusion of taxa richness and diversity metrics was tested, but resulted in significantly poorer correlations with the stressor.

2.2. Sampling and data processing

All MS methods have similar sampling and data processing methods (see Table 2.3).

Table 2.3 Overview of the sampling of the national benthic invertebrates assessment methods.

MS	Sampling methodology
BE/FL	1/year: April to November; handnet: standard handnet with 500 µm mesh size • With the handnet, a stretch of approximately 10-20 meters is sampled during 3-5 minutes. Sampling effort is proportionally distributed over all accessible aquatic habitats. kick sampling is performed by vertically positioning the handnet on the bed and turning over bottom material located immediately upstream by foot or hand. • In addition to the handnet sampling, animals are manually picked from stones, leaves or branches along the same stretch. If a site is too deep to be sampled with the handnet method, macroinvertebrates can alternatively be sampled using the so-called Belgian artificial substrates. These are composed of a plastic netting filled with medium-sized (4-8 cm) pieces of brick, with a total volume of approximately 5 L. Per sampling site, three substrates are placed in the water, anchored with a rope to a fixed point located on the bank. After an exposure time of at least 3 weeks, the substrates are lifted from the water and transferred into a closed container.
EE	1/year: April - May or September-October Standard Handnet with 25 cm edge length, 0.5 mm mesh size From the most typical bottom at the sampling site, five 1-m long kick- or sweep replicates are taken and kept separately. A separate qualitative sample is collected from all available substrates, not considering their area.
DE	Minimum one occasion per year: February to April (lowland) / to May (alpine) or September to October. Mainly hand net (eulittoral method); Eckman grab (sublittoral method) Habitat specific sampling designed for sampling all available habitats at up to 1.2 m depth of water. 0.6 to 1,0 m2 should be sampled per habitat. The area sampled and the relative presence of each habitat is determined for a later combination to a multi-habitat taxa list. Sampled are sorted out in the field or sieved, fractionated and preserved in ethanol for sorting in the laboratory. • eulittoral sampled with handnet for most habitats or suitable device for substrates not suited for handnets (e.g. modified Surber sampler for sand or scratcher for concrete surfaces); • sublittoral sampled with Ekman grabs
LT	1/year: April to November; Hand net (frame 25X25 cm>sub>2) All available habitats per site (Multi-habitat)

MS	Sampling methodology
	Standard method: 12 kick or sweep replicates from different microhabitats; Additional survey: Semi-quantitative sampling procedure is carried out using a standard dip-net (25x25 cm). Sampling can be performed in either of the two core eulittoral mesohabitats: a bottom (preferably hard) kick sample or a vegetation (preferably submerged) sweep sample. Within a stand of either mesohabitat, a stretch of about 15-20 meters long is sampled while moving along the shore in a trajectory of a zigzag curve (from the very shoreline to the depth of 1 m) in a way to result in 3 minutes of actual catching time. A semi-quantitative sample is supported by qualitative (search) sample (duration 1 minute) within the same mesohabitat.
NL	Minimum one occasion per year (spring), but classification preferably averaged over three years: March until 15 June. Handnet 30x15 cm Van Veen or Eckman-Birge grab, boxcorer. All available habitats per site (Multi-habitat) Multihabitat sampling in all habitats present in proportion to their presence. Active moving of handnet through vegetation and bottom substrates.
UK	Sampling occasions: 4 recommended, 2 minimum, during April to October Hand net with 250 µm mesh net Collect floating debris at leeward shore (to which wind is blowing), where floating pupal exuviae will be accumulating from across the lake over the previous 2 days.

National reference conditions

Reference condition setting was performed on the national level:

- As reference sites are not available in most of countries, other approaches (expert judgement, modeling. Historical data) were used;
- Only Estonia, Lithuania and UK used ref sites in setting their reference conditions (+ other approaches).

Table 2.4 Overview of the methodology used to derive national reference conditions

Member State	Methodology used to derive the reference conditions
BE/FL	Expert judgement (based on data and expert validation)
EE	Existing near-natural reference sites, expert knowledge, least disturbed conditions
DE	Modelling (extrapolation, percentiles) combined with expert-judgement
LT	Percentiles based on existing near-natural reference sites, expert knowledge
NL	Expert knowledge, historical data, least disturbed conditions (no actual existing natural sites in lakes)
UK	Existing near-natural reference sites, expert knowledge, historical data, modeling (extrapolating model results)

2.2.1. National boundary setting

Boundary setting was performed on the national level (see Table 2.5):

- Most countries used equal division of the EQR range in combination with expert judgement to set the boundaries (see Annex A for details for the countries);
- The GIG group accepted all procedures of the countries as valid and decided that there was no need to do a boundary setting on GIG-level. Therefore, the average view of the countries' class boundaries was used as common view of the group.

Table 2.5 Overview of the methodology used to derive ecological class boundaries

MS	Methodology used to set class boundaries
BE/FL	Boundaries used for most river types (resulting from intercalibration exercise) are applied to lakes as well: Boundaries for rivers were derived by equidistant division of the EQR gradient was applied (boundaries at 0.8, 0.6, 0.4, 0.2. Based on the river intercalibration results (phase 1, same method is applied for rivers), the boundaries were upgraded a little bit.
DE	The single metrics were standardised from 1 (90%tile decreasing metric, 10%tile increasing metric) to 0 (10%tile decreasing metric, 90%tile increasing metric). The average of all metrics - the resulting multimetric index - is standardised from 1 (extrapolated reference condition) to 0 (extrapolated bad condition). This range was equally split up into the 5 quality classes (very good 1-0,8; good 0,8-0,6; moderate 0.6 to 0.4; poor 0,4-0,2; bad 0.2-0).
EE	Boundaries used for most river types (resulting from intercalibration exercise) were applied to lakes as well. Boundaries were adjusted by expert judgement based on pre-classified sampling sites. All index values of high quality were assigned five points, values of good quality, four points, values of moderate quality, two points, and values of poor and bad quality, zero points. The difference between good level and moderate level was intentionally emphasized in order to underline the principal difference between them in terms of the Water Framework Directive. Multimetric quality (MMQ) was then calculated by adding up the corresponding points. Hence, for small lakes, the reference value was 25 and the sum 23−25 was considered to indicate high, 18−22, good, 10−17, moderate, 6−9, poor and <6, bad quality.
LT	Reference values for each of the metrics were derived as 90% percentiles of metric values of samples from reference sites. Same reference values apply for CB1 and CB2 lake types. Equidistant division of the multimetric EQR gradient (boundaries at 0.8, 0.6, 0.4, 0.2). No relation to the pressure has been used.
NL	Boundaries are derived from a theoretical reference and deviding the range into five equal classes. There has been no comparison with pressure data yet. The pressure data from the current IC database provide a good basis for that

MS	Methodology used to set class boundaries
UK	CPET: Using paired metrics (percent relative frequency of sensitive and tolerant species) that respond in different ways to the influence of the pressure (GM boundary); other boundaries by equidistant division and expert judgement

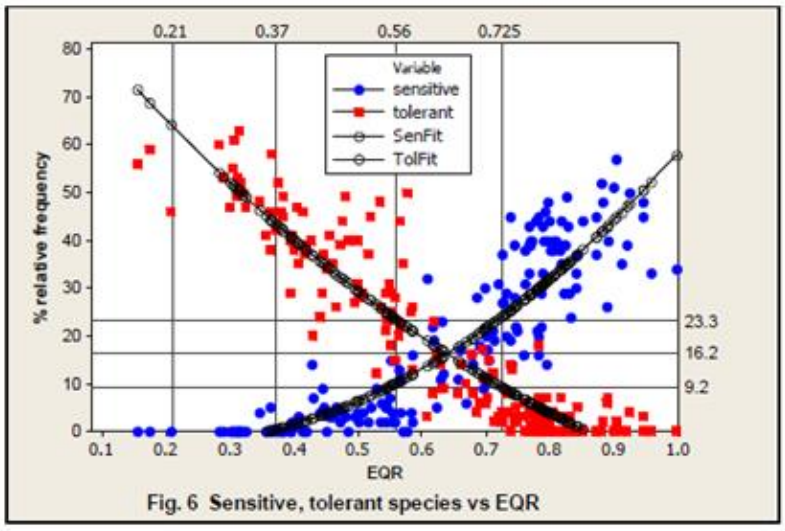


Figure 2.1 EQR class boundaries defined by crossover of sensitive and tolerant species frequency regression fits. Empty squares are tolerant species frequency. Filled circles are sensitive species frequency with fitted regression lines

3. Results of WFD compliance checking

All methods include wide range of metrics, still, UK include only taxonomic composition metrics reflecting sensitive/ tolerant taxa. UK could justify that the criteria of the WFD were not completely covered :

- The UK method CPET is species richness metric for eutrophication that comprises the annual species composition of chironomiidae as represented by the pupal exuviae sampled from a lake. It has a strong relationship to the pressure gradient for N and P (R^2 0.78), this often being greater than that of other benthic invertebrate metrics that include abundance and diversity indices;
- While relative species abundance parameters are recorded for CPET, they are not currently included in the metric, as tests showed that they do not improve the relationship with these pressures.
- The inclusion of taxa richness and diversity metrics was tested, but resulted in significantly poorer correlations with the stressor.

In conclusion, all methods are considered to be compliant. The table below lists the criteria from the IC guidance and compliance checking conclusions.

Table 3.1 List of the WFD compliance criteria and the WFD compliance checking process and results

Compliance criteria	Compliance checking conclusions
1. Ecological status is classified by one of **five classes** (high, good, moderate, poor and bad).	Yes for all methods
2. High, good and moderate ecological status are set in line with the WFD's **normative definitions** (**Boundary setting procedure**)	Yes for all methods
3. **All relevant parameters** indicative of the biological quality element are covered (see Table 1 in the IC Guidance). A **combination rule** to combine para-meter assessment into BQE assessment has to be defined. If parameters are missing, Member States need to demonstrate that the method is sufficiently indicative of the status of the QE as a whole.	Yes for all methods*
4. Assessment is adapted to **intercalibration common types** that are defined in line with the typological requirements of the WFD Annex II and approved by WG ECOSTAT	Yes, all methods are suited for the assessment of the intercalibration common types
5. The water body is assessed against **type-specific near-natural reference conditions**	Yes for all methods
6. Assessment results are expressed as **EQRs**	Yes for all methods
7. Sampling procedure allows for **represent-tative** information about water body quality/ ecological status **in space and time**	Yes for all methods
8. All data relevant for assessing the biological **parameters** specified in the WFD's normative definitions are covered by the **sampling procedure**	Yes for all methods
9. Selected taxonomic level achieves adequate **confidence and precision** in classification	Yes for all methods

4. Results IC Feasibility checking

Typology

Intercalibration feasible in terms of typology (see Table 4.1) :

- **IC is feasible** for L-CB1 and L-CB2;
- **IC is not feasible for L-CB3**, because of missing methods (as FR and LV did not participate).

Even when methods will be established for FR and LV, the differences between the environmental characteristics of LV and FR lakes might hamper a successful comparison.

Table 4.1 Description of common intercalibration water body types and the MS sharing each type

Common IC type	Type characteristics	MS sharing IC common type
L-CB1	Lowland shallow stratified calcareous	All, except FR
L-CB2	Lowland very shallow stratified calcareous	All, except FR
L-CB3	Lowland shallow stratified siliceous	Only FR and LV

Pressures addressed

Intercalibration is feasible in terms of pressures addressed by the methods:

- However, focus on hydromorphological alterations and eutrophication differs a little among the MS (see table below);
- UK method focuses on eutrophication only.

Table 4.2 Description of the pressures addressed by the MS assessment methods

Method	Pressure
BE/FL: Multimetric Index	Hydromorphological alterations Eutrophication
EE: Multimetric Index	Hydromorphological alterations Eutrophication
DE: Multimetric Index (AESCHNA)	Hydromorphological alterations (Eutrophication_
LT: Multimetric Index	(Hydromorphological alterations) Eutrophication
NL: WFDi	Hydromorphological alterations (Eutrophication)
UK: CPET	Eutrophication

Most of MS have tested their national method against pressures and found significant reslationships (see table below)

More details to pressure-response relationships:

EE: Pressure – national assessment for CB lakes

Ecological data from 20 LCB1 and LCB2 lakes were examined to establish pressure-impact relationship between macroinvertebrate metrics, eutrophication and green land use in catchment area. There occurred significant correlations between multimetric quality (on the basis of 5 indices) and the following parameters: total phosphorus content in water (r=-0.32), water quality on the basis of phytoplankton (0.58), and green land cover (0.64).

DE: Pressure – national assessment for CB lakes (graphs from IC exercise)

The German AESHNA method mainly responds to morphological alterations, and only weakly to nutrients and the land use in the catchments. However, R^2 for combined morphology and nutrients is slightly larger than for morphology alone.

LT: Pressure – national assessment for lakes

The Lithuanian LLMI responds is best correlated with BOD, but also significantly with several chemical and morphological parameters (see table below). Correlations with the chemical parameters are somewhat stronger than with morphology; however correlations are not very strong, due to the lack of pressure gradient.

Table 4.3 Information on pressure-response relationships

MS	Metrics tested	Pressure	Pressure indicators	Strength of relationship
BE/FL	Multimetric Index	Relationship was thus far only tested for rivers; it is assumed that the relationship is similar for lakes but this was not tested yet		
EE	Multimetric Index	Hymo Eutro	Natural and semi natural landuse, TP	Landuse: R^2 = 0.41 (n=20) TP: R^2 = 0.1 (n=14)
DE	Multimetric Index (AESHNA)	Hymo (Eutro)	several morphological parameters, nutrients and trophic status	Morphology: R^2 = 0.1 – 0.25 (significant) Eutrophication: R^2 <= 0.1 (some significant) Combined Morphology-TP: R^2 = 0.30 (significant)
LT	Multimetric Index	(Hymo) Eutro	tP, tN, chl-a, BOD	BOD: significant (Spearman)
NL	WFDi	Hymo (Eutro)	Shore alteration%, Shore artificial%, Shore natural vegetation%, tP	Shore alteration%: R^2 = 0.,45 (n=32) Shore artificial%: R^2 = 0.15 (n=32) Shore natural vegetation%: R^2 = 0.16 (n=32) TP: higher metric scores (>0.5) only at TP <0,1 mg/l
UK	CPET	Eutro	pressure metric: tN / mdepth * tP	R^2 = 0.78 (n=166)

Table 4.4 Correlations of the Lithuanian LLMI with some pressure parameters.

	N	Spearman R	p
Transparency	58	0,45	0,000292
Chlorophyll *a*	66	-0,40	0,000742
TP	66	-0,43	0,000294
TN	66	-0,27	0,026321
BOD_7	50	-0,51	0,000128
Morphoindex	64	-0,32	0,009595
Morpho-TP	64	-0,39	0,001306

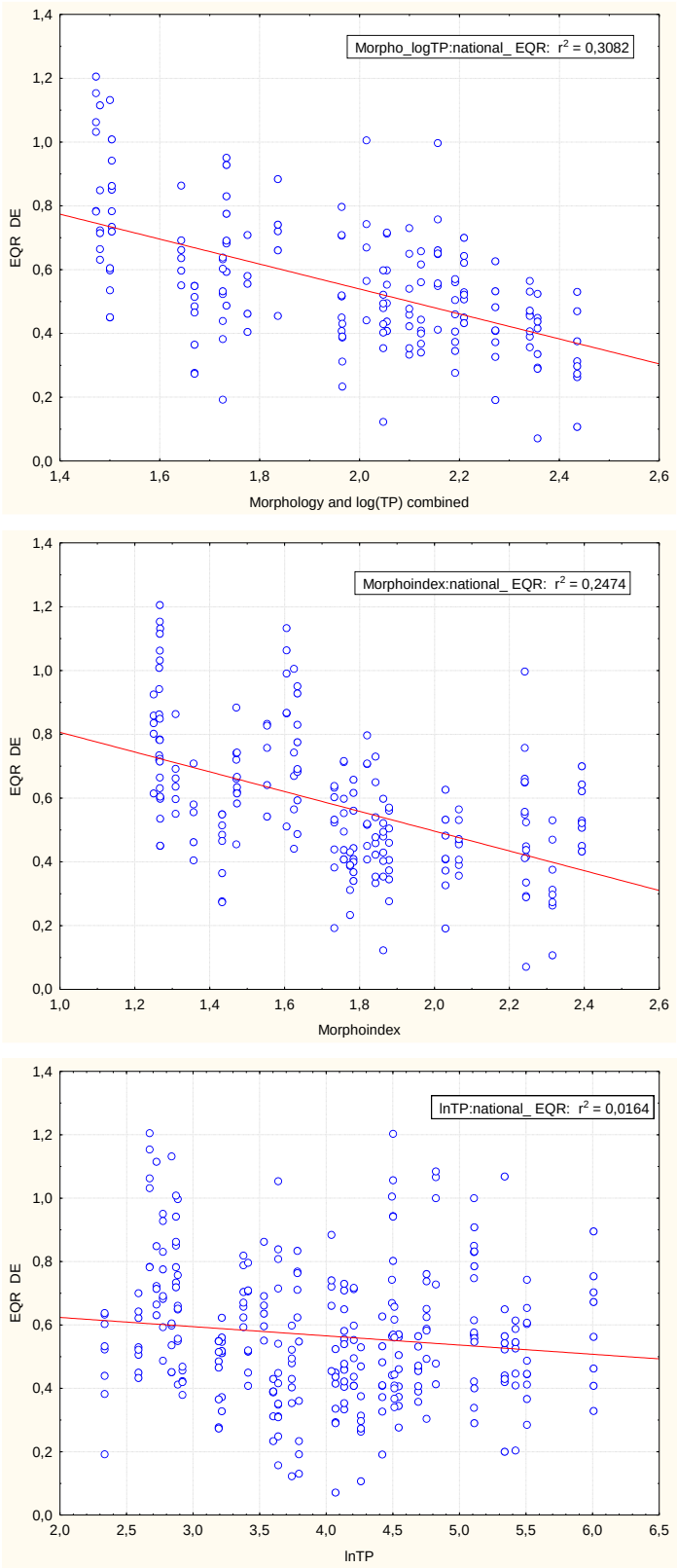


Figure 4.1 Response of German national lake benthic fauna method to eutrophication pressure, morphological alterations and combined pressure

NL: Pressure – national assessment for lakes and streams together

The Dutch WFDI responds well to general degradation, especially hydromorphology (see Figure below).

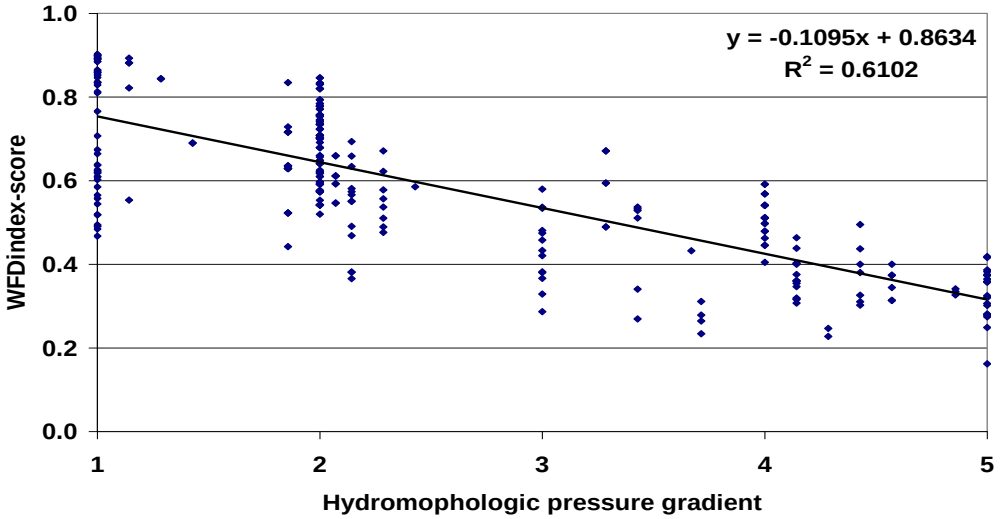


Figure 4.2 Correlation of WFDi with hydromophological pressure gradient (n = 279)

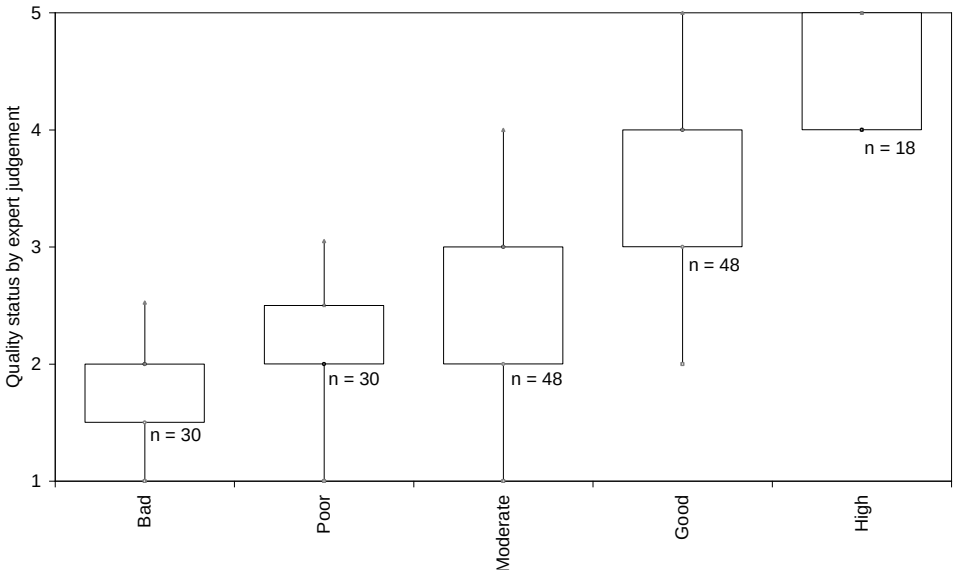
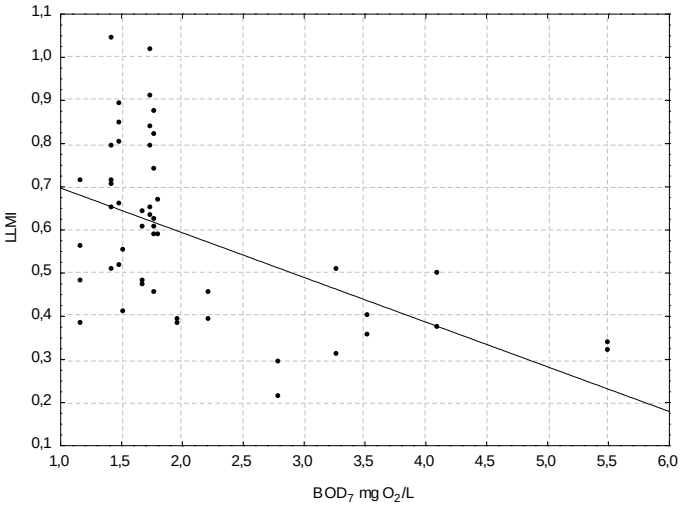
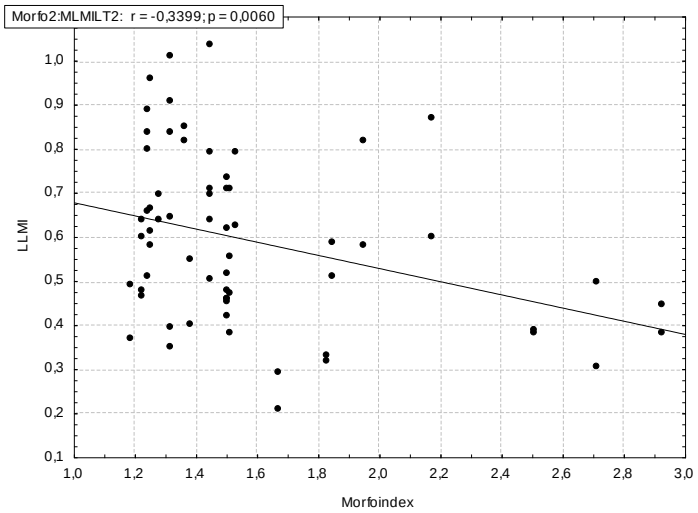


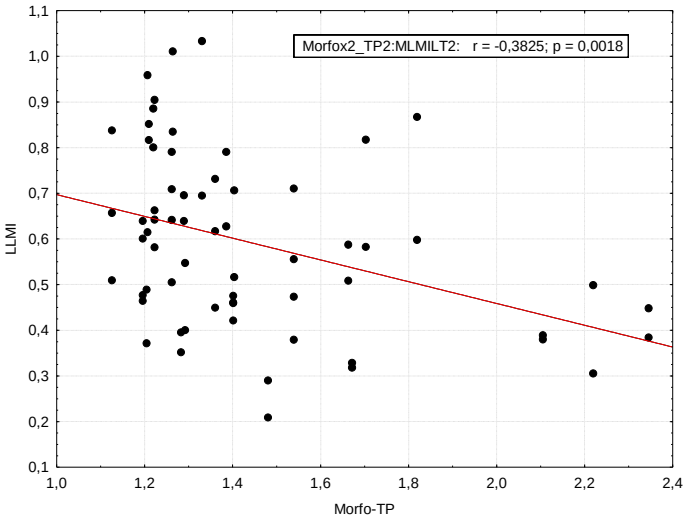
Figure 4.3 Distribution of mean WFDi-scores with expert judgement (n = 6 expert judgements on 29 samples)



R2=0.279



R2=0.116



R2=0.146

Figure 4.4 Response of Lithuanian national lake benthic fauna method to eutrophication pressure, morphological alterations and combined pressure

UK: CPET for lakes

The UK CPET method responds best to eutrophication.

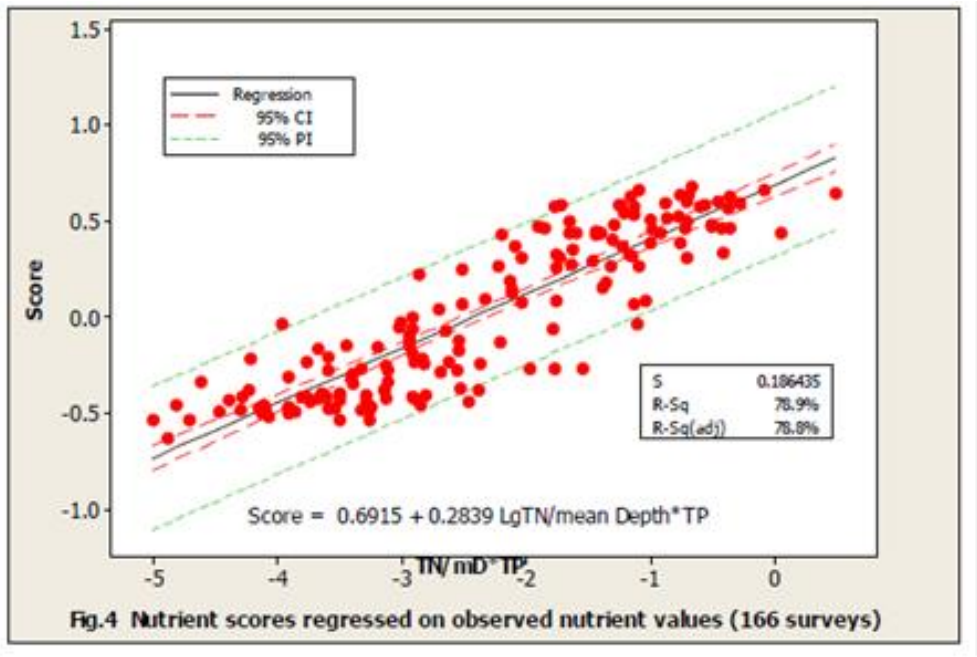


Figure 4.5 Response of UK CPET method to the eutrophication pressure gradient

Assessment concept

All national methods follow a similar assessment concept:

- Therefore the intercalibration is feasible for eulittoral methods;
- The UK method is not directly comparable due to differences in sampling design, but a linkage via parallel eulittoral samples was made.

Table 4.5 Assessment concepts of national assessment methods

Method	Assessment concept	Remarks
BE, DE, EE, NL, LT	Eulittoral macroinvertebrates, sampled by handnet	
UK	Chironomid exuviae representing all lake zones	Parallel eulittoral samples were collected for a number of lakes

5. Collection of IC dataset

A huge dataset was collected in the Central Baltic benthic invertebrate GIG (see Table 5.1)

Table 5.1 Overview of the Central Baltic Benthic invertebrate GIG dataset

Member State	Number of sites or samples or data values		
	Biological data	Physico- chemical data	Pressure data
BE	12 lakes / 55 sites	12	12
DE	54 lakes / 410 sites	54	(54)
DK	17 lakes / 79 sites	17	(17)
EE	20 lakes / 20 sites	20	20
UK	26 lakes / 26 sites (CB-GIG) + 47 lakes / 70 sites (NGIG)	26 (CB-GIG) + 47 (NGIG)	(26) + 47 (NGIG)
LT	26 lakes / 26 sites	26	26
LV	23 lakes / 23 sites	23	23
NL	32 lakes / 113 sites	32	32
PL	6 lakes / 36 sites (littoral) 11 lakes (CPET)	6 11 (CPET)	6 11 (CPET)

For UK only lakes with both eulittoral and CPET samples were taken into account; only eulittoral samples were selected for other countries.

Since all taxonomic data could be assigned to the relevant taxonomic codes, and differences in taxonomic precision could be largely minimised by applying a taxonomic harmonisation list, all datasets were accepted for the analysis. All datasets fulfilled the requirements, although many are incomplete (mainly with regard to stressor variables). Therefore, not all data could be included in each analysis.

6. Common benchmarking

The common approach for setting reference conditions

The standard approach by applying reference criteria (see Table 6.1) was tried: Due to the scarcity of resulting reference sites, alternative approaches became necessary. There was not a suitable set of alternative benchmarks to cover all MS, therefore 'continuous benchmarking' (named 'full regression curve procedure' in the previous reports) was applied .

Table 6.1 Reference criteria for screening of sites in near-natural conditions

	Criteria [1]	Notes [2]	Reference threshold	Rejection threshold
Catchment characteristics	(1) Reference threshold > 85 % nature (i.e. "natural" forests, wetlands, moors, meadows, pasture); NOTE: Rejection threshold = 70 %	Land use is determined using CORINE categories, if more accurate national maps are not available. "Not natural" (opposite to "nature") are agricultural land and urban areas. Forest that are planted and fertilized (e.g. spruce cultures used as christmas trees etc.) are "not natural". They should be regarded as agricultural land. Pasture are extensively grazed grassland.	>=90	>=70
	(2) No intensive crops (incl. vines) within in the near surroundings (i.e. within a zone of 200 m from the lake shore)	provide numerical value		
	(3) ≤ 5 % urbanisation and peri-urban areas in the near surroundings (i.e. within a zone of 200 m from the lake shore)		<=5 *	<=5 *
	(4) No direct inflow of treated or untreated waste water			
	(5) Impact of wastewater from scattered dwellings low (i.e. < 10 inhabitants km^{-2}) within the whole catchment	Inferred from national maps; number of houses multiplied by the national average of inhabitats per household; provide numerical value	<10 *	<10 *
Morphology	(6) ≤ 5 % artificial modification of the shore line	provide numerical value	<=5 *	<=5 *
Trophic state	(7) Generally: No (or insignificant) deviation of the actual from the natural trophic state			
Other pressures	(8) No mass (or significant) recreation activities (camping, swimming, roing, coarse fish angling, put			

	Criteria [1]	Notes [2]	Reference threshold	Rejection threshold
	and take angling, releasing and feeding of ducks for hunting)			
	(9) No actively invading (and reproducing) plant or animal species that may negatively impact the structure, productivity, function and diversity of the ecosystem			
10.	(10) no evidence for one of the following pressures: - Significant changes in the hydrological and sediment regime of the tributaries (larger than the range between the natural mean low water level and the natural high water level) - Fish farm activities or other fishing operations that negatively impact the structure, productivity, function and diversity of the ecosystem - Introduction of non-native fish species, unless their abundance and biomass is insignificant - Sigificant changes in status parameters prior to major changes in industrialisation, urbanisation and intensification of the agriculture - Substances mentioned in Annex X and/or in annex VIII of the WFD in concentrations above the limits of detection of the most advanced analytical techniques in general use or presence of possible and important sources of pollutants. - Measured values of other anthropogenic, synthetic substances above quality objectives and not near natural background concentrations, except for those from atmospheric sources			

1. The criteria are provided based on: /1/ CIRCA, Feb. 2008, "WFD Intercallibration technical report, Part 2 - lakes, section 3 - phytoplankton composition"; /2/ CIRCA, Feb. 2008, "WFD Intercallibration technical report, Part 2 - lakes, section 3 - macrophytes"; /3/ "CB GIG Rivers reference criteria"
2. Some of the criteria are difficult to assess - due to the lack of data, and/or because there are qualitatively rather than quantitatively defined

Values were preliminary and it had been intended to further elaborate them; but this did not happen, since the reference approach had been dropped.

Since no sufficient number of reference sites / alternative benchmarks could be found, the whole pressure response relationship was used as kind of "multiple continuous benchmark". This approach was described by Boehmer et al. 2011 (see Annex B). The adjustment is done by selecting the standardisation value (offset or factor) for a type/country which adjusts it to the dose response regression for all types/countries

together. This procedure is also suitable (and probably the only way) to benchmark types/countries with only very few or even no reference or alternative benchmark values.

Benchmark standardisation

Benchmark standardization as described in the Guidance has not been applied. Instead continuous benchmarking has been applied to standardise all single common metrics (see Annex B for detailed description of the principle). The offset has been determined using Linear Mixed Models with the biological metrics as dependent variable, the combined pressure variable as covariates and the country as random factor. For this purpose the package 'lme4' of the 'R'-software was used.

To obtain standardised metrics the offsets given by the model were subtracted from the metric values in most cases (see Figure 6.1 as example).

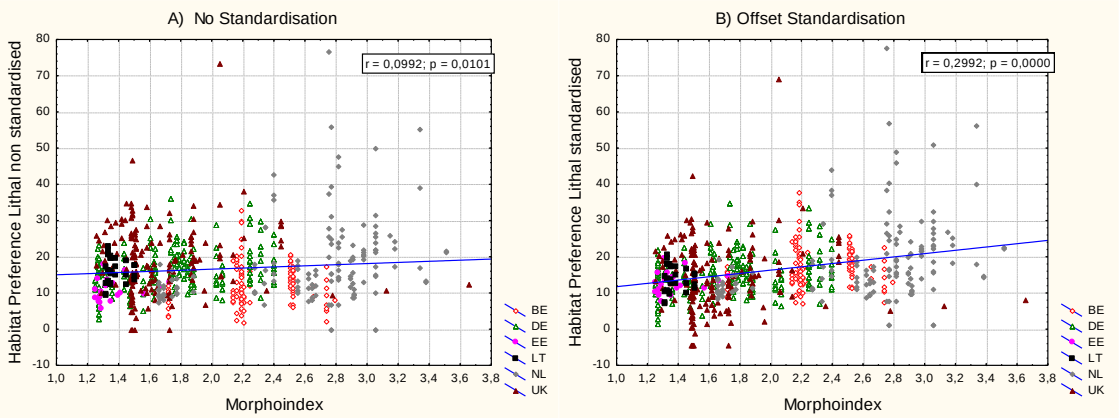


Figure 6.1 Metric for habitat preference lithal in dependence of the morphology index. A) before and B) after offset standardisation.

However for taxa number metrics (e.g. number of ETO-taxa), this was inappropriate, because negative values occurred after standardisation, and because the slopes of the dose response curves differed too much (s. Figure 6.2B and example of NL in Figure 6.3B below). Therefore a factor was applied instead in that way, that the intersection of the standardised metric with the common regression line of all countries together remained the same as when subtracting the offset (s. Figure 6.2C and Figure 6.3C below). As result this procedure gives identical standardised values at the intersection point of the regression lines, which is usually near the centre of the data spread of the country. The figure below illustrates the effects of this procedure.

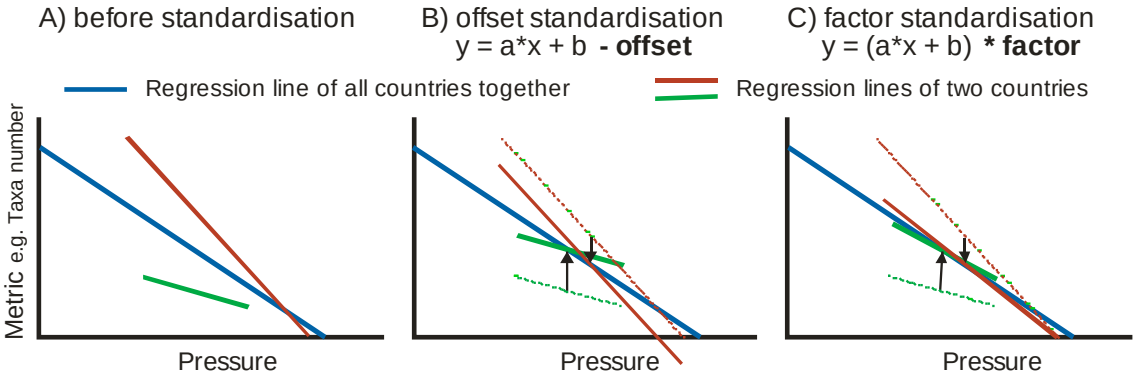


Figure 6.2 Effects of standardisation on common metrics with strongly deviating slopes of the dose-response-curves (A). Offset standardisation does not change the slopes, but may lead to negative values (B). Factor standardisation gives the same corrections at the intersections of the standardised metrics, but also changes the slopes (C). Dotted lines = regression lines before standardisation.

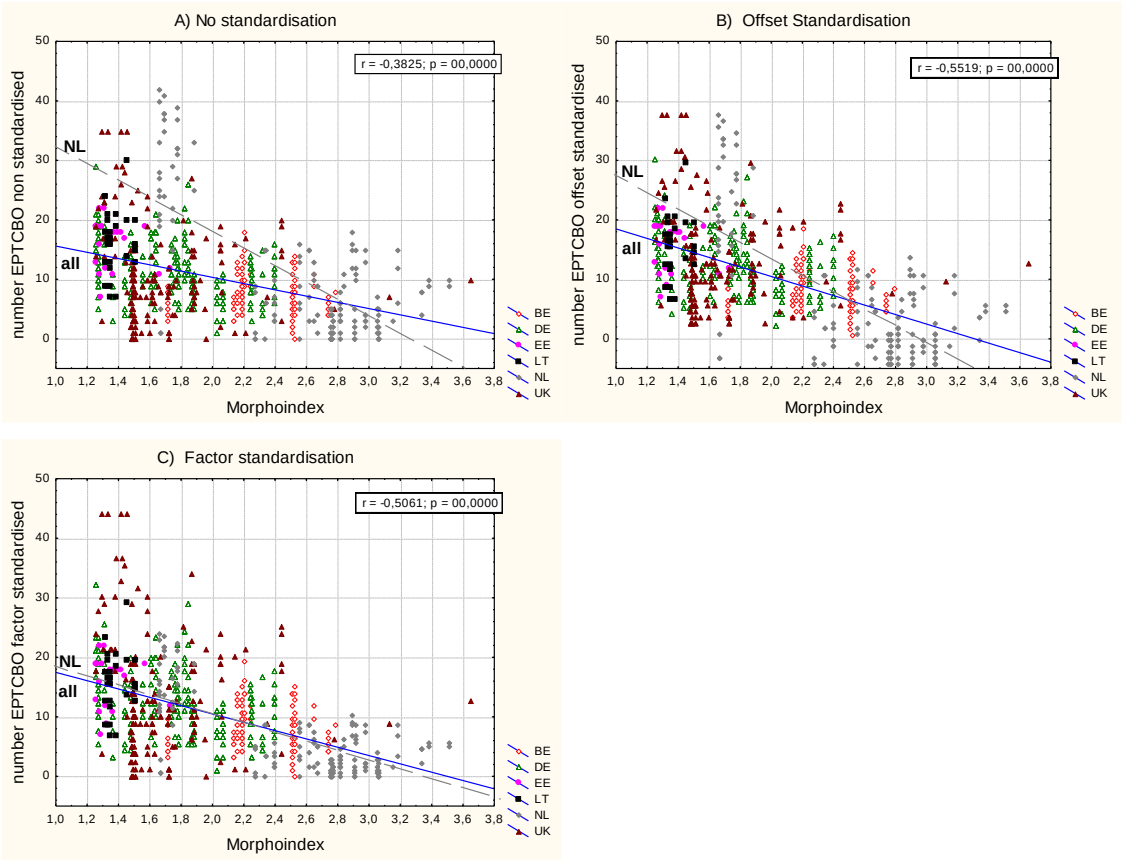


Figure 6.3 Effects of standardisation on common metrics with strongly deviating slopes of the dose-response-curves – Number of EPTCBO taxa as Example (A); note the deviating regression curve for NL in comparison for the regression for all countries together. Offset standardisation does not change the slopes, but leads to negative values (B). Factor standardisation gives the same corrections at the

intersections of the standardised metrics, but also changes the slopes (C). Consequently the dose response curves for the countries become more similar.

Theoretically, it would be better to standardise the values at the reference condition (= intersection of the regression curves at near zero pressure. But in reality the slope of the regression lines is too uncertain to give realistic results. This is mostly due to the small range which is covered by the countries' data (like the green line in Figure 6.2).

The standardisation results were controlled graphically. After standardisation the slopes were much more similar for the countries (like in Figure 6.2C).

The resulting standardisation values (offsets or factors) are given in

Table 6.2.

Table 6.2 Standardisation values derived by continuous benchmarking

Country	Factor no_EPTCBO	Offset ASPT	Offset %ETO	Offset %Lithal
BE	1,08	-0,45	-7,49	-5,14
DE	1,11	-0,04	-4,83	1,35
EE	1,00	-0,10	4,23	-1,84
LT	0,98	0,41	14,17	2,40
NL	0,57	-0,08	-4,92	-1,11
UK	1,26	0,26	-1,16	4,34

After combination of the standardized single metrics into a common multimetric index for boundary comparison, all countries followed one common dose response curve (Figure 6.4).

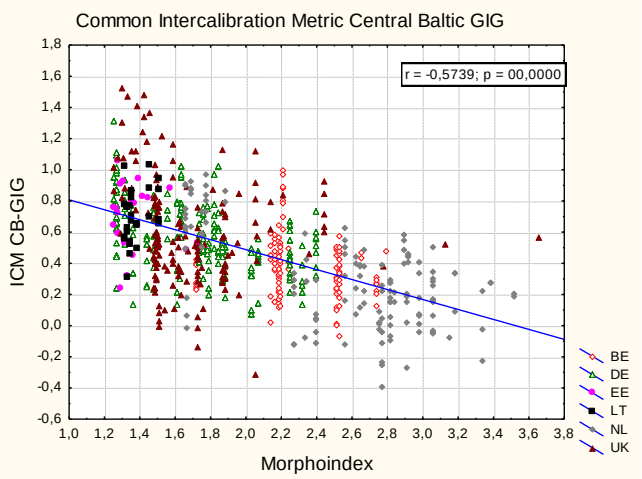


Figure 6.4 Correlation of the final intercalibration common metric (ICM) with the morphology index. The ICM is composed of the standardised single metrics. Therefore all countries follow the same dose response curve.

7. Comparison of methods and boundaries

7.1. IC Option and Common Metrics

Option 2 was selected, because the sampling and evaluation procedures of the methods are too different for option 1 and 3:

- The most important differences include differences in identification level of important taxonomic groups, differences in sampling season and differences of habitats covered;
- The UK CPET method is based on a completely different sampling procedure (collection of floating pupal exuviae instead of eulittoral handnet samples).

IC Common Metrics (see Annex B) was weighted average of normalised values of number of EPTCBO taxa, ASPT, % ETO, % Habitat preference lithal (all % in relation to abundance classes), thus covering all WFD criteria (diversity, sensitive/tolerant taxa, composition):

ICM = (2* no_EPTCBO + ASPT + %ETO +%lithal)/5

Metric were normalised using 10- and 90-%tiles of all metric values as anchors (see Table 7.1 below).

Table 7.1 Upper and lower anchors for the normalisation of the common metrics

Metric	Upper anchor (= value close to reference condition)	Lower anchor (= approximate median value at bad status)
no_EPTCBO	20.1	2.8
ASPT	5.5	3.6
%ETO	48.1	9.8
%Lithal	8.7	25.1

7.2. Results of the regression comparison

All correlations of each method with the common metric are highly significant (see Table 7.2.) The Pearson R is >0.5 for all except LT and UK, but no country was excluded. The slopes of the regression curves were all significantly different from 0 and all within the acceptable range (0.5-1.5).

Table 7.2 Correlation coefficient (r) and the probability (p) for the correlation of each method with the common metric.

Member State/Method		R	P	slope
BE-FL		0.56	<0.001	1.0
DE		0.63	<0.001	0.6
EE		0.63	0.009	1.0
LT		0.36	0.007	0.7
NL		0.70	<0.001	1.4
UK	ICM on site level	0.43	<0.001	1.1
	ICM on lake level	0.66	<0.001	1.1

The low R for LT was accepted, because the reason is a lack of pressure gradient (only undisturbed or very little disturbed lakes) and hence a lack of EQR variance. The low R for UK was accepted, because when the ICM-results are aggregated on lake level, the resulting Pearson R is 0.66, thus exceeding the requirement of 0.5 (because the benthic samples give an assessment for each sampling site whereas the UK CPET method is always one value per lake). Considering the fact that ICM is from littoral samples and EQR from CPET sampling, the correlation is amazingly high. Using this correlation, at least a satisfactory boundary comparison is possible instead of just declaring the method as incomparable.

7.3. Evaluation of comparability criteria

The comparison was carried out following the option 2 procedure:

- Common Metrics were standardised, normalised and combined into a multimetric common index (= IC common metric = ICM), expressed as EQR;
- National EQRs of status class boundaries were translated into ICM using regression lines of the ICM in dependence of the national EQR;
- ICM class boundary values were averaged to get a common view;
- The deviation of the countries was expressed in terms of the countries' status class width.

This procedure was carried out manually as well as using the Excel template provided ("IC_Opt2_sub.xlsx" - Intercalibration Excel Template Sheets v1.23).

Since the ICM had been already standardised by continuous benchmarking, no further benchmarking was used within the Excel template (manual offsets were set to 0).

Results of boundary comparison for H/G class boundaries (see Table 7.3):

- Too stringent for EE, LT, NL;
- Satisfactory for UK;
- Too relaxed for BE (1.07 classes outside tolerance) and DE (0.01 class width outside tolerance).

Results of boundary comparison for G/M class boundaries (see Table 7.3):

- Too stringent for EE, LT;
- Satisfactory for DE, NL, UK;
- Too relaxed for BE (0.91 classes outside tolerance),

Table 7.3 Boundaries and boundary bias of MS assessment systems (before system adjustment). Red cells – boundary too relaxed, yellow – boundary too precautionary

	BE	DE	EE	LT	NL	UK
H/G boundary	0.90	0.80	0.90	0.80	0.80	0.77
G/M boundary	0.70	0.60	0.70	0.60	0.60	0.64
H/G bias_CW	-1.322	-0.260	0.782	0.675	0.326	-0.238
G/M bias_CW	-1.162	0.112	0.278	0.897	-0.062	-0.027

Adjustments of boundaries and methods:

BE could not adjust its H/G-boundary by raising it from 0.9 to 1.15, because the maximal possible EQR for the method was 1.0. The high national EQR-values for the boundaries revealed problems with the BE reference values. Therefore the adjustment had to be done by revising the reference values instead of changing the boundary values. After the changes of the reference values (see Table 7.5), BE was within the harmonisation band with a bias of -0.125 for the G/M boundary and -0.033 class width for the H/G boundary (less than the tolerable bias of -0.25).

Table 7.4 a/b Original and adapted Belgian reference values and thresholds for metric scores, which were derived by equal division of the interval between reference condition and the lowest threshold. The lowest threshold was based on expert judgement. Values have to be exceeded for the higher scores.

a) Original Belgian reference values and thresholds for metric scores

	Reference value	Threshold between score 4 and 3	Threshold between score 3 and 2	Threshold between score 2 and 1	Threshold between score 1 and 0
TAX	33	26	19	12	5
EPT	6	4.5	3	1.5	0
OST	10	7.5	5	2.5	0
SWD	3.5	2.675	1.85	1.025	0.2
MTS	6	5	4	3	2

b) Adapted as result of the IC exercise:

	Reference value	Threshold between score 4 and 3	Threshold between score 3 and 2	Threshold between score 2 and 1	Threshold between score 1 and 0
TAX	54	41.75	29.5	17.25	5
EPT	12	9	6	3	0
OST	18	13.5	9	4.5	0
SWD	4	3.05	2.1	1.15	0.2
MTS	6.8	5.6	4.4	3.2	2

Adjustments for the other countries were optional (boundaries too stringent). Finally EE and LT decided to lower their boundaries to values just slightly above the harmonisation band:

- EE: HG boundary to 0.86, no change for GM boundary;
- LT: HG boundary to 0.74, GM boundary to 0.5.

Necessary adjustment for DE H/G-boundary is within the rounding error (National EQR 0,800 - 0.802).

The final boundary bias for all countries is listed in Table 7.5.

Table 7.5 Final boundary bias of the countries after the boundary adjustments

	BE	DE	EE	LT	NL	UK
H/G bias in class width	-0.033	-0.260	0.273	0.288	0.326	-0.238
G/M bias in class width	-0.125	0.112	0.347	0.331	-0.062	-0.027

The final boundaries are given below.

Table 7.6 H/G and G/M boundary EQR values for the national methods

MS	National classification systems intercalibrated	Ecological Quality Ratios	
		High-good boundary	Good-moderate
BE-FL	Multimetric Macroinvertebrate Index Flanders (MMIF)	0.90	0.70
DE	German Macroinvertebrate Lake Assessment (AESHNA)	0.80	0.60
EE	Estimation of Freshwater Quality Using Macroinvertebrates	0.86	0.70
LT	Lithuanian Lake Macroinvertebrate Index	0.74	0.5
NL	WFDi - Metric for Natural Watertypes	0.80	0.60
UK	Chironomid Pupal Exuvial Technique (CPET)	0.77	0.64

Gaps of the current Intercalibration: DK, FR, PL and LV could not be intercalibrated because they are still developing their assessment methods.

8. Description of IC type-specific biological communities

Good status is characterised by high diversity and abundance of sensitive insect taxa (mainly Ephemeroptera, Trichoptera and Odonata), a dominance of sensitive versus tolerant taxa (leading to a decrease in ASPT, for example), low ratios of r-strategists in relation to k-strategists and a low portion of indifferent taxa.

More than 200 out of the 1692 taxa of the the IC dataset showed preferences for high to good or moderate to bad status. Examples of frequently found taxa with higher abundances at high or good status include *Siphonoperla sp., Sericostoma sp., Leptocerus tineiformis, Gomphus vulgatissimus and Leptophlebia vespertina*.

Moderate or worse status is characterised by high diversity and abundance of insensitive taxa (mainly Crustaceans and many Chironomid taxa), a dominance of tolerant versus sensitive taxa, higher ratios of r-strategists in relation to k-strategists and a high portion of indifferent taxa.

Examples of frequently found taxa with higher abundances at moderate or worse status include *Corbicula sp., Mysidae Gen. sp., Physa sp., Glossiphonia sp., Chironomini Gen. sp.* and *Asellidae Gen. sp.* as well as some alien taxa like *Potamopyrgus sp., Dikerogammarus sp.* and *Corbicula sp.*

All in all, this reflects a change from more specialised and sensitive taxa towards generalist and tolerant taxa. All single metrics as well as the common multimetric index respond to the pressure in a more or less linear way, without certain changes indicative of class boundaries. Additionally abiotic factor combinations vary, leading to specific responses of each single metric as well as each taxon. Consequently no borderline communities can be described properly.

Description the biological communities at reference sites:

Reference status has the same characteristics as good status, but with a gradual difference: high diversity and abundance of sensitive insect taxa (mainly Ephemeroptera, Trichoptera and Odonata leading to a number of EPTCBO >20 and a proportion of ETO taxa >50%), a dominance of sensitive versus tolerant taxa (ASPT >5.5, for example), and a low portion of the habitat preference lithal (<9%).

Examples of frequently found taxa with highest abundances at reference status include: *Sericostoma sp., Siphonoperla sp., Triaenodes sp., Platycnemis pennipes, Limnephilus lunatus, Cordulia aenea, Lepidostoma hirtum and Leptocerus tineiformis*.

Pressure response relationships were also analysed for the combined IC dataset:

- In general both, the relative abundance and the number of sensitive taxa, decreased with increasing pressure;

- Diversity indices and taxa number also decreased, but less pronounced;
- Taxonomic composition changed from sensitive insect taxa (Ephemeroptera, Trichoptera, and Odonata) to tolerant insects (mainly tolerant chironomids) and non-insects (e.g. crustaceans and tolerant molluscs).

Annexes

A. Lake Benthic Fauna classification systems of Member States

A.1 Belgium-Flanders: Multimetric Macroinvertebrate Index Flanders (MMIF) for assessment of surface water bodies compliant to the European Water Framework Directive

Introduction

A first version of the Multimetric Macroinvertebrate Index Flanders (MMIF) for assessment of rivers and lakes compliant to the European Water Framework Directive (WFD; EU, 2000) was described by Gabriels et al. (2010). Specifically for lakes, a number of adaptations were later introduced to the calculation of the metrics, in order to comply to the results of the European intercalibration exercises. Because these adaptations were not yet incorporated in the description of Gabriels et al. (2010), the present paper provides a complete overview of the MMIF method in lakes compliant to the intercalibration exercises.

Lake types

In Flanders, four general lake types are distinguished. These types are further assigned to several sub-types, but all sub-types within a general type are assessed in the same way for macroinvertebrates. Therefore, only the general types are discussed in the present paper.

The four general lake types are summarised in Table A.1.

Table A.1 General lake types in Flanders

Symbol	Type	Properties
A	Alkaline	pH $\geq$ 7.5
C	Circumneutral	6.5 $\leq$ pH < 7.5; no clay
Z	Acidic	pH < 6.5; only sand/sandy loam/loam
Bzl	Very slightly brackish	Na > 250 mg/l; no sand/sandy loam/loam

All large lakes in Flanders (with surface area > 50 ha) which are relevant for reporting under the WFD, belong to the general type A. These were subject to the intercalibration exercises.

Selection of sampling sites

At least one representative sampling station is selected within the water body. Depending on the heterogeneity of the water body, up to three sampling stations are selected within the water body.

Sampling

The samplings should be carried out during spring, summer or autumn. It is recommended to avoid sampling macroinvertebrates during winter in order to avoid

extreme conditions, both of hydrological regime and temperature, to ensure a reliable water quality assessment.

Macroinvertebrates are sampled using a standard handnet, as described by De Pauw and Vanhooren (1983) and NBN (1984). This handnet consists of a metal frame of approximately 0.2 m by 0.3 m to which a conical net is attached with a mesh size of minimum 300 and maximum 500 mm. The frame is attached to a 2 m long shaft with two handles enabling it to be handled in a similar way as a scythe. With the handnet, a stretch of approximately 10–20 m is sampled during 3 minutes for watercourses less than 2 m wide or up to 5 minutes for larger rivers. Sampling effort is proportionally distributed over all accessible aquatic habitats. This includes the bed substrate (stones, sand or mud), macrophytes (floating, submerged, emerged), immersed roots of overhanging trees and all other natural or artificial substrates, floating or submerged in the water. Each aquatic habitat is explored, either with the handnet or manually, in order to collect the highest possible diversity of macroinvertebrates. For this purpose, kicksampling is performed by vertically positioning the handnet on the bed and turning over bottom material located immediately upstream by foot or hand. In addition to the handnet sampling, animals are manually picked from stones, leaves or branches along the same stretch (De Pauw and Vanhooren, 1983). For lakes, macroinvertebrates are sampled using the same method, distributing the sampling effort proportionally over all accessible aquatic habitats within a stretch of 10–20 m.

If a site is too deep to be sampled with the handnet method, macroinvertebrates can alternatively be sampled using the so-called Belgian artificial substrates as described by De Pauw et al. (1986) and De Pauw et al. (1994). These substrates are composed of a plastic netting filled with medium-sized (4–8 cm) pieces of brick, with a total volume of approximately 5 litre. Per sampling site, three substrates are placed in the water, anchored with a rope to a fixed point located on the bank. The substrates should not be placed in open water but along the banks: in protected sites among the vegetation near the surface, in unprotected sites, which are exposed to surface turbulence, in deeper water. After an exposure time of at least three weeks, the substrates are lifted from the water and transferred into a closed container (De Pauw et al., 1986).

Identification

The used identification levels are:

- Hydracarina: presence;
- Oligochaeta, Crustacea, Coleoptera, Trichoptera, Diptera (except Chironomidae): family;
- Chironomidae: groups thummi-plumosus and non thummi-plumosus;
- Plathelminthes, Hirudinea, Mollusca, Ephemeroptera, Odonata, Plecoptera, Hemiptera, Megaloptera: genus.

A standard list with all taxa included in index calculation can be found in Table A.4.

The recommended identification key is De Pauw and Vannevel (1991), except for Ampharetidae, Janiridae, Sphaeromatidae, *Corbicula*, which are not included in the cited work, and *Physa* and *Physella*, which are not distinguished in this work.

For all observed taxa, abundances are recorded. For very abundant taxa (>10 individuals) the abundance may be estimated instead of counted.

Index calculation

Calculation of metrics

Metric taxa richness

The metric taxa richness is calculated as the total number of taxa (according to the specified levels of identification) of which one or more individuals were found in the sample.

Metric number of EPT taxa

The metric number of EPT taxa is calculated as the total number of taxa (according to the specified levels of identification) belonging to Ephemeroptera, Plecoptera and/or Trichoptera of which one or more individuals were found in the sample.

Metric number of other sensitive taxa

The metric number of other sensitive taxa is calculated as the total number of taxa (according to the specified levels of identification), other than the EPT taxa, with a tolerance score of six or more. The list of tolerance scores (ranging from 10 for very intolerant to 1 for very tolerant) for all taxa is given in Gabriels et al. (2009).

Metric Shannon-Wiener Index

The metric Shannon-Wiener Index is calculated using the following formula (Shannon & Weaver, 1949):

$H' = - \Sigma s\ i=1\ [\ pi\ .\ ln\ pi\ ]$

With: S = the taxa richness;

pi = the relative abundance of the i-th taxon.

When no taxa are encountered in a sample at all, this metric is set equal to zero.

Metric Mean Tolerance Score

The metric mean tolerance score is calculated as the sum of the tolerance scores of taxa of which one or more individuals were found in the sample, divided by the total number of taxa. The list of tolerance scores (ranging from 10 for very sensitive to 1 for very tolerant) for all taxa can be found in Table A.4. When no taxa are encountered in a sample at all, this metric is set equal to zero.

Total index calculation

Calculation of the index for one sampling station

In order to integrate the values of the five metrics into one index, they are first each converted into a score of 0 to 4. For each lake type, criteria are set for each metric by which the value can be converted into the corresponding score. These criteria are summarized per lake type in Table A.2.

Table A.2 Scoring criteria for calculating the Multimetric Macroinvertebrate Index Flanders for all lake types in Flanders. Columns and rows respectively correspond to the lake types and the scores that are assigned based on the respective metric values.

Lake type	Type A	Type C	Type Z	Type Bzl
Score	**Metric values – taxa richness**			
0	≤ 5	≤ 5	≤ 5	≤ 5
1	≤ 17.25	≤ 12.5	≤ 10.75	≤ 11.25
2	≤ 29.5	≤ 20	≤ 16.5	≤ 17.5
3	≤ 41.75	≤ 27.5	≤ 22.25	≤ 23.75
4	> 41.75	> 27.5	> 22.25	> 23.75
Score	**Metric values – number of EPT taxa**			
0	0	0	0	0
1	≤ 3	≤ 2	≤ 1.25	≤ 1.25
2	≤ 6	≤ 4	≤ 2.5	≤ 2.5
3	≤ 9	≤ 6	≤ 3.75	≤ 3.75
4	> 9	> 6	> 3.75	> 3.75
Score	**Metric values – number of other sensitive taxa**			
0	0	0	0	0
1	≤ 4.5	≤ 2.5	≤ 2	≤ 2.25
2	≤ 9	≤ 5	≤ 4	≤ 4.5
3	≤ 13.5	≤ 7.5	≤ 6	≤ 6.75
4	> 13.5	> 7.5	> 6	> 6.75
Score	**Metric values – Shannon-Wiener index**			
0	≤ 0.2	≤ 0.2	≤ 0.2	≤ 0.2
1	≤ 1.15	≤ 1.025	≤ 0.9	≤ 0.95
2	≤ 2.1	≤ 1.85	≤ 1.6	≤ 1.7
3	≤ 3.05	≤ 2.675	≤ 2.3	≤ 2.45
4	> 3.05	> 2.675	> 2.3	> 2.45
Score	**Metric values – mean Tolerance Score**			
0	≤ 2	≤ 2	≤ 2	≤ 2
1	≤ 3.2	≤ 3	≤ 3	≤ 3
2	≤ 4.4	≤ 4	≤ 4	≤ 4
3	≤ 5.6	≤ 5	≤ 5	≤ 5
4	> 5.6	> 5	> 5	> 5

The overall index for a sampling station equals the sum of the five metric scores, which is a number between 0 and 20, divided by 20. This results in an EQR value that is comprised within the interval 0-1.

Determination of the index for the whole water body

The index value for the whole lake is the average of the index values of the different representative sampling points.

Determination of the quality class

The criteria used for determining the quality classes are summarized in Table A.3.

Table A.3 Class boundaries for the macroinvertebrate index for lakes

MMIF	Class	Colour code
≥ 0,90	High	Blue
< 0,90 and ≥ 0,70	Good	Green
< 0,70 and ≥ 0,50	Moderate	Yellow
< 0,50 and ≥ 0,30	Poor	Orange
< 0,30	Bad	Red

References

De Pauw, N., Lambert, V., Van Kenhove, A. & Bij de Vaate, A. (1994). Performance of two artificial substrate samplers for macroinvertebrates in biological monitoring of large and deep rivers and canals in Belgium and The Netherlands. Environmental Monitoring and Assessment 30(1): 25-47.

De Pauw, N., Roels, D. & Fontoura, P. (1986). Use of artificial substrates for standardized sampling of macroinvertebrates in the assessment of water quality by means of the Belgian Biotic Index. Hydrobiologia 133(3): 237-258.

De Pauw, N. & Vanhooren, G. (1983). Method for biological quality assessment of watercourses in Belgium. Hydrobiologia 100(1): 153-168.

De Pauw, N. & Vannevel, R. (eds.) (1991). Macro-invertebraten en waterkwaliteit. Stichting Leefmilieu, Antwerp, Belgium. 316 p.

EU (2000). Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. Official Journal of the European Communities L327: 1-72.

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.

NBN (1984). Biologische kwaliteit van de waterlopen. Bepaling van de Biotische Index steunende op aquatische macro-invertebraten. NBN T92-402. Belgian Institute for Normalisation, Brussels, Belgium.

Table A.4 Taxa taken into account for calculating the Multimetric Macroinvertebrate Index Flanders, and their respective tolerance scores

Taxon	Tolerance score	Taxon	Tolerance score
Plathelminthes		Dytiscidae	5
Bdellocephala	5	Elminthidae	7
Crenobia	7	Gyrinidae	7
Dendrocoelum	5	Haliplidae	6
Dugesia s.l.	5	Hydraenidae	6
Phagocata	5	Hydrophilidae	5
Planaria	6	Hygrobiidae	5
Polycelis	6	Noteridae	5
Polychaeta		Psephenidae	6
Ampharetidae	3	Scirtidae	7
Oligochaeta		Hemiptera	
Aelosomatidae	2	*Aphelocheirus*	8
Branchiobdellidae	2	*Arctocorisa*	5
Enchytraeidae	2	*Callicorixa*	5
Haplotaxidae	4	*Corixa*	5
Lumbricidae	2	*Cymatia*	6
Lumbriculidae	2	*Gerris* s.l.	6
Naididae s.s.	5	*Glaenocorisa*	5
Tubificidae	1	*Hebrus*	6
Hirudinea		*Hesperocorixa*	5
Cystobranchus	4	*Hydrometra*	6
Dina	4	*Ilyocoris*	5
Erpobdella	3	*Mesovelia*	6
Glossiphonia	4	*Micronecta*	6
Haementeria	4	*Microvelia*	7
Haemopis	4	*Naucoris*	6
Helobdella	4	*Nepa*	6
Hemiclepsis	4	*Notonecta*	5
Hirudo	4	*Paracorixa*	5
Piscicola	5	*Plea*	6
Theromyzon	4	*Ranatra*	6
Trocheta	4	*Sigara*	5
Mollusca		*Velia*	7
Acroloxus	6	Odonata	
Ancylus	7	*Aeshna*	6
Anisus	5	*Anax*	6

Taxon	Tolerance score
Anodonta	6
Aplexa	6
Armiger	6
Bathyomphalus	5
Bithynia	5
Bythinella	8
Corbicula	5
Dreissena	5
Ferrissia	7
Gyraulus	6
Hippeutis	6
Lithoglyphus	6
Lymnaea s.l.	5
Margaritifera	10
Marstoniopsis	5
Menetus	5
Myxas	7
Physa s.s.	5
Physella	3
Pisidium	4
Planorbarius	5
Planorbis	6
Potamopyrgus	6
Pseudamnicola s.l.	5
Pseudanodonta	6
Segmentina	6
Sphaerium	4
Theodoxus	7
Unio	6
Valvata	6
Viviparus	6
Acari	
Hydracarina s.l.	5
Crustacea	
Argulidae	5
Asellidae	4
Astacidae	8
Atyidae	7
Cambaridae	6

Taxon	Tolerance score
Brachytron	7
Calopteryx	8
Cercion	7
Ceriagrion	7
Coenagrion	6
Cordulegaster	9
Cordulia	7
Crocothemis	7
Enallagma	7
Epitheca	7
Erythromma s.s.	7
Gomphus	7
Ischnura	6
Lestes	7
Leucorrhinia	7
Libellula	7
Nehalennia	7
Onychogomphus	7
Ophiogomphus	7
Orthetrum	7
Oxygastra	7
Platycnemis	7
Pyrrhosoma	7
Somatochlora	7
Sympecma	7
Sympetrum	7
Ephemeroptera	
Baetis	6
Brachycercus	7
Caenis	6
Centroptilum	7
Cloeon	6
Ecdyonurus	9
Epeorus	10
Ephemera	8
Ephemerella s.l.	8
Ephoron	9
Habroleptoides	8
Habrophlebia	8

Taxon	Tolerance score
Chirocephalidae	6
Corophiidae	5
Crangonyctidae	4
Gammaridae	5
Janiridae	5
Leptestheriidae	6
Limnadiidae	6
Mysidae	5
Palaemonidae	5
Panopeidae	4
Sphaeromatidae	4
Talitridae	5
Triopsidae	6
Varunidae	4
Diptera	
Athericidae	7
Blephariceridae	7
Ceratopogonidae	3
Chaoboridae	3
Chironomidae:	
-group non *thummi-plumosus*	3
-group *thummi-plumosus*	2
Culicidae	3
Cylindrotomidae	3
Dixidae	6
Dolichopodidae	3
Empididae	3
Ephydridae	3
Limoniidae	4
Muscidae	3
Psychodidae	3
Ptychopteridae	3
Rhagionidae	3
Scatophagidae	3
Sciomyzidae	3
Simuliidae	5
Stratiomyidae	4

Taxon	Tolerance score
Heptagenia s.l.	10
Isonychia	7
Leptophlebia s.s.	8
Metreletus	7
Oligoneuriella	7
Paraleptophlebia	8
Potamanthus	8
Procloeon	7
Rhitrogena	10
Siphlonurus	7
Trichoptera	
Beraeidae	9
Brachycentridae	9
Ecnomidae	6
Glossosomatidae	9
Goeridae	9
Hydropsychidae	6
Hydroptilidae	8
Lepidostomatidae	9
Leptoceridae	8
Limnephilidae s.l.	8
Molannidae	9
Odontoceridae	9
Philopotamidae	6
Phryganeidae	9
Polycentropodidae	6
Psychomyiidae	7
Rhyacophilidae	8
Sericostomatidae	8
Plecoptera	
Amphinemura	9
Brachyptera	10
Capnia s.l.	10
Chloroperla s.l.	10
Dinocras	10
Isogenus	10
Isoperla	10

Taxon	Tolerance score	Taxon	Tolerance score
Syrphidae	1	*Leuctra*	9
Tabanidae	3	*Marthamea*	10
Thaumaleidae	3	*Nemoura*	8
Tipulidae	3	*Nemurella*	8
Megaloptera		*Perla*	10
Sialis	5	*Perlodes*	10
Coleoptera		*Protonemura*	9
Dryopidae	6	*Rhabdiopteryx*	10
		Taeniopteryx	10

s.l.: The notation 's.l.' (*sensu lato*) was added to those taxa that comprise one or more additional taxa than the taxon itself. These taxa were considered as such by De Pauw and Vannevel (1991), but due to new taxonomic insights they were split up into two or more taxa. For example, the genus *Lymnaea*, *Stagnicola*, *Radix* and *Galba*, were, at the time of publication of De Pauw and Vannevel (1991), all considered as *Lymnaea*. This lumping is maintained here and therefore the notation *Lymnaea* s.l. should be interpreted here as *Lymnaea* sensu De Pauw and Vannevel, in this case *Lymnaea* + *Stagnicola* + *Radix* + *Galba*.

A.2 Estonia: A multimetric index to assess quality of lake littoral on the basis of macroinvertebrates [Järvede litoraali seisundi hindamise liitindeks suurselgrootute järgi]

General information

Detected pressure(s)

Eutrophication, General degradation, Hydromorphological degradation, Riparian habitat alteration

Pressure-impact-relationship*:*

Ecological data from 20 LCB1 and LCB2 lakes were examined to establish pressure-impact relationship between macroinvertebrate metrics, eutrophication and green land use in catchment area. There occurred significant correlations between multimetric quality (on the basis of 5 indices) and the following parameters: total phosphorus content in water (r=-0.32), water quality on the basis of phytoplankton (0.58), and green land cover (0.64).

Internet reference: https://www.riigiteataja.ee/ert/act.jsp?id=13210253&replstring=33 (in Estonian only)

Pertinent literature of mandatory character:

Pinnaveekogumite moodustamise kord ja nende pinnaveekogumite nimestik, mille seisundiklass tuleb määrata, pinnaveekogumite seisundiklassid ja seisundiklassidele vastavad kvaliteedinäitajate väärtused ning seisundiklasside määramise kord.

https://www.riigiteataja.ee/akt/125112010015

Scientific literature: Timm, H., 2003. Typology and classification of freshwaters in Estonia: preliminary results using shallow-water macroinvertebrates. In Ruoppa, M., P. Heinonen, A. Pilke, S. Rekolainen, H. Toivonen & H. Vuoristo (eds), How to assess and monitor ecological quality in freshwaters. TemaNord 547: 164-169.

Timm, H., 2005. Benthic invertebrates as a tool to classify ecological status of inland waters. Estonian experiences. In Lääne, A. & P. Heinonen (eds), Presentations of three training seminars about Quality Assurance (QA), Biological methods of Water Framework Directive and Waste water sampling techniques. Suomen ympäristökeskuksen moniste, Helsinki 328: 89-94.

Timm, H. & E. Mälton, 2006. Littoral macroinvertebrates in large lakes: can they tell us something about the status of lake? - European Large Lakes Symposium 2006. Ecosystem changes and their ecological and socioeconomic impacts. Programme and abstracts. Tartu, Estonia: 54-55.

Timm, H., K. Mardi & T. Möls, 2008. Macroinvertebrates in Estonian streams: the effect of habitat, season and sampling effort on some common metrics of biological quality. Estonian Journal of Ecology 57 (1): 37-57.

Timm, H. & T. Möls, 2008. Do shallow-water macroinvertebrate assemblages correspond to physico-chemical habitats of streams and lakes? Verh. Internat. Verein. Limnol 31 (1): 138-140.

Wasson, J.-G., B. Villeneuve, A. Iital, J. Murray-Bligh, M. Dobiasova, S. Bacikova, H. Timm, H. Pella, N. Mengin & A. Chandesris. Large-scale relationships between basin and riparian land cover and ecological status of European rivers: examples with invertebrate indices from France, Estonia, Slovakia and United Kingdom. Freshwater Biology (accepted).

Data acquisition

Field sampling/surveying

From the most typical bottom at the sampling site, five 1-m long kick- or sweep replicates are taken and kept separately. A separate qualitative sample is collected from all available substrates, not considering their area. Standard handnet with 25 cm edge length, 0.5 mm mesh size is used, all available habitats per site (Multi-habitat) are inspected.

Sampling time: April - May or September-October, one occasion per sampling season

Total sampled/surveyed area or volume or total sampling duration to classify site or area: 1.25 m^2 + qualitative sample (5-10 min.)

Sample treatment: Subsampling is used only to estimate the abundance of dominants. The decision which is a dominant is made separately for each replication. No subsampling is used for qualitative sample. Organisms of the complete sample are identified.

Level of taxonomical identification: Chironomidae, Oligochaeta, Hydrachnidia, Pisidium on group level, other taxa to species where possible.

Unit of the record of abundance: Number of individuals per one square-metre

Evaluation

List of biological metrics: Total taxon richness, Ephemeroptera, Plecoptera and Trichoptera taxon richness, Shannon diversity, British Average Score Per Taxon, Swedish Acidity Index

Combination rule for multi-metrics: Average metric scores

Reference conditions

Approach to derive reference conditions: Existing near-natural reference sites, expert knowledge

Reference site characterisation: 50 sites from 50 Estonian lakes of LCB1 or LCB2 types Reference criteria. Water pollution and shoreline modification are absent, catchment area without significant stress factors.

Boundary setting

All index values of high quality were assigned 5 points, values of good quality – 4 points, values of moderate quality - 2 points, and values of poor and bad quality - 0 points. The difference between good level and moderate level was intentionally emphasized in order to underline the principal difference between them in terms of the Water Framework Directive. Multimetric quality (MMQ) was then calculated by adding up the corresponding points. Hence, for small lakes:

- reference value was 25
- sum 23−25 was considered to indicate high;
- 18−22 - good;
- 10−17 - moderate;
- 6−9 - poor;
- <6 - bad quality.

A.3 Lithuania

Lithuanian lake eulittoral macroinvertebrate assessment system (Lithuanian eulittoral macroinvertebrate system) [Lietuvos ežerų eulitoralės makrobestuburių vertinimo sistema (Lietuvos eulitoralės makrobestuburių sistema)]

Data acquisition

Field sampling/surveying

Standard method: 12 kick or sweep replicates from different microhabitats;

Additional survey: Semi-quantitative sampling procedure is carried out using a standard dip-net (25x25 cm). Sampling can be performed in either of the two core eulittoral mesohabitats: a bottom (preferably hard) kick sample or a vegetation (preferably submerged) sweep sample. Within a stand of either mesohabitat, a stretch of about 15-20 meters long is sampled while moving along the shore in a trajectory of a zigzag curve (from the very shoreline to the depth of 1 m) in a way to result in 3 minutes of actual catching time. A semi-quantitative sample is supported by qualitative (search) sample (duration 1 minute) within the same mesohabitat.

Hand net (frame 25X25 cm>sub>2) is used, all available habitats per site (Multi-habitat)

Sampling time : April to November, one occasion per sampling season and per lake

Number of spatial replicates per sampling/survey occasion to classify site or area: Standard sampling: 12 replicates in proportion to microhabitat coverage; additional survey: 1 replicate per lake (15-20 m stretch within a stand of either of the two mesohabitats)

Total sampled/surveyed area or volume or total sampling duration to classify site or area: 40 x 25 cm2 = 0.1 m2 x 12 = 1.2 m2 per site, plus qualitative collection sample: 3 minutes from each mesohabitat

Standard sampling: sum of 12 spatial replicates 0.1 = 1.2 square-meters; additional survey: 3 minutes from each mesohabitat

Specification of level of determination: Chrironomidae to subfamily level, other dipterans to family level, Oligochaeta to class level, Coleoptera - genus, Hydrachnidia – ignored, other animals to species (if not possible, to genus) level

Data evaluation

List of biological metrics:

1. Hill's number (exp(-Sum[1..TR](p[i]*ln(p[i]))), where TR - taxa richness, p - relative abundance of taxon i);
2. Average Score Per Taxon ASPT;
3. Number of Ephemeroptera, Plecoptera, Coleoptera taxa;
4. Percentage of Odonata, Plecoptera and Coleoptera individuals in respect of a total number of individuals;

Combination rule for multi-metrics: Average of EQRs of all comprising metrics

Reference conditions

Key sources to derive reference conditions**:** Existing near-natural reference sites, expert knowledge; Least Disturbed Conditions. Reference values for the metrics were derived as 90% percentiles of metric values of samples from reference sites.

Reference Criteria: Open eulittoral sites of lakes satisfying all of these criteria: 1) annual average TP < 0.04 mg/L; 2) annual average TN < 1.3 mg/L; 3) average for the vegetation

period of chlorophyll a < 5 ug/L; 4) positive expert judgement (natural surroundings, no pollution sources, no substantial hydromorphological alteration and no heavy biocontamination)

Reference community description: Macroinvertebrate communities of mesotrophic lakes, showing high diversity, presence of many pollution sensitive taxa, high richness of pooled Ephemeroptera, Plecoptera or Coleoptera orders, and large numbers of pooled individuals that belong to Odonata, Plecoptera or Coleoptera groups.

Setting of ecological status boundaries: No relation to the pressure has been used to set the boundaries. Lakes are classified using equidistant division of the EQR gradient of multimetric macroinvertebrate index with boundaries at 0.8, 0.6, 0.4, 0.2.

"Good status" community: Macroinvertebrate communities of mesotrophic lakes, showing high diversity, presence of many pollution sensitive taxa, high richness of pooled Ephemeroptera, Plecoptera or Coleoptera orders, and large numbers of pooled individuals that belong to Odonata, Plecoptera or Coleoptera groups.

A.4 Netherlands: WFD-metrics for natural watertypes [KRW-maatlatten voor natuurlijke watertypen]

General information

Detected pressures: Eutrophication, general degradation, hydromorphological degradation, pollution by organic matter

Specification of pressure-impact-relationship: The metric for invertebrates in lakes is validated for chemical and hydromorphological pressures (n = 32 lakes, 113 samples). High nutrient concentrations limited the metric score, but low nutrient concentration does not automatically result in a high metric score. Additionally, a distinct relation between hydromorphological shore alteration and EQR was observed (r=0.67)

Internet reference:

http://themas.stowa.nl/thema/ecologische_beoordeling/krw-maatlatten.aspx?mId=7213&rId=817

Pertinent literature of mandatory character: Besluit Kwaliteitseisen en Monitoring Water (2009). Ministry of Housing, Spatial Planning and the Environment (presently under public consultation).

Data acquisition

Field sampling/surveying

Multihabitat sampling in all habitats present in proportion to their presence. Active moving of handnet through vegetation and bottom substrates.

Sampling/survey device: Handnet 30x15 cm, Van Veen or Eckman Birge grab, box-corer

Sampled habitat: All available habitats per site (Multi-habitat)

Sampling/survey time: March until 15 June, minimum one occasion per year (spring), but classification preferably averaged over three years.

Sample treatment

Organisms of the complete sample are identified. only if some organisms occur in extreme high number, subsampling is done and total number is estimated.

Specification of level of determination: Oligochaetes and Hydracarina may sometimes be determined to genus/family level., other- to species level

Data evaluation

List of biological metrics: EQR=[200 x (KM%/KMmax) + (100-DN%) + (KM%+DP%)]/400 where KM% = relative number of typical (for watertype) species in a sample KMmax - maximum achievable number of typical species under reference conditions %DN = relative abundance of dominant negative species %(DP+KM) = sum of relative abundances of dominant positive species and typical species Abundances are converted first to abundance (log) classes The metric for invertebrates in lakes is based on the littoral zone and not the pelagic or benthic zone.

Reference conditions

Key source(s) to derive reference conditions: Expert knowledge, Historical data, Least Disturbed Conditions, as no actual existing natural sites in lakes

All lakes in The Netherlands are (very) high hydromorphological impacted, level fluctuation is completely controlled (less than 5 cm) and most of them are moderately to highly impacted by eutrophication. Too few lakes are assumed to meet the criteria of (almost) unimpacted.

Reference community description:

Regarding the metric: High status of lakes is characterized by a high abundance of dominant positive species and a high diversity and abundance of typical species. Dominant negative species are nearly absent. Furthermore a general description is given (in Dutch) in: STOWA (2009) Referenties en maatlatten voor natuurlijke watertypen. report 2007-32

Boundary setting procedure:

- The boundaries for the different EQR-classes (bad, poor, moderate, good and high) are set based on expert judgement and follow a more or less equal division of quality.
- The WFDi and its class-boundaries were validated by experts judging species lists from anonymous sites, using normative definitions. Validation was done based on existing data on shallow lakes from the Netherlands (Naardermeer, Randmeren, Vollenhovermeer and Wijchens Ven).
- In the validation of the method the response of the WFD-classes to pressures was tested. WFD-classes responded negatively to hydromorphological pressure. Of the chemical pressures studied, EQR is

most related to oxygen content. EQR and oxygen availability are positively correlated. Influences of other chemical pressures considered (phosphate and nitrogen content) were less clear. Water bodies in the Netherlands are hydromorphologically altered, making physical pressure an important factor in assessment of Dutch water bodies.

"Good status" community:

Good status is characterized by a high diversity and abundance of typical species and an increasing abundance of dominant positive species. The abundance of dominant negative species is low.

Uncertainty

Precision and uncertainty is regarded in Van Herpen, van Tongeren, Knoben, Baggelaar, van Loon (2009). Quick scan precision and confidence of KRW assessment (in Dutch). This study resulted in a statistical method to assess the level of precision and confidence monitoring results and status classifications (including identifying outliers and estimates for missing values). The confidence of a status classification is expressed as the probability of exceeding a chemical limit value or the biological status classification moderate/good. Recommendations from this study are incorporated in the Instructie; Richtlijn Monitoring Oppervlaktewater en Protocol Toetsen & Beoordelen (28 april 2009) (see question B.0). In the metric abundance is expressed in abundance classes to reduce the impact of extreme abundance of one species on the calculated EQR.

A.5 AESHNA - German Lake Macroinvertebrate Assessment System for the Water Framework Directive

The ecological classification system for eulittoral macroinvertebrates will be used in the second RBMP (2015). The sublittoral classification will be used for additional information only.

Lake types

The method is suited to all natural alpine lakes and lowland lakes in Germany. These include all intercalibration lake types occurring in Germany (L-CB1, L-CB2, L-AL3, L-AL4). Within the national typology these comprise the general types 1, 2, 3, 4, 10, 11, 12, 13 and 14, or the benthic fauna lake types small alpine, large alpine, riverine lowland and non-riverine lowland respectively.

AESHNA also includes multimetric assessment indices for other natural and artificial German lake types > 50 ha. But since these have not been officially accepted yet, they were not included here.

Detected pressures

AESHNA was designed for the detection of all kinds of pressures, but the focus was laid on hydromorphological degradation.

Based on **eulittoral** macroinvertebrate samples of 491 central-baltic sampling sites (55 lakes) and 131 alpine sampling sites (12 lakes) pressure-impact and lake morphology data relationships were established for a variety of candidate metrics. Finally multimetric indices were developed consisting of several metrics, which cover all criteria of the WFD. Multimetric indices and lake morphology indices were significantly correlated (Spearman R ranging from 0.5 to 0.8, depending on the pressure index and lake type). Correlations with a combined morphology-TP index were slightly higher, whereas correlations with eutrophication alone or with catchment landuse were significantly lower with Spearman R up to 0.5.

For the **sublittoral** assessment however Spearman R is similar for all landuse in lake surroundings, catchment landuse and eutrophication with values up to 0.6.

Sampling

Short description of eulittoral sampling

A multihabitat sampling procedure is carried out for eulittoral macroinvertebrates in February to April (lowland) / to May (alpine) or September to October.

A minimum of 4 sampling sites per lake (N=4+shorelength^-2) is selected by expert judgement according to the occurrence of shoreline types.

At each sampling site all available habitats have to be covered at up to 1.2 m depth of water. Hand nets (500 µm mesh-size) are used whenever suitable or other devices when more appropriate (e.g. scrapers).

There are two options:

A) The area sampled for each habitat is proportional to the percentage of occurrence at the sampling site. At minimum total of 1 m^2 is sampled.

B) All habitats are sampled with the same intensity, covering 0.6 to 1,0 m^2 per habitat. The area sampled and the relative presence of each habitat is determined for a later combination to a multihabitat taxa list.

Short description of sublittoral sampling

Sublittoral sampling is carried out once in February to April (lowland) / to May (alpine) or September to October.

A minimum of 8 stations per lake (>=12 for lakes >200 ha) is selected by dividing the lake in equal sectors and placing them in the center of the upper sublittoral zone. At each sampling site 3 Ekman grabs are taken.

Sample processing

Samples are sorted out in the field or sieved, fractionated and preserved in Ethanol for sorting in the laboratory. Sublittoral samples may be subsampled.

Level of taxonomical identification

Taxa to be identified are given in a detailed "operational taxa list". The level is mostly species or achievable level for all but the following: Family for oligochaeta and most non-chironomid dipterans, mostly genus for chironomids.

Abundances are recorded as number of individuals per m^2.

Multimetric index / EQR calculation

Multimetric index (MMI) composition and standardisation values differ between benthic fauna lake types. In order to obtain EQR values comparable to other biological quality elements the EQR-values are obtained from the MMI values by linear transformations.

Metric standardisation

Two anchor points for metric standardisation were derived from the data distribution along the pressure gradient using 10%tiles of the whole distribution in combination with extrapolated values (for incomplete pressure gradients): The near reference value and the bad status value. Using the following formula the each metric value (M) is standardised from 0.0 for the bad status value (M0) to 1.0 for the near reference value (M1):

Standardised metric = (M - M0) / (M1 –M0).

The anchor values are specific for each benthic fauna lake type.

Eulittoral alpine MMI/EQR

Five standardised metrics are averaged: relative abundance of Odonata (% of abundance classes), relative abundance of feeding type collectors (% of abundance classes,

reproduction strategy r/k, Shannon diversity and littoral faunaindex, with double weighting of the fauna index;

Formula: MMI = (2*fauna index + odonata+ Shannon diversity + gatherer + rk)/6

EQR = MMI*4/3-1.2

Eulittoral non-riverine lowland MMI/EQR

Four standardised metrics are averaged with equal weighting:

Faunaindex, relative abundance of habitat type lithal (% of abundance classes), relative abundance of Odonata (% of abundance classes) and number of ETO-Taxa;

EQR = MMI*4/3-1.2

Eulittoral Riverine lowland MMI/EQR

Three standardised metrics are averaged with equal weighting:

Faunaindex; relative abundance of Chironomidae (% of abundance classes), Margalef-diversity);

EQR = MMI*4/3-1.2

Sublittoral Alpine MMI/EQR

Seven standardised metrics are averaged with equal weighting:

ETO-taxa (% based on taxa number), insecta (% based on individual numbers), habitat preference phythal (% based on abundance classes), feeding type collectors and predators (each in % based on abundance classes), locomotion type sessile (% based on abundance classes), alpha-Mesosaprobic (% based on Individual numbers);

EQR=MMI

Ecological status classification

Ecological status classes are obtained from the EQR values using the following class boundaries:

Boundary	EQR
High/Good	0,8
Good	0,6
Moderate/Poo	0,4
Poor /Bad	0,2

References

http://www.laenderfinanzierungsprogramm.de/cms/WaBoAb_prod/WaBoAb/Vorhaben/LAWA/Vorhaben_des_Ausschusses_Oberflaechengewaesser_und_Kuestengewaesser_%28AO%29/O_5.10/O_5-10_2010_Endbericht.pdf

Miler, O., Brauns, M., Böhmer, J. & Pusch, M. (2011): LAWA-Abschlußbericht „Praxistest des Verfahrens zur Bewertung von Seen mittels Makrozoobenthos" (Projekt-Nr. O 5.10)

Miler, O., Brauns, M., Böhmer, J. & Pusch, M. (2013): LAWA-Abschlußbericht „Feinabstimmung des Bewertungsverfahrens von Seen mittels Makrozoobenthos“ (Projekt-Nr. O 5.10)

Zenker, A., Baier, B. & Böhmer, J. & (2005): LAWA-Abschlußbericht „Feinabstimmung des Bewertungsverfahrens für Makrozoobenthos in stehenden Gewässern“ (Projekt-Nr. O 4.05)

A.6 United Kingdom:Method: Chironomid pupal exuvial technique [Chironomid pupal exuvial technique]

General information

Detected pressure: Eutrophication

Ecological data from 203 lakes (representing all WFD types available in UK and 0.2 - 315 mg/l CaCO3) were explored using Canonical Correspondence Analysis to produce optima scores and niche breadth from species abundance-weighted data. Monte Carlo randomisation tested (Bonferroni-adjusted) and validated for significant taxa response to nitrogen and phoshorus impact (eutrophication). Nutrient impact scores were significantly related to a compound pressure metric (Total Nitrogen x Total Phosphorus/mean depth) r2= 0.78, n = 166

Internet reference:

http://www.wfduk.org/bio_assessment/bio_assessment/lakes_cpet

http://publications.environment-agency.gov.uk/epages/eapublications.storefront/

4b13bb3403a11362273fc0a80296065b/Product/View/SCHO0609BQFJ&2DE&2DE <

Pertinent literature of mandatory character:

http://www.wfduk.org/bio_assessment/bio_assessment/lakes_cpet

Scientific literature: Ruse, L.P., 2002. Lake reference state deduced from chironomid pupal skin data. International Symposium of Chironomidae, University of Minnesota, USA. Chironomid pupal exuviae as indicators of lake status. Arch. Hydrobiol., 3: 367-390. Ruse, L.P., 2009. Classification of nutrient impact using the chironomid pupal exuvial technique. Ecological Indicators (in press). Ruse, L.P. & S. Brooks, 2005. A guide to the identification of chironomid pupal exuviae occurring in Britain and Ireland. Freshwater Biological Association, UK. Ruse, L.P. & S. Brooks, 2009. Lake reference state deduced from chironomid pupal skin data. International Symposium of Chironomidae, University of Minnesota, USA.

Field sampling/surveying

Short description of sampling: Collect floating debris at leeward lake shore (to which wind is blowing), where floating pupal exuviae will be accumulating from across the lake over the previous 2 days. Best to compose sample from several points along the leeward shore.

Sampling/survey device: Hand net with 250 µm mesh net

Sampled habitat: All available habitats per site (Multi-habitat)

Sampling/survey time: during April to October, 4 times recommended, 2 minimum

Specification of level of determination: assessment is valid both at full species level but also at the genus/species-group level possible using the identification key according to

Wilson & Ruse (2005). Data at genus/species group is used for classifications used for WFD.

Comments: Abundance data were used to develop the assessment tool and determine each species impact score but for lake assessment it has been demonstrated that qualitative data are as efficient as quantitative data in measuring impact of anthropogenic nutrient enrichment

Data evaluation

List of biological metrics: Average impact score of all taxa collected from lake survey of 4 samples. Aggregated data from multiple sampling/survey occasions in time is used

Reference conditions

Key source(s) to derive reference conditions: Existing near-natural reference sites

Reference site characterisation: 20 sites from England, Wales and Scotland, 1999-2007.

Reference criteria: Chosen from their relative proportions of impact-sensitive and tolerant species across all WFD lake types available in UK where possible. Reservoirs were not used because their anthropogenic physical characteristics would distort models of reference condition. Lakes were also not considered as reference if urban land-use was greater than 10 per cent or if they had a catchment population density greater than 10/km^2. Acidified lakes were not used as reference lakes for nutrient tool.

Reference community description: Reference community not described. Relationship between lake nutrient impact score and pressure metric developed. Reference nutrient impact scores determined by regression model using best sub-set regression Log lake area, log mean lake depth, log retention time, log catchment area R2 = 0.79

Boundary setting

Boundaries were set using paired metrics that respond in different ways to the influence of the pressure. Boundaries derived from a plot of the relative frequency of sensitive and tolerant species for all surveys. The best fit describing each data set was a quadratic equation. For species-level nutrient assessment the relative frequency of sensitive species exceeded tolerant species at an EQR of 0.64 at 16 % with a SD 7.07 for the fit of tolerant species:

- In terms of frequency of tolerant species the High/Good boundary was placed at the crossover point minus 7.07 which has an EQR of 0.725.
- The Good/Moderate boundary occurred at a tolerant species frequency of 16.24+7.07 which has a EQR of 0.56.
- The Moderate/Poor boundary was taken as the fitted 0 % sensitive species at EQR 0.37.
- Below an EQR of 0.21, where no sensitive species occurred and observed scores were well below reference scores, was taken as the Poor/Bad boundary.

- Generic level boundaries were derived from the species boundaries as in a linear regression genus-level EQR equalled 0.1854+0.8105* species EQR with an r2 of 93.6 per cent (p<0.001).

"Good status" community: No prescriptive taxa description, CPET provides a surrogate assessment for all benthic macroinvertebrates. Good status species EQR 0.560> <0725, generic EQR 0.639> <0.773.

Uncertainty: CPET overcomes spatial and operator sampling error due to the passive collection of material from a large area of the lake. The largest source of variation unrelated to ecological status was the contingency of which months samples were collected in. Full details of methods are provided by Ruse, L. (2006). Sampling efficiency using the chironomid pupal exuvial technique in a survey of Cotswold Water Park Lake 12 [online]. Available from: http://www.freshwaterlife.org/ .

CPET data for the UK study were collected over four visits among the seven months from April to October. There are 35 possible combinations of 4 months from April to October. EQRnutr were calculated for all 35 combinations to measure the variation due to the contingency of which four months were sampled. This would include spatial variation as samples were taken at the leeward shore on each visit and not necessarily at the same point on the lake. The frequency distribution of all possible EQRnutr was normal about the mean with a Ryan and Joiner correlation of 0.996, normal probability > 10 per cent . The seasonal sampling exercise with Cotswold Lake data provided an EQR Standard Deviation which was used to plot per cent confidence of class points based on a symmetrical SD v mean EQR curve (Ellis J. and Adriaenssens V. (2006) Uncertainty estimation for monitoring results by the WFD biological classifications tools [online]. Environment Agency, Science Report, February 2006.)

B. Common Metric development for Lake Eulittoral Benthic Fauna

According to the intercalibration guidance common metrics had to be used for boundary comparisons as sort of a 'common currency' if the assessment methods are different and cannot be applied to the data of the countries to compare with. In order to translate assessment results into common metrics, it is essential that they are correlated well enough. But since the WFD requires a correlation between assessment results and stressors, the intercalibration guidance also requires a correlation of the common metrics with relevant stressors.

Dataset

Dataset description

The data basis was compiled within the lake macroinvertebrate groups of the AL- and CB-GIG. 7 countries with existing assessment systems for eulittoral macroinvertebrates and 4 additional countries contributed data. 9 of them are represented in the CB-GIG (Table B.1) and 3 in the AL-GIG (Table B.2)

Table B.1 Number of lakes, sites and samples used for the development of eulittoral MMIs for the Central/Baltic GIG

Member State	Method	N lakes	N sites	N samples
Belgium/Flanders (BE/FL)	Multimetric Macroinvertebrate Index Flanders (MMIF)	12	55	119
Germany (DE)	German Macroinvertebrate Lake Assessment (AESHNA)	54	410	410
Denmark (DK)	No method yet	17	79	79
Estonia (EE)	Estimation of Freshwater Quality Using Macroinvertebrates	20	20	20
United Kingdom (UK)	Chironomid Pupal Exuviae Technique (CPET) with parallel eulittoral samples	82	105	175
Latvia (LV)	No method yet	23	23	25
Lithuania (LT)	Lithuanian Lake Macroinvertebrate Index	26	29	56
Netherlands (NL)	WFD-Metrics for Natural Watertypes	32	113	149
Poland (PL)	No method yet	6	21	21

Table B.2 Number of lakes, sites and samples used for the development of eulittoral MMIs for the Alpine GIG

Member State	Method	N lakes	N sites	N samples
Austria (AT)	No method yet	5	14	14
Germany (DE)	German Macroinvertebrate Lake Assessment (AESHNA)	12	131	131

Slovenia (SI)	Slovenian ecological status assessment system for lakes using littoral benthic invertebrates	2	28	28

A variety of environmental data was collected within the GIG groups to characterise the lakes and to check for typolological differences. The basic parameters were ecoregion, intercalibration type, national type, coordinates, lake area, catchment size, altitude above sea level, mean depth and conductivity.

Stressors

Stressor parameters were compiled within the GIG groups in dependence of their importance for intercalibrating the assessment systems and their availability. Different Parameters were collected within the Al- and the CB-GIG. Both GIGs started with whole lake parameters, because the WFD assessment was done on water body level. Due to the low number of data for alpine lakes the AL-GIG decided to work on sampling site level and to collect sampling site specific data in addition.

Table B.3 Stressor Variables for the development of the AL-ICM

Variable	Explanation
Shore alteration%	% altered shore length of total shore length
Landuse_surround	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)
Landuse_catchment	Land-use index from the % of land uses in the lake catchment (1 * % extensive agriculture + 2 * % intensive agriculture + 4 * % urban areas)
Naturalness_site_national	National naturalness classification by expert judgement, based on morphology and landuse of the shoreline and adjacent areas at the sampling sites (5 classes)
Urban_agr%_site	% of non-natural[#] landuses (mainly urban and agricultural areas) directly adjacent to the site (15 m belt at 100 m shore length)
Urban_agr%_site100	% of non-natural[#] landuses (mainly urban and agricultural areas) directly adjacent to the site (100 m belt at 100 m shore length)
Morpho_AT_DE_SI_all	Combined stressor index[##] consisting of 2* Naturalness_site_national, Urban_agr%_site, Urban_agr%_site100, Landuse_surround and Shore alteration%
Morpho_AT_DE_SI_TP	Combined parameters[##] of 2* Morpho_AT_DE_SI_all and TP
TP	Total Phosphorous concentration in mg P/l

[#] all anthropogenically altered areas, except woodlands, successional areas (e.g. scrublands) and natural marshes;

[##] All combined indices are weighted averages of standardised single parameters; for standardisation the parameters were transformed into a range from 1,0 to 5,0;

Since the focus of most assessment methods was clearly on hydromorphological pressure (

Table B.4) morphological stressor parameters dominated (Table B.3 and Table B.5).

The stressor parameters used for the development of the Intercalibration Common Metrics comprise different variables describing landuse and alteration of shore structure (see Table B.3 and Table B.5 for more details).

Table B.4 Pressures indicated by the MMIs of member states (pressures in brackets are minor pressures indicated by the respective MMI)

Member State	Method	Pressure
Belgium/Flanders (BE/FL)	Multimetric Macroinvertebrate Index Flanders (MMIF)	hydromorphology, eutrophication
Germany (DE)	German Macroinvertebrate Lake Assessment (AESHNA)	hydromorphology, (eutrophication)
Estonia (EE)	Estimation of Freshwater Quality Using Macroinvertebrates	hydromorphology, eutrophication
United Kingdom (UK)	Chironomid Pupal Exuviae Technique (CPET)	eutrophication
Lithuania (LT)	Lithuanian Lake Macroinvertebrate Index	eutrophication, (hydromorphology)
The Netherlands (NL)	WFD-Metrics for Natural Watertypes	hydromorphology, (eutrophication)
Slovenia	Slovenian ecological status assessment system for lakes using littoral benthic invertebrates	hydromorphology

Table B.5 Stressor Variables for the development of the CB-ICM

Variable	Explanation
Shore_alteration	% altered shore length of total shore length
Landuse_surround	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)
Landuse_catchment	Land-use index from the % of land uses in the lake catchment (1 * % extensive agriculture + 2 * % intensive agriculture + 4 * % urban areas)
Landuse_shore	Lnduse in the 15m belt around the whole lake (4* [%artificial] + 1,5* [%agriculture])
Morphoindex	Combined stressor index[##] consisting of shore_alteration, landuse_surround, 2* landuse_shore
Morpho_TP	Combined stressor index[##] consisting of 2* morphoindex and TP

[#] all anthropogenically altered areas, except woodlands, successional areas (e.g. scrublands) and natural marshes;

[##] All combined indices are weighted averages of standardised single parameters; for standardisation the parameters were transformed into a range from 1,0 to 5,0;

Metric selection

Metric calculation

Metric results will be dependent on the taxonomic resolution of the taxa list. The differences in determination level between the countries was analysed and harmonised within the GIGs. In the AL-GIG the taxonomic level was maintained on mostly species level for all taxa with the exception of Chironomidae and Oligochaeta, which were transformed to family level. In the CB-GIG Oligochaeta were also transformed to family level, Chironomidae to subfamily/tribe and most other taxa were left unchanged on mostly species level. For both GIGs, all meio- or microfauna, as well as Acari and parasites were excluded.

Using the harmonised taxa lists, over 120 biological indices were calculated within the Access-databases of the GIG groups. The algorithms and ecological information was identical to the current Asterics software (version 3.1), developed by the EU projects AQEM and EUROLIMPACS. Some additional indices for lakes were created on the basis of that information (e.g. "no_ETO"= number of Ephemeroptera + Trichoptera + Odonata taxa). Only the Alpine Faunaindex was based on a different indicator list, which was originally derived during the development of the German assessment system AESHNA and further extended using the intercalibration data.

From these indices many were excluded, for which there was no rationale why a metric is supposed to increase or decrease with the degradation of a water body. For example some stream indices were considered as unsuitable for lakes. Finally 71 indices were tested.

More details on these can be found in the Table B.13.

Selection of candidate metrics

To ensure a successful intercalibration, the metrics have to be well correlated with the national assessment systems of all countries. At the same time it is desirable to have a good correlation with the stressor parameters.

Since the pressure situation differs between countries, the biological indices were analysed for the whole dataset as well as for each country separately. The data of some countries however, do not cover a wide range of the pressure gradient. This leads to weaker correlations for these countries.

The criteria for the selection of candidate metrics were in descending order:

- Overall correlation with the national Ecological Quality Ratios (EQR values),
- correlation with the national EQRs for each country separately,
- overall correlation with the stressor variables and
- correlation with the stressor variables for each country separately.

To judge the strength and quality of the correlation Spearman's and Pearsson's R were calculated and Scatterplots were inspected for separation quality near the presumable

good-moderate boundary. The focus was laid on the combined morphological indices, which yielded the highest R^2 values.

Within the Alpine GIG only 2 countries had a method and 3 countries supplied data. R^2 values >0,5 between the national EQR and metrics for all data together could only be obtained for the Faunaindex (FI_AL, Figure B.1), while all other metrics had much weaker correlations. For single countries, especially Slovenia, there were metrics with stronger correlations, but these metrics did not work that well in other countries.

The reason for these differences is to be attributed most likely to differences in sampling design, because metric responses were very similar between the countries with a similar sampling design (Austria and Slovenia with multihabitat sampling, Germany with habitat specific sampling).

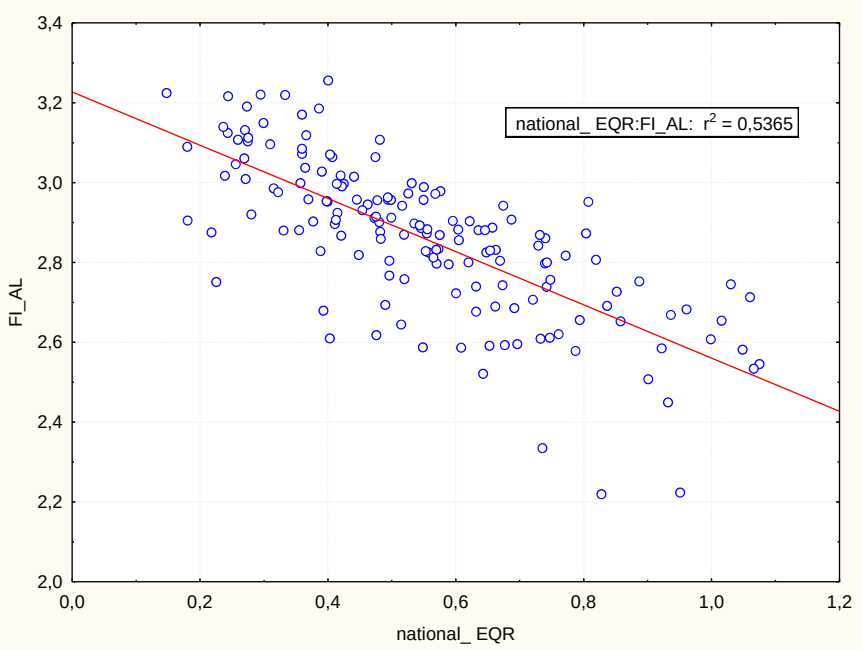


Figure B.1 Correlation between national assessment result (national_EQR) and the Faunaindex FI_AL.

Based on all results the following metrics were selected for the alpine lakes as candidates for combination into multimetric indices:

- Faunaindex FI_AL
- % Odonata (% in relation to abundance classes)
- % ETO (% in relation to taxa number)
- Shannon-Wiener diversity
- number of taxa
- % feeding type gatherer (% in relation to abundance classes)
- rk (reproduction strategy r / k)
- % indifferent taxa (% in relation to abundance classes).

For Central Baltic Lakes the candidate metrics were:

- number of EPTCBO-taxa
- number of ETO-taxa
- ASPT
- % Odonata (% in relation to abundance classes)
- % ETO (% in relation to abundance class)
- Reproduction strategy: r-/k-strategists
- % ETO (% in relation to abundance class)
- % indifferent taxa (% in relation to abundance class)

Metric standardisation and normalisation

Metric values need to be standardised to account for biogeographical and/or methodological differences. Also type specific differences have to be considered, if necessary.

The evaluation of Scatterplots revealed no differences for the intercalibration types CB1 and CB2 for Central/Baltic lakes and AL3 and AL4 for Alpine lakes. However within the Alpine Lakes the differences between the smaller (<5 km2) and larger lakes (> 5 km2) according to the German national macroinvertebrate assessment had to be taken into account. This was only relevant for the German data, because only one of the Austrian lakes and no Slovenian lake was larger than 5 km2.

As long as sufficient data are available for each data group to be normalised, which cover the whole stressor gradient, the normalisation can be performed using upper and lower anchor points. Hereby the upper anchor corresponds to the upper limit of the metric's value under reference conditions, and to the lower limit of the metric's value under the worst attainable conditions.

Since reference lakes were scarce in most countries and several countries covered only a small part of the stressor gradient. This approach could not be satisfactorily applied to many countries and an alternative approach was applied to standardise the metrics in a first step, before normalising the data. This approach uses the full dose response curve of a metric to adjust for country differences in metric responses. The procedure will be described in detail within the final intercalibration report of the CB-macroinvertebrate group. Instead of using only parts of the metric responses to the stressor (benchmarks or references) it uses the full regression curve to calculate the differences between the countries. This is done with linear mixed models. We used the R statistics package lme4 with Morpho_AT_DE_SI_all as stressor gradient and the offset as random factor. Table B.6 gives the resulting offsets for the AL-lakes and Table B.7 for the CB-lakes.

Table B.6 Offsets for the alpine lake metric standardisation calculated with linear mixed models.

Group (country_laketype)	FI_AL	no_Taxa	gatherer	rk
AT_small(<5)	0.049	1.19	0.30	-0.0416
DE_large(>5)	0.100	6.47	-4.14	0.0307
DE_small(<5)	-0.054	-6.91	0.91	0.0483
SI_small(<5)	-0.095	-0.75	2.40	-0.0374

Table B.7 Factors/offsets for the central/baltic lake metric standardisation calculated with linear mixed models.

country	factor no_EPTCBO	factor no_ETO	offset ASPT	offset Odo_HK	offset %ETO	offset rk	offset %Lithal	offset IN_HK
BE	1.08	1.26	-0.45	0.21	-7.49	-0.0275	-5.14	5.15
DE	1.11	1.12	-0.04	-1.93	-4.83	0.0316	1.35	8.42
EE	1.00	0.99	-0.10	4.58	4.23	-0.0187	-1.84	-3.64
LT	0.98	0.84	0.41	0.84	14.17	-0.0294	2.40	-6.15
NL	0.57	0.63	-0.08	0.60	-4.92	0.0152	-1.11	-3.54
UK	1.26	1.36	0.26	-4.29	-1.16	0.0289	4.34	-0.23

Table B.8 Anchor points of the candidate metrics for alpine lakes.

Metrics	upper	lower
ETO_Art%	55	20
FI_AL	3.1	2,6
Shannon-Wiener index	3,0	1,5
r/K	0,3	0,12
IN_HK	45	32
Odo_HK	3	0
Gather_HK	52	31
No_taxa	35	14

Table B.9 Anchor points of the candidate metrics for central/baltic lakes.

Metrics	upper	lower
no_EPTCBO	20.0	2.3
no_ETO	15.0	1.3
ASPT	5.5	3.6
Odo_HK	10.2	-0.6
ETO_HK	48.0	9.8
rk	0.26	0.05
rk_HK	0.42	0.04
LIT_HK	25.1	8.7

IN_HK	66.3	26.6

Using these anchors, the formula for normalisation is

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

for metrics decreasing with increasing impairment, and

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

for metrics increasing with increasing impairment.

Generation of Multimetric Indices

Normalised Metrics can be simply averaged to generate multimetric indices. Equal Weight was preferably given to all metrics. But to improve the correlations with the national methods the MMIs were calculated with both, single and double weighting of the faunaindex for alpine lakes and the number of EPTCBO taxa for Central/Baltic lakes, because these metrics were the best correlating with both, the national methods and the stressor parameters.

41 MMI-variants were tested for Central/Baltic and 2*17 for alpine lakes. These variants contained 3 to 6 Metrics, where at least one metric belonged to one of the three WFD-types required (sensitivity, taxonomic composition, diversity; Table B.10 and Table B.11).

Table B.10a and b 24 Metric combinations tested for Common Multimetric Index development for the Central/Baltic lakes.

Variant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Sensitivity metrics																								
ASPT	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Taxonomic composition and functional groups																								
Odo_HK	x	x			x	x			x				x				x	x	x	x				
ETO_HK			x	x			x	x		x				x			x	x			x	x		
rk	x	x	x	x							x				x		x	x	x	x	x	x		
LIT_HK	x	x	x	x	x	x	x	x				x				x	x	x						
IN_HK																			x	x	x	x	x	x
Diversity																								
no_EPTCBO	x		x		x		x		x	x	x	x					x		x		x		x	
no_ETO		x		x		x		x					x	x	x	x		x		x		x		x
Variant			**25**		**26**		**27**		**28**		**29**		**30**		**31**		**32**		**33**		**34**		**35**		**36**	
Sensitivity metrics																								
ASPT_norm			x		x		x		**x**		x		x		x		x							
Taxonomic composition and functional groups																								
Odo_HK_norm			x		x						x		x		x									
ETO_HK_norm							x		**x**				x		x									
rk_norm																								
LIT_HK_norm					x				**x**															
IN_HK_norm			x				x																	
stage%Hkadult																								
dissem%HKair																								
Diversity																								
no_EPTCBO			2x		2x		2x		2x		2x		2x		2x		2x		2x		2x		2x	

Table B.11Metric combinations tested for Common Multimetric Index development for the alpine lakes. All variants were calculated with single and double weighting of the Faunaindex.

Variant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Sensitivity metrics																	
FI_AL	x	x	x	x	x	x	x	x	x	x	x	x	**x**	x	x	x	x
Taxonomic composition and functional groups																	
Odo_HK	x	x	x	x	x	x	x	x	x	x				x	x	x	x
ETO_Art%	x	x	x	x					x		x						
gather_HK%	x	x	x	x	x	x	x	x			x	x	**x**	x	x	x	x
rk		x		x		x		x			x	x	**x**		x		x
IN_HK			x	x			x	x								x	x
Diversity																	
Shannon-Wiener					x	x	x	x		x		x					
Taxazahl													**x**	x	x	x	x

All MMI-variants were correlated with the national methods and the stressor variables. Using the same criteria as for the single metrics the final common multimetric indices were selected.

The resulting Alpine Intercalibration Common Metric was variant 13 with double weighting of the faunaindex:

ALP-ICM = (2*FI_AL+gather_HK+rk+no_taxa)/5.

The resulting Central/Baltic Intercalibration Common Metric was variant 28:

ICM = (2* no_EPTCBO + ASPT + %ETO +%lithal)/5

Correlations of the ICM with the national methods are stronger for the AL- than for the CB-lakes (

Table B.12). This can be explained by the higher number and larger heterogeneity of the countries and lakes within the CB-GIG. The correlations of the ICM with the stressors were slightly stronger than that for the CB-GIG, but weaker for the AL-GIG.

Exemplary graphs for the correlation between the ICM and the national methods as well as with the morphological stressor are given in Figure B.2, Figure B.3 and Figure B.4.

The correlations within the CB-GIG are higher for the individual countries which cover most of the pressure gradient (e.g. Pearsson R for CB-ICM with national EQR = 0,66 for NL and 0,68 for DE) and much weaker for countries covering too small parts of the gradients.

Table B.12Correlation coefficients (Pearsson's R/Spearman's R) of the developed Intercalibration Common Metrics.

	CB-ICM		ALP-ICM	
	Pearsson's R	**Spearman's R**	**Pearsson's R**	**Spearman's R**
National methods	0.54	0.52	0.79	0.77
Naturalness_site			-0.49	-0.44
Morphology index	-0.57	-0.59	-0.42	-0.35
Morphology-TP index	-0.62	0.64		
Ln TP	-0.47	-0.49	0.11	0,17

Correlations with TP and other water chemistry parameters are not significant within the AL-GIG. Within the CB-GIG the correlations with TP are almost 0.5, but weaker than with morphology. The combined morphology-TP index used within the CB-GIG gave stronger correlations than morphology and TP alone.

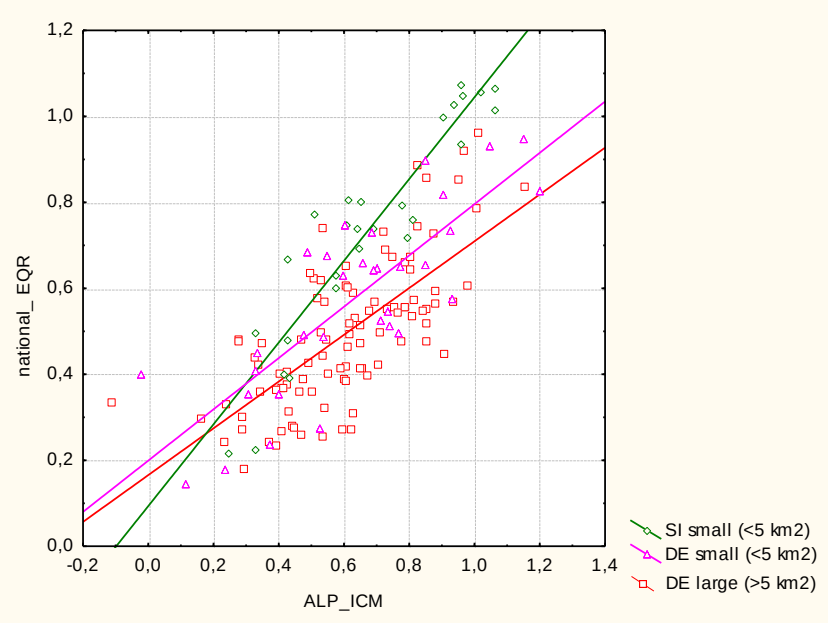


Figure B.2 Correlation between the selected ALP-ICM and the national assessment result (national_EQR).

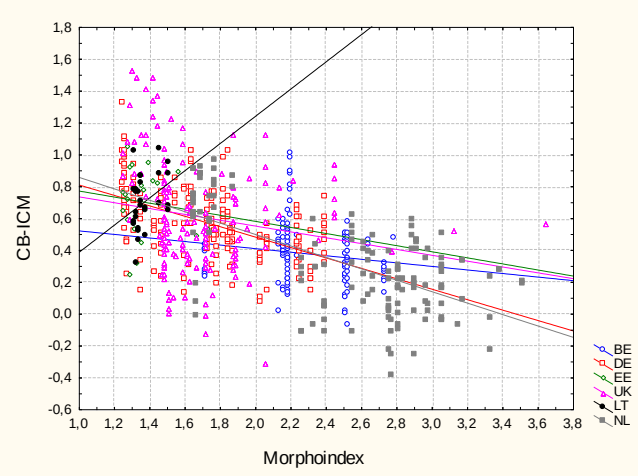


Figure B.3 Correlation between the combined morphological stressor (Morphoindex) and the selected CB-ICM. Note that some countries cover only very small parts of the stressor gradient. For LT this leads to a regression curve very much deviating from the others.

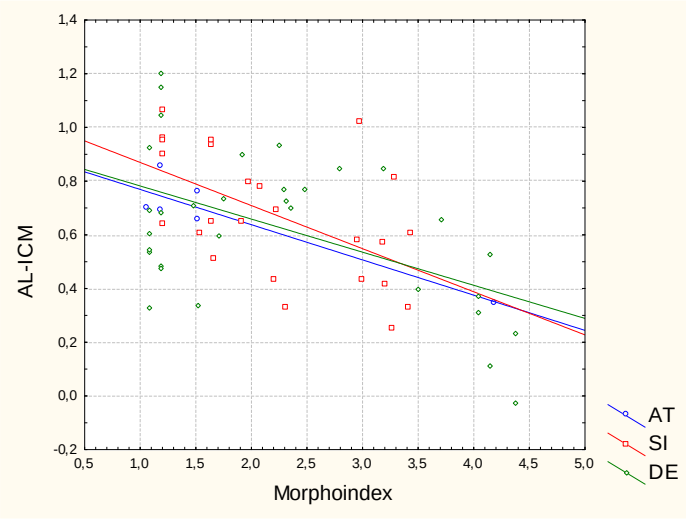


Figure B.4 Correlation between the combined morphological stressor (Morpho_AT_DE_SI_all) and the selected ALP-ICM. For the small lakes (< 5,0 km2).

References

Mathes, J., Plambeck, G., Schaumburg, J. (2002) Das Typisierungssystem für stehende Gewässer in Deutschland mit Wasserflächen ab 0,5 km² zur Umsetzung der Wasserrahmenrichtlinie. In: Implementierung der EU-WRRL in Deutschland: Ausgewählte Bewertungsmethoden und Defizite, 15-23, (Deneke, R. & Nixdorf, B., eds).

Michels, U., Böhmer, J. (2007) Bestandserfassung der benthischen wirbellosen Fauna in ausgewählten Seen des Landes Brandenburg im Jahr 2007. Projektbericht im Auftrag des Ministeriums für Ländliche Entwicklung, Umwelt und Verbraucherschutz des Landes Brandenburg, 1-105.

Table B.13Example for a correlation table between biological indices and stressor parameters for the alpine GIG; Spearman's R for all countries together.

biological index	National_ EQR	Naturalness_ site_ national	Morpho_AT_ DE_SI_all_TP	Morpho_AT _ DE_SI_all	urb_ agr% _site	urb_ agr 100 _site	shore_ alteration%	landuse_ surround	landuse_ca tchment	t-P _mg_l
afil%	-0,06	-0,07	0,06	0,05	-0,09	0,09	0,26	0,30	0,38	0,35
AKA%	-0,35	0,04	0,06	0,04	-0,09	-0,02	0,15	0,30	0,13	0,26
AKA_HK%	-0,25	-0,01	0,04	0,02	-0,12	-0,04	0,21	0,38	0,25	0,16
ASPT	-0,06	0,04	0,02	-0,01	-0,03	-0,02	0,12	0,19	0,18	0,17
ASPT_IZ	-0,15	0,03	0,16	0,14	-0,01	0,08	0,29	0,38	0,20	0,24
BMWP_Score	0,02	-0,04	0,08	0,08	-0,04	0,10	0,23	0,30	0,15	0,17
chiro%	-0,20	0,13	0,01	0,02	0,11	0,05	-0,13	-0,06	-0,05	-0,07
chiro_HK	-0,06	0,15	-0,05	-0,04	0,16	-0,04	-0,28	-0,35	-0,12	-0,19
Chironominae%	-0,06	0,10	-0,03	-0,02	0,08	0,03	-0,19	-0,12	-0,07	-0,04
Coleoptera%	-0,24	-0,04	-0,07	0,02	-0,08	-0,03	0,26	0,15	-0,16	-0,15
Crust%	0,06	-0,08	0,08	0,09	0,00	0,04	0,25	0,41	0,00	0,08
DomFam%	0,01	0,01	-0,07	-0,13	-0,05	-0,09	-0,18	-0,14	0,16	0,01
EPT%	-0,32	0,16	0,21	0,28	0,15	0,16	0,35	0,24	-0,03	0,00
EPT_HK%	-0,32	0,16	0,23	0,21	0,09	0,14	0,28	0,29	0,28	0,20
EPTCBO%	-0,22	0,02	0,15	0,17	-0,01	0,15	0,35	0,31	0,25	0,21
ETO_Art%	-0,13	0,13	0,14	0,08	0,02	0,08	0,12	0,18	0,28	0,30
ETO_HK%	-0,15	0,10	0,19	0,13	0,03	0,12	0,21	0,24	0,28	0,28
faf_fpf	-0,11	0,09	0,18	0,19	0,07	0,19	0,23	0,23	0,28	0,20
famrich	0,06	-0,07	0,06	0,08	-0,04	0,11	0,22	0,25	0,09	0,12
FI_AL	-0,76	0,41	0,43	0,45	0,35	0,37	0,23	0,35	0,11	0,21
FI_nat	-0,58	0,32	0,30	0,41	0,27	0,29	0,33	0,12	-0,07	-0,02
FI_nat_Ind	-0,15	0,20	0,13	0,28	0,23	0,16	0,19	0,02	-0,41	-0,29

biological index	National_ EQR	Naturalness_ site_ national	Morpho_AT_ DE_SI_all_TP	Morpho_AT _ DE_SI_all	urb_ agr% _site	urb_ agr 100 _site	shore_ alteration%	landuse_ surround	landuse_ca tchment	t-P _mg_l
Gastropoda%	0,44	-0,21	-0,08	-0,18	-0,16	-0,09	-0,29	-0,21	-0,04	0,25
gather%	-0,06	0,19	0,16	0,20	0,26	0,11	-0,11	-0,14	-0,15	-0,23
gather_HK	-0,09	0,16	0,03	0,15	0,18	-0,02	-0,03	-0,08	-0,21	-0,45
grazer%	-0,15	0,00	-0,06	0,02	-0,07	-0,03	0,17	0,11	-0,21	-0,23
grazer_HK	-0,05	-0,05	-0,06	0,01	-0,04	0,04	0,16	-0,03	-0,05	-0,22
IN%	0,12	-0,13	0,00	-0,04	-0,11	-0,05	0,14	0,19	0,37	0,14
Insecta%	-0,33	0,16	0,05	0,03	0,06	-0,01	-0,07	-0,01	-0,06	0,05
Lake%	0,43	-0,11	0,00	-0,08	-0,05	0,00	-0,08	-0,14	0,05	0,16
LB_HK%	-0,29	-0,01	0,00	0,01	-0,02	-0,01	0,02	0,06	-0,07	0,13
LIT_HK	-0,34	0,06	0,02	0,06	0,01	0,01	0,23	0,20	0,02	-0,10
litoral%	-0,08	0,18	0,13	0,11	0,16	0,10	-0,25	-0,19	-0,15	0,00
lse%	-0,14	0,00	0,16	0,12	0,00	0,12	0,28	0,35	0,44	0,28
lsw_HK	0,38	-0,06	-0,04	-0,05	-0,04	0,01	0,00	-0,13	-0,11	-0,05
no_C	-0,21	-0,06	0,05	0,07	-0,07	0,04	0,25	0,28	-0,04	0,06
no_E	-0,32	0,11	0,16	0,23	0,07	0,18	0,36	0,43	0,19	0,10
no_EPT	-0,25	0,09	0,26	0,21	0,02	0,20	0,31	0,47	0,34	0,40
no_EPTCBO	-0,24	0,06	0,27	0,20	-0,01	0,21	0,30	0,48	0,36	0,47
no_ETO	-0,15	0,07	0,26	0,18	-0,01	0,20	0,26	0,44	0,38	0,48
no_individuals	-0,21	0,01	0,15	0,10	-0,07	0,14	0,16	0,26	0,35	0,27
no_P	0,17	-0,03	-0,13	-0,10	0,02	-0,13	-0,13	-0,12	-0,28	-0,14
no_Taxa	-0,14	0,02	0,24	0,17	-0,01	0,20	0,27	0,45	0,33	0,40
no_Tricho	-0,19	0,08	0,27	0,17	0,01	0,20	0,25	0,41	0,38	0,47

biological index	National_ EQR	Naturalness_ site_ national	Morpho_AT_ DE_SI_all_TP	Morpho_AT _ DE_SI_all	urb_ agr% _site	urb_ agr 100 _site	shore_ alteration%	landuse_ surround	landuse_ca tchment	t-P _mg_l
Odo%	0,60	-0,27	-0,22	-0,30	-0,17	-0,15	-0,30	-0,24	-0,17	0,22
Odo_HK	0,63	-0,25	-0,17	-0,29	-0,17	-0,12	-0,31	-0,23	-0,09	0,28
oligo_HK%	-0,04	-0,10	-0,10	-0,07	-0,18	-0,02	0,20	0,21	0,16	0,03
orthoclad/chir%	-0,31	0,08	0,10	0,07	0,02	0,04	0,08	0,14	0,10	0,01
Orthocladinae%	-0,36	0,13	0,08	0,07	0,07	0,05	0,01	0,09	0,03	-0,03
PEL%	0,15	-0,01	-0,08	-0,04	-0,03	-0,04	-0,11	-0,31	-0,24	-0,12
pfil%	0,00	-0,09	-0,01	-0,09	-0,08	0,00	-0,05	0,05	0,10	0,24
PHY%	0,07	-0,10	-0,20	-0,15	-0,06	-0,15	-0,13	-0,20	-0,27	-0,25
Pleco%	0,17	-0,02	-0,13	-0,10	0,02	-0,12	-0,15	-0,14	-0,30	-0,15
POM%	0,23	-0,12	-0,05	-0,02	-0,06	-0,03	0,20	0,14	-0,03	-0,01
prodiamesinae %	-0,30	0,19	0,17	0,21	0,15	0,14	0,09	0,26	0,12	-0,07
PSA%	0,00	0,09	0,02	0,03	0,05	0,05	-0,13	-0,15	-0,02	-0,07
PTI	0,01	0,02	-0,01	0,00	0,11	0,08	-0,24	-0,26	-0,27	-0,14
rk	-0,53	0,16	0,25	0,19	0,08	0,19	0,13	0,08	0,24	0,31
rk_HK	-0,44	0,13	0,20	0,13	0,03	0,15	0,07	0,03	0,30	0,32
RTI	-0,07	0,01	-0,09	0,06	-0,01	-0,07	0,25	0,20	-0,29	-0,53
shred%	0,17	-0,06	0,06	0,03	0,00	0,00	0,10	0,21	-0,02	0,07
ShW	-0,24	0,02	0,06	0,10	0,02	0,06	0,24	0,29	-0,09	0,10
SI	0,26	-0,13	-0,01	-0,12	-0,02	0,00	-0,19	-0,20	0,07	0,29
sza%	0,25	-0,05	0,00	-0,06	0,03	-0,05	-0,18	-0,16	-0,23	-0,04
szo%	-0,10	-0,05	-0,02	0,01	-0,11	0,05	0,23	0,30	0,25	0,14

biological index	National_ EQR	Naturalness_ site_ national	Morpho_AT_ DE_SI_all_TP	Morpho_AT _ DE_SI_all	urb_ agr% _site	urb_ agr 100 _site	shore_ alteration%	landuse_ surround	landuse_ca tchment	t-P _mg_l
szp_HK	0,29	-0,13	-0,05	-0,08	0,00	-0,02	-0,09	-0,15	-0,06	-0,06
Tanypodinae%	0,09	-0,05	-0,09	-0,07	-0,06	-0,04	0,04	0,03	-0,16	-0,02
Tricho%	-0,12	-0,04	0,08	0,08	-0,01	0,02	0,38	0,32	0,17	0,05
xeno%	0,12	-0,09	-0,03	0,02	-0,02	0,05	0,13	0,06	0,08	-0,07
xeno_HK%	0,15	-0,09	-0,03	0,03	0,01	0,07	0,14	0,03	0,01	-0,14
xenoligo	-0,09	-0,05	-0,02	0,01	-0,10	0,06	0,22	0,30	0,25	0,13

C. Approaches for Metric Standardisation in Intercalibration: Reference Benchmarking, Alternative Benchmarking and Continuous Benchmarking in comparison

Introduction – the need for standardisation

Due to biogeographical and typological reasons as well as differences in data acquisition biological data of different countries or different water types cannot be compared without concern. As an example the number of taxa might be generally higher in a country than in others, because the sampling covers much more area per site. Additionally, the national assessment metrics differ between countries. Hence, they cannot be compared directly.

For this reason the Water Framework Directive (WFD) demands the use of **reference conditions** as a benchmark to standardise biological assessment metrics: assessment results have to be expressed as Ecological Quality Ratios (EQR), i.e. the ratio between the observed index value and the index value which is typical at reference sites.

The standardisation of biological metrics is also crucial for the comparison and harmonisation of ecological status class boundaries in intercalibration. Within the intercalibration "common metrics" or "pseudo-common metrics" as well as national assessment results are standardised. Due to the scarcity of reference sites also **alternative benchmarks** at a certain level of pressure have been applied as a second option.

However, both options rely on the availability of undisturbed or similarly disturbed sites among countries within a common type. If one or more countries lack sites featuring similar levels of anthropogenic pressure alternative benchmarking is not possible. This will be a common problem if, for instance, countries featuring contrasting population densities or land use practices, like Poland and the Netherlands, are involved in the same exercise. In such cases **continuous benchmarking** allows for the metric standardisation required in intercalibration.

In this paper we describe continuous benchmarking in comparison with the other approaches.

Data availability and approaches to determine the differences between countries

Figure C.2a illustrates the basic problem: An assessment or intercalibration metric responds differently to a gradient of anthropogenic pressure for two countries. Therefore, the values of the metrics cannot be compared directly. Ideally, the available data covers the whole pressure gradient. The difference between the metric values is indicated by the arrow in the centre. In order to account for this difference, it first has to be determined. The subsequent step of standardisation which is the same for all benchmarking approaches is described further below.

1. Reference benchmarking: Prerequisite is the availability of references, independent of the completeness of the remaining pressure gradient Figure C.2a

and Figure C.2b). The average metric values at the references are used to determine the differences between countries (s. Figure C.1). Usually more than 10 to 15 independent reference data points are considered as necessary to determine a precise average.

With insufficient references (Figure C.2c and Figure C.2d) other approaches are necessary:

1. Alternative benchmarking: This approach was already established in the intercalibration guidance. Prerequisite is the availability of benchmark samples for each country within a narrow window of pressure. This pressure has to be specified with the same set of relevant pressure parameters for each country. The average metric values of the benchmarks are used to determine the differences between countries (s.Figure C.1). As for references, 10 to 15 independent benchmark data points per country and water type are considered necessary.
2. Continuous benchmarking: In some intercalibration exercises (e.g. CBlakeGIG Benthic fauna[1], CBrivGIG Macrophytes[2]) we encountered cases with insufficient references and benchmark sites (Figure C.2d). Continuous benchmarking was thus developed as a third option. Prerequisite is the availability of samples with relevant pressure data. Similar to alternative benchmarking all countries need to provide the same (set of) pressure-variables along with the biological data. All data points (summarised by individual regression curves) are used to determine the differences between countries (see Figure C.1 and Figure C.2).

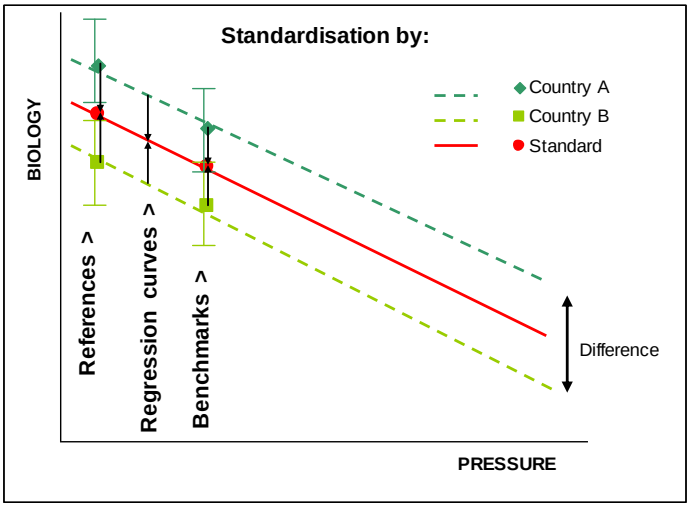


Figure C.1 Adjustment of national metric values based on the offset results from averaged references, averaged alternative benchmarks or regression curves. The length of the arrow at each option indicates the necessary adjustments that have to be

[1] Böhmer (2010 & 2011) CBlakeGIG Benthic fauna - Milestone Reports 3-5. September 2010 – June 2011.

[2] Birk, S. & N. Willby (2011) CBrivGIG Macrophytes - Milestone 5 Report. June 2011. 30 pp.

applied to the whole dataset of a country. Note the slight differences for the three approaches in this example indicates that the number of reference samples (N=5) and benchmark samples (N=6) is too low.(same data as in previous graphs).

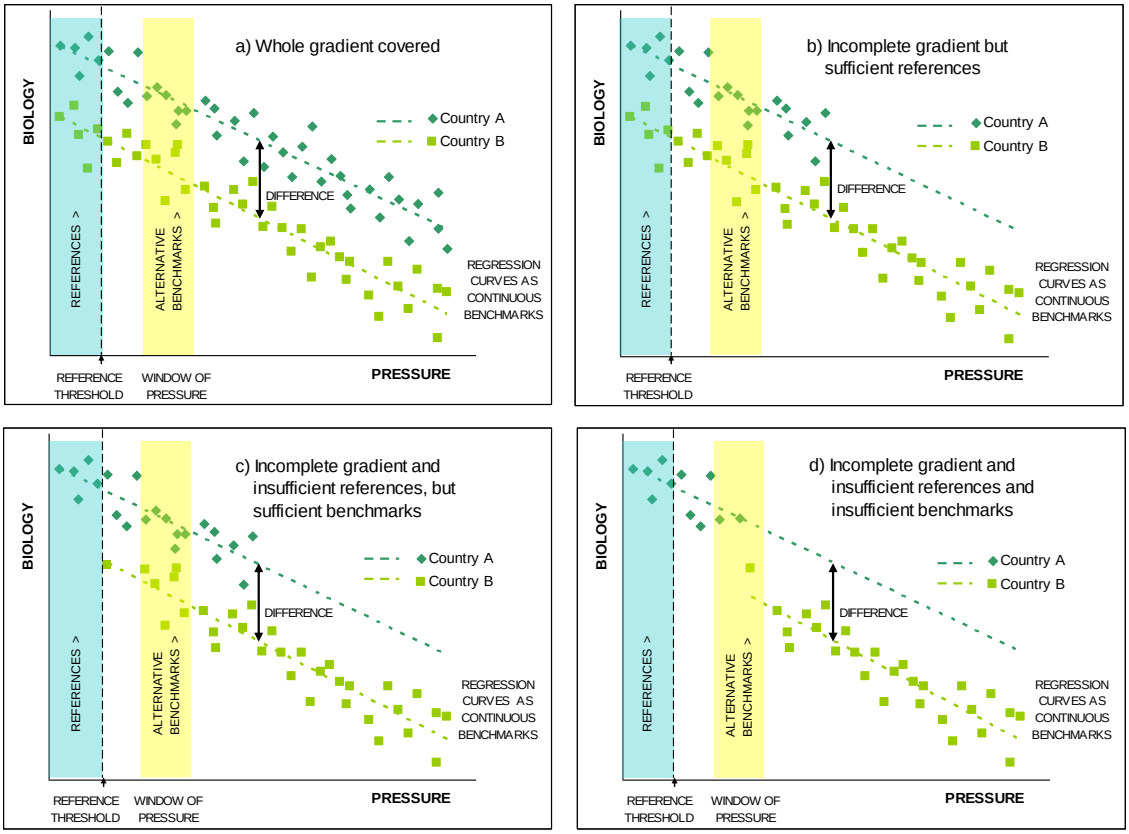


Figure C.2 Three possibilities for standardisation in dependence of data availability.

Note: For demonstration purposes 5 sites are considered to be sufficient to precisely determine the average value of a reference or benchmark. Usually at least 10 to 15 independent data points are considered to be necessary, depending on the scatter of the data.

a. *Ideal: Whole Gradient covered by all countries – All benchmarking approaches possible*

b. *Incomplete gradient but references for all countries – All benchmarking approaches possible*

c. *Incomplete gradient and insufficient references for one or more countries but sufficient alternative benchmarks within a window of pressure – reference benchmarking impossible but alternative benchmarking possible*

d. Incomplete gradient and insufficient references and insufficient alternative benchmarks within a window of pressure – reference and alternative benchmarking impossible but continuous benchmarking still possible

Continuous benchmarking can principally be applied in all cases shown in Figure C.2, but we recommend the reference benchmarking whenever possible (Table C.1), because it is the basic principle for all WFD assessments.

Without sufficient references but sufficient alternative benchmark samples both, alternative and continuous benchmarking can be applied. Both will give the same results if many data points are available for all countries and the pressure-impact-relationship is very strong. If these conditions are not fulfilled continuous benchmarking will give more reliable results, especially if there are countries with reference sites or if the scatter between alternative benchmark samples is high. This is because the references are then included in the determination of the country differences and more points contribute to continuous benchmarking, leading for smaller standard errors in comparison to alternative benchmarking.

In the ongoing intercalibration exercises the metric value within an alternative benchmarking window span almost the whole possible gradient, e.g. EQRs from 0.1 to 0.9. There are many reasons behind it: The "pressure window" might bee too broad, data variability to high, biological response too variable in a selected range (alternative states) etc. Consequently it is often not only a problem to find alternative benchmark sites within a certain window of pressure, but also to find sites with a similar biological impairment level.

Table C.1 Data availability and possibilities to apply the benchmarking approaches.

Data availability	Preconditions	Sufficient References (> 10-15 independent reference samples for each IC-type in each country)	Insufficient references but sufficient alternative benchmarks (>10-15 independent benchmark samples for each IC-type in each country)	Insufficient references and insufficient alternative benchmarks
Reference benchmarking	Reference sites	Best option, recommended	No	No
Alternative benchmarking	Benchmark sites; accompanying pressure data	Possible	Possible, recommended when most countries lack references and standard error of alternative benchmark sites is low	No
Continuous benchmarking	accompanying pressure data	Possible	Possible, recommended when many countries have references or standard error of benchmarks is high	Only option

Standardisation procedure

How can the national assessment metrics be standardised after the differences between the countries were determined by the averaged reference or alternative benchmarks, or the regression curves? The easiest approach to think of is to calculate the offset from the common standard (see Figure C.1) and then subtract this offset from all corresponding data points.

Since benchmarking aims at defining abiotic baselines that standardise the different national metrics across their full range, their response pattern to human pressure is important. Do the metrics only differ at (relatively) undisturbed conditions but converge at the more disturbed end of the gradient, or do these differences persist throughout the gradient? Often also diverging curves can be found.

This distinction determines how to calculate the benchmark standardisation (Figure C.3). It can be obtained either by directly subtracting the offsets yielded by one of the benchmarking approaches from the observed metric-values (=offset correction) or by dividing the observed values by a divisor (= slope correction). If necessary, also a combination of offset and slope correction can be applied.

For EQR-standardisation the divisor can be calculated as 1 + country offset with the country offset being the country value minus the global mean of all countries. For example if the offset was +0.1 for a country, the offset correction would mean to subtract 0,1 from all EQRs, while the divisor correction would mean to divide all EQRs by 1+0,1=1,1. This leads to identical corrections of both approaches at the reference condition (where standardised EQR=1.0 and non standardised EQR=1.1): For subtraction 1.1 - 0.1 = 1.0 as well as for division 1.1 / 1.1 = 1.0.

In intercalibration so far, slope correction has been almost exclusively used for reference benchmarking and offset correction most commonly for alternative and continuous benchmarking.

Although the creators of the WFD demanded the division by a factor (the reference value) to obtain standardised EQR-values. and division was always used for benchmark standardisation in earlier exercises it was important to introduce this distinction here as the calculation affects the relative position of the class boundaries to be compared in intercalibration. In division, for instance, distances increase if the actual benchmark value is smaller than the national reference value; in subtraction all distances stay the same.

Standardisation is possible to different standards as long as the same standard is applied to all data: Usually it is 1 for the reference condition (= EQR), or 1 at an alternative benchmark (when division is applied), or the common view at the alternative benchmark or reference benchmark (when subtraction is applied), or the common view of the whole regression curve (for continuous benchmarking). These standards can be easily transformed to any other scale, but to calculate EQR-values a reference value is necessary. This can be easily derived for the standardised metrics if all countries other

have at least 10-15 independent reference values or otherwise by using the regression curve for all standardised data together to calculate the value at zero pressure.

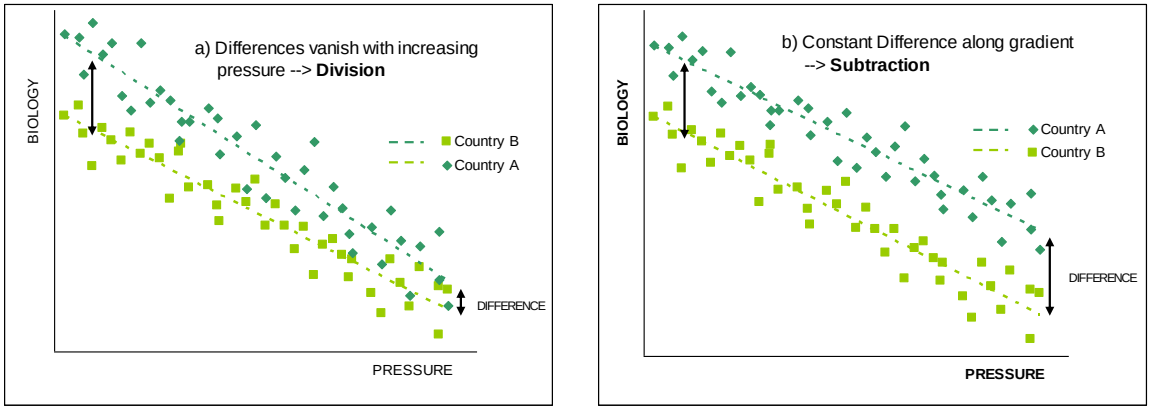


Figure C.3 Types of pressure-impact-relationships and options of calculating the benchmark standardisation

a. Standardisation is done by division if differences between metrics vanish with increasing pressure.

b. Standardisation is done by subtraction if differences remain throughout the entire pressure gradient.

For standardisation of national EQRs in stream macrophyte intercalibration the appropriate calculation (slope or offset correction) was selected by testing if the average value of all per survey in the full dataset was significantly correlated with its standard deviation. In case of a significant positive relationship, i.e. national EQRs converge towards the bad end of the quality gradient, division was used. An insignificant relationship, i.e. constant distances between EQRs across the full gradient, required subtraction.

When the standardised national metric boundary is compared to the common metric national EQR-boundaries for each country, it is also possible to leave the EQR-values for the country in consideration as they are and express the offset of the other countries relative to it. This has the advantage that the national boundaries do not need to be adjusted for comparison. For example when UK method was compared to DE and PL and the offset values relative to the average of all were -0.01 for UK, -0.02 for DE and +0.02 for PL, then UK can be expressed as zero offset from itself, -0.01 from DE (=-0.02 - -0.01) and +0,03 from PL (=+0.02 - -0.01). Then the UK good/moderate boundary remains at 0.6 for comparison. This is done for each country. It is simpler to plot and to explain.

Applying the continuous benchmarking approach

The principle of continuous benchmarking is to adjust all national regression curves (national metric versus pressure gradient) to a common regression curve for all data together (Figure C.3).

The simplest way to determine the differences **without further statistics** is to calculate the metric values at a selected level of pressure with the regression formula of each

regression curve. This would equal the alternative benchmarking, but using the regression curve based on all available data points instead of the average of some alternative benchmark points.

However, the differences vary along the pressure gradient if the curves are not parallel. In this common case several aspects have to be taken into account:

1. Slope correction (Divisor = non-standardised / standard) might be better than offset correction.
2. The slope of the regression curve might be imprecise if the curve covers only a small part of the pressure gradient or if the correlation is weak → use centre part of regression cloud only.
3. If larger parts of the gradient are covered the adjustment might be most appropriate in the range of the boundaries to be intercalibrated. This yields the highest precision in most important range.
4. More overlap of pressure ranges of the country groups to be standardised yields more precise results → two groups with no overlap may be problematic if the correlations are weak.

One disadvantage of this "manual" approach is that the regression curve of all data is not modelled and may change after standardisation. This depends on the data distribution and might require a graphical control (which is always recommended - also for more sophisticated statistical approaches) and a repetition of the process.

In order to model the standard and to receive the correction values directly, **statistical models** may be used, especially General Linear Models (GLM, available e.g. in SPSS and R) and Linear Mixed Models (LMM, e.g. package lme4 in R). Which option is best is not clear: Statisticians advised the CBlakeGIG-Phytoplankton group to use mixed models but the CBrivGIG Macrophyte group to use general linear models. Probably the best model depends on the data, but most likely the differences in results are minor.

To apply the models the biological metric (e.g. national EQR) is set as dependent variable, the pressure variable(s) form the covariates in the model and the country is a fixed or random factor.

Depending on whether offset or offset and slope are modelled as random factors, the output will yield the correction values for offsets or offset and slopes which are then used for standardisation.

The exact steps to perform the statistics depend on the software used, e.g. using the package lme4 in R the model can be specified as "fit.mm2 <- lmer(Metric~ Pressure + (1|country_type),data=data)" with "Metric" being the metric variable, "Pressure" the pressure variable and "country_type" the groups for standardisation (country and water body type).

The **advantages and disadvantages of the statistical models** versus the "manual" approach is that the models give a better standard curve for adjustment and a more

profound standardisation for the complete pressure range, but they are a black box which cannot be easily explained. The advantage of the manual approach is that the focus for highest precision may be laid onto the most important range of the pressure (the boundaries of good status) and that problems with the data distribution may be judged graphically and then taken into account.

First comparisons with phytobenthos data of very large rivers in eight countries gave no significant differences between the manual and the mixed model standardisation by offset correction.

Most **further questions** about continuous benchmarking are either related to general intercalibration issues or specific software used or related to specific situations within the GIG groups:

- When should continuous benchmarking be used?

Answer: it is an alternative for all cases in intercalibration, where reference or alternative benchmarking is to be applied. See Table C.1 for details which benchmarking option is best in dependence of data availability. Usually standardisation is needed for common metrics in option 2 and for EQRs in option 3.

- What to do with offsets?

Answer: The offset gained by continuous benchmarking are used in the same way as offsets obtained by other benchmarking approaches (see above)."

Concluding remarks

Standardisation can have a significant influence on the position of the boundaries. It was found that the country offsets for EQRs are very often at least as high as the change in EQR needed for a boundary to be within the harmonisation band. So they are critical in reaching the correct decision.

J R C T E C H N I C A L R E P O R T S

Water Framework Directive Intercalibration Technical Report

Central Baltic Lake Benthic invertebrate ecological assessment methods

Jürgen Böhmer, Kestutis Arbaciauskas
Rachel Benstead, Wim Gabriels,
Gwendolin Porst, Bart Reeze, Henn Timm

Edited by Sandra Poikane

2014

Report EUR 26504 EN

Joint Research Centre

European Commission
Joint Research Centre
Institute for Environment and Sustainability

Contact information
Sandra Poikane
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 46, 21027 Ispra (VA), Italy
E-mail: sandra.poikane@ec.europa.eu
Tel.: +39 0332 78 9720
Fax: +39 0332 78 9352

http://ies.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

This publication is a Technical Report by the Joint Research Centre of the European Commission.

JRC88273

EUR 26504 EN

ISBN 978-92-79-35457-1

ISSN 1831-9424

doi: 10.2788/7103

Cover photo: Sandra Poikane

Luxembourg: Publications Office of the European Union, 2014

Printed in Ispra, Italy