XGIG Large River Intercalibration Exercise

WFD Intercalibration Phase 2: Milestone 6 report

Report-Version 1.2 (6 September 2012)

This report informs about the intercalibration activity carried out for the national classification schemes of very large rivers using benthic invertebrates and phytobenthos (diatoms).

We identified specific uncertainties related to the intercalibration of individual national methods. On this basis only those classifications were intercalibrated that showed a high level of confidence with regard to the intercalibration analysis. For benthic invertebrates the following countries were included in the analysis: *Estonia*, *Finland*, *Germany*, *Hungary*, *Slovenia* and *Spain*. For diatoms the methods of *Austria (type me1), Belgium (Wallonia), Czech Republic, Estonia, Germany, Hungary, Slovakia* and *Slovenia* were intercalibrated. Beside specific uncertainties the overall level of confidence in invertebrate intercalibration justified an acceptable boundary bias of ≤ 0.50 (instead of ≤ 0.25 usually applied in intercalibration exercises).

We also tested the performance of methods with low confidence in the intercalibration analysis. However, we did not use these outcomes for boundary harmonisation. They rather provide indication of a possible boundary bias that should be tested and validated by alternative intercalibration analyses, if feasible.

ECOSTAT approval (21 March 2012)

All participating countries approved the results of the diatom exercise presented in this report. With reference to the general uncertainties related to the IC procedure the outcomes of the exercise using benthic invertebrates were not approved.

Document history

Version 0.9 (13 January 2012): Description of IC analysis and results
Version 1.0 (12 March 2012): Major update: Uncertainty evaluation of IC outcomes with relevance for comparability
Version 1.1 (28 April 2012): Correction of minor details; new analysis of Dutch comparability (diatoms) based on corrected dataset; inclusion of French diatom assessment method into compliance analysis
Version 1.2 (see above): Correction of minor details; Member States' decisions on proposed boundary adjustments for diatom classifications; analysis of adjusted Slovak and Slovene diatom classifications

Table of contents

EUROPEAN COMMISSION
DIRECTORATE GENERAL JRC
JOINT RESEARCH CENTRE
Institute of Environment and Sustainability

WFD Intercalibration Phase 2: Milestone 6 report

Water category/GIG/BQE/ horizontal activity:	**XGIG Large Rivers**
Information provided by:	Franz Schöll, Sebastian Birk[1], Jürgen Böhmer

1. Organisation

1.1. Responsibilities

Indicate how the work is organised, indicating the lead country/person and **the list of involved experts of every country:**

Franz Schöll (DE, lead), Sebastian Birk (DE), Jürgen Böhmer (DE), Wim Gabriels (BE-FL), Kris van Looy (BE-FL), Christine Keulen (BE-WL), Jukka Aroviita (FI), Stina Drakare (SE), Joakim Lücke (SE), Irene Carrasco (ES), Nuno Caiola (ES), Carles Ibanez (ES), Yorick Reyjol (FR), Juliette Rosebery (FR) , Béla Csányi (HU), Gábor Várbíró (HU), Denisa Nemejcova (CZ), Libuse Opatrilova (CZ), Alena Slavikova (CZ), Franz Wagner (AT), Gisela Ofenböck (AT), Richard Hemsworth (UK), Rachel Benstead (UK), Simone Ciadamidaro (IT), Andrea Buffagni (IT), Audrone Pumputyte (LT), Gorazd Urbanic (SI), Matus Haviar (SK), Emilia Misikova Elexova (SK), Roel Knoben (NL), Eddy Lammens (NL), Herman van Dam (NL), Marcel van den Berg (NL), Irja Truumaa (EE), Sirje Vilbaste (EE), Henn Timm (EE), Steinar Sandøy (NO), Elin M. Lien (NO), Ann Kristin Schartau (NO), Joao Ferreira (PT), Samantha Jane Hughes (PT), Maria Teresa Ferreira (PT), Gabriel Chiriac (RO)

1.2. Participation

Austria Belgium (Flanders), Belgium (Wallonia), Czech Republic, Estonia, Finland, France, Germany, Hungary, Italy, Lithuania, Netherlands, Romania, Slovakia, Slovenia, Spain, Sweden

1.3. Meetings

First workshop – Koblenz 22-23 Sept. 2009
Second workshop – Koblenz 19-20 April 2010
Third workshop – Koblenz 22-23 Sept. 2010
Forth workshop – Koblenz 10-11 Febr. 2011
Fifth workshop – Koblenz 22-23 Sept. 2011

[1] Author for correspondence: sebastian.birk@uni-due.de

2. Overview of Methods to be intercalibrated

Identify for **each** MS the national classification method that will be intercalibrated and the status of the method

1. finalized formally agreed national method,
2. intercalibratable finalized method,
3. method under development,
4. no method developed

BQE	Country	Method name	Status
Benthic Invertebrates	Austria[a]	Assessment of the Biological Quality Elements - part benthic invertebrates	1
	Belgium (Flanders)	Multimetric Macroinvertebrate Index Flanders	1
	Belgium (Wallonia)	Global biological index adapted to large watercourses and deep rivers	1
	Estonia	Estimation of freshwater quality using macroinvertebrates	1
	Finland	Finnish multimetric index	1
	Germany	Assessment method for rivers using benthic invertebrates	1
	Hungary	Hungarian Multimetric Macroinvertebrate Index for large and very large rivers	1
	Lithuania	Assessment system for rivers using macrozoobenthos indicators (Danish Stream Fauna Index)	1
	Netherlands	WFD-metrics for natural water types	1
	Romania	Assessment method for ecological status of water bodies based on macroinvertebrates	1
	Slovakia	Slovak assessment of benthic invertebrates in large rivers	1
	Slovenia	Slovenian assessment system for rivers using benthic invertebrates	1
	Spain	Iberian Biological Monitoring Working Party	1
Benthic diatoms	Austria[b]	Assessment of the Biological Quality Elements - part phytobenthos	1
	Belgium (Flanders)	Proportions of Impact-Sensitive and Impact-Associated Diatoms	1
	Belgium (Wallonia)	Pollution Sensitivity Index	1
	Czech Republic	Assessment system for rivers using phytobenthos	1
	Estonia	Assessment system for rivers using phytobenthos in Estonia	1
	Finland	Indice de Polluosensibilité Spécifique (Specific Pollution Sensitivity Index SPI)	1
	France[c]	Indice Biologique Diatomées 2007 (IBD2007)	1
	Germany	German assessment system for Macrophytes and Phytobenthos according to the WFD	1
	Hungary	Improvement of the Hungarian ecological water qualification system - Phytobenthos in Rivers	1
	Netherlands	WFD-metrics for natural water types	1
	Slovakia	Ecological status assessment system for rivers using phytobenthos	1
	Slovenia	Ecological status assessment system for rivers using phytobenthos and macrophytes – part phytobenthos	1
	Sweden	Benthic algae in running water - diatom analysis	1

[a] Only the multimetric index for large Alpine rivers of Austria was considered.

[b] For Austria two national diatom classifications with different reference conditions and boundary settings were intercalibrated (me1 – basic trophic state: meso- to eutrophic 1, me2 – basic trophic state: meso- to eutrophic 2)

[c] Only included in the analysis of WFD-compliance.

In Annex I of this report further descriptions of the national assessment methods for large rivers are included. Furthermore, relevant issues of large river bioassessment are summarised in the "*Conceptual paper on large river bioassessment*" (Annex II of this report).

3. Checking of compliance of national assessment methods with the WFD requirements (April 2010 + update in October 2010)

Do all national assessment methods meet the requirements of the Water Framework Directive? (Question 1 in the IC guidance)
Do the good ecological status boundaries of the national methods comply with the WFD normative definitions? (Question 7 in the IC guidance)

List the WFD compliance criteria and describe the WFD compliance checking process and results (the table below lists the criteria from the IC guidance, please add more criteria if needed)

Compliance criteria	Compliance checking conclusions
1. Ecological status is classified by one of **five classes** (high, good, moderate, poor and bad).	All methods are compliant.
2. High, good and moderate ecological status are set in line with the WFD's **normative definitions** (**Boundary setting procedure**)	All methods are compliant.
3. **All relevant parameters** indicative of the biological quality element are covered (see Table 1 in the IC Guidance). A **combination rule** to combine parameter 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.	All methods are compliant except for: ***Benthic Invertebrates*** - Spain: parameter *abundance* is not covered (but see explanation on page 5).
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	All methods are compliant.
5. The water body is assessed against **type-specific near-natural reference conditions**	All methods are compliant.
6. Assessment results are expressed as **EQRs**	All methods are compliant.
7. Sampling procedure allows for **representative** information about water body quality/ ecological status **in space and time**	All methods are compliant.
8. All data relevant for assessing the biological **parameters** specified in the WFD's normative definitions are covered by the **sampling procedure**	All methods are compliant.
9. Selected taxonomic level achieves adequate **confidence and precision** in classification	All methods are compliant.

Clarify if there are still gaps in the national method descriptions information.
Summarise the conclusions of the compliance checking:

Annex I of this report contains further descriptions of the national assessment methods for large rivers. Furthermore, relevant issues of large river bioassessment are summarised in the "Conceptual paper on large river bioassessment" (Annex II). ***Explanation on the compliance of the Spanish method*** The Spanish IBMWP that does not consider *abundance* was nevertheless already accepted for intercalibration in the MEDGIG exercise because the method was highly correlated with multimetric indices based on quantitative data (Munné & Prat 2009[2]). Furthermore, the IBMWP works adequately in Spanish rivers including the type of very large rivers (see demonstration of the pressure-impact relationship).

4. Methods' intercalibration feasibility check

Do all national methods address the same common type(s) and pressure(s), and follow a similar assessment concept? (Question 2 in the IC guidance)

4.1. Typology

What is the outcome of the feasibility evaluation in terms of typology? Are all assessment methods appropriate for the intercalibration water body types, or subtypes?

Method	Appropriate for IC types / subtypes	Remarks
Very large rivers were generally defined as running waters exceeding a total catchment area of 10,000 km^2. No typological differentiation was made for the intercalibration of national methods using benthic invertebrates. Due to the benchmark standardization applied to the common metrics prior to boundary comparison (see explanations in Section 6) typological differences were minimized. Diatom intercalibration distinguished between: R-L1: Very large low alkalinity rivers (<0.5 meq $CaCO_3$/l) R-L2: Very large medium to high alkalinity rivers (≥0.5 meq $CaCO_3$/l).		

[2] Munne A, Prat N (2009) Use of macroinvertebrate-based multimetric indices for water quality evaluation in Spanish Mediterranean rivers: an intercalibration approach with the IBMWP index. Hydrobiologia 10:203-225.

4.2. Pressures

Benthic Invertebrates

Austria

The Austrian multimetric index for large Alpine rivers was tested against a pre-classification for hydromorphological alterations (worst case approach, approx. 50 samples; see figure below).

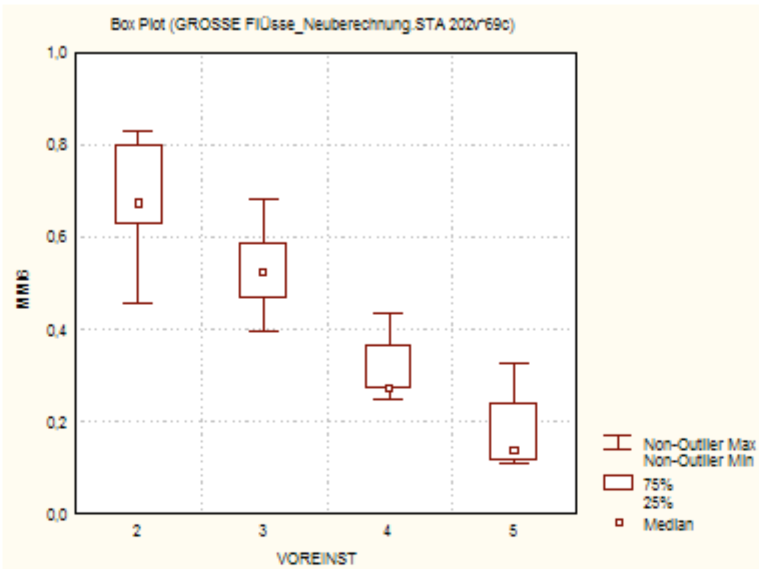


Figure: Ranges of Austrian multimetric index scores (MMI6) for sites pre-classified according to hydromorphological pressure

For Danube March and Thaya it was not possible to develop multimetric indices for general degradation due to (i) difficulties to define reference conditions, (ii) typological difficulties (large variety of different habitats) and (iii) difficulties to establish pressure-impact relationships (multitude of anthropogenic pressures, impacts and or positive effects are mask by dominance of alien species). To overcome (at least some of) the problems the following steps will be taken in the next years: (1) Floodplains: a new index will be developed during the next years, a research project has already been started; (2) Data availability: A sampling campaign in the Danube which will be carried out in 2012 (especially at least disturbed sites) will deliver more comparable data and will allow for the development of (a simple and pragmatic) assessment system.

Belgium (Flanders)

The Flemish macroinvertebrate index is assumed to respond to physical-chemical as well as hydromorphological pressures. The index uses the same metrics for all river types but with different reference values for each type. Because actual reference conditions are not available for Flanders, these reference values were set by a combination of data examination and expert judgment. For the very large river type, only a few sampling stations are available for Flanders, all belonging to the same river. It was therefore not possible to test the relation of the index with pressure parameters using a wide range of pressure values for this river type alone. It is however assumed that the response of the index to pressure parameters is similar for all river types. Using a large database with data from different river types, Gabriels et al. (2010)[3] found a positive correlation of MMIF with oxygen concentration (Spearman R=0.45, n=304) and with oxygen saturation (Spearman R=0.46, n=304) and a negative correlation with Kjeldahl nitrogen (Spearman R=-0.66, n=282), total nitrogen (Spearman R=-0.43, n=301), ammonium (Spearman R=-0.69, n=297), nitrite (Spearman R=-0.41, n=301), total phosphorous (Spearman R=-0.61, n=296), orthophosphate (Spearman R=-0.53, n=170), 5 day biochemical oxygen demand (Spearman R=-0.62, n=261) and chemical oxygen demand (Spearman R=-0.43, n=237) ($p<0.001$ in all cases).

Belgium (Wallonia)

The IBGA applied to Wallon large rivers responds to physical-chemical pressure. Gosselain et al. (2004)[4] demonstrate good correlations between the IBGA and the parameters : nitrogen compounds excluding nitrate, oxidisable organic matter, total phosphorus, suspended matter ($R^2 = 0.68$).

Estonia

Due to the availability of a constraint gradient of anthropogenic pressure existing for Estonian large rivers (Narva river only) a pressure-impact relationship could not be established. However, it is assumed that the pressure-response observable in small to medium-sized rivers is also valid for large rivers.

Finland

The method was intercalibrated in IC-Phase 1, where it correlated strongly (Pearson r = 0.837, p < 0.01) with the IC Common Metric (ICMi) that was shown to respond to a range of stressors (Buffagni et al. 2005[5]). The method thus has a very similar stressor-response as the ICMi. Due to lack of enough data points from very large rivers the method was not tested specifically for very large rivers, but given that species tolerances are not river type-specific, the method is expected to have a similar stressor-response as in smaller rivers.

Germany

The assessment method for large rivers in Germany (PTI) was validated based on 55 case-studies including data on river structure, current velocity, water quality and salinity. Validation results were evaluated against criteria of nature conservation. The PTI complied in 87 % of cases. For the remaining 13 % three additional metrics were considered to fully validate the assessment method (Schöll et al. 2005[6]). The pressure-impact relationship was not tested statistically but by expert judgement.

Hungary

Results of the Hungarian multimetric index for benthic invertebrates (HMMI_II) were tested against various chemical pressures (BOD, COD, N-NO_3, N-NO, TP, TN, conductivity, dissolved oxygen) using 45 sampling locations (including 194 samples). Spearman correlation coefficients ranged from 0.20 to 0.35 (see figure below).

[3] Gabriels W, Lock K, Pauw N De, Goethals PLM (2010) Multimetric Macroinvertebrate Index Flanders (MMIF) for biological assessment of rivers and lakes in Flanders (Belgium). Limnologica - Ecology and Management of Inland Waters 40:199-207.

[4] Gosselain V, Fauville C, Descy J-P (2004) PIRENE Programme Intégré de Recherche-Environnement-Eau - Corrélation entre qualité physico-chimique et qualité biologique. Facultés Universitaires Notre-Dame de la Paix, Namur. 26 pp.

[5] Buffagni A, Erba S, Birk S, Cazzola M, Feld C (2005) Towards European Inter-calibration for the Water Framework Directive: procedures and examples for different river types from the E.C. project STAR. Quad. Ist. Ric. Acque 123:1-467.

[6] Schöll F, Haybach A, König B (2005) Das erweiterte Potamontypieverfahren zur ökologischen Bewertung von Bundeswasserstraßen (Fließgewässertypen 10 und 20: kies- und sandgeprägte Ströme, Qualitätskomponente Makrozoobenthos) nach Maßgabe der EU-Wasserrahmenrichtlinie. Hydrologie & Wasserbewirtschaftung 49:234-247.

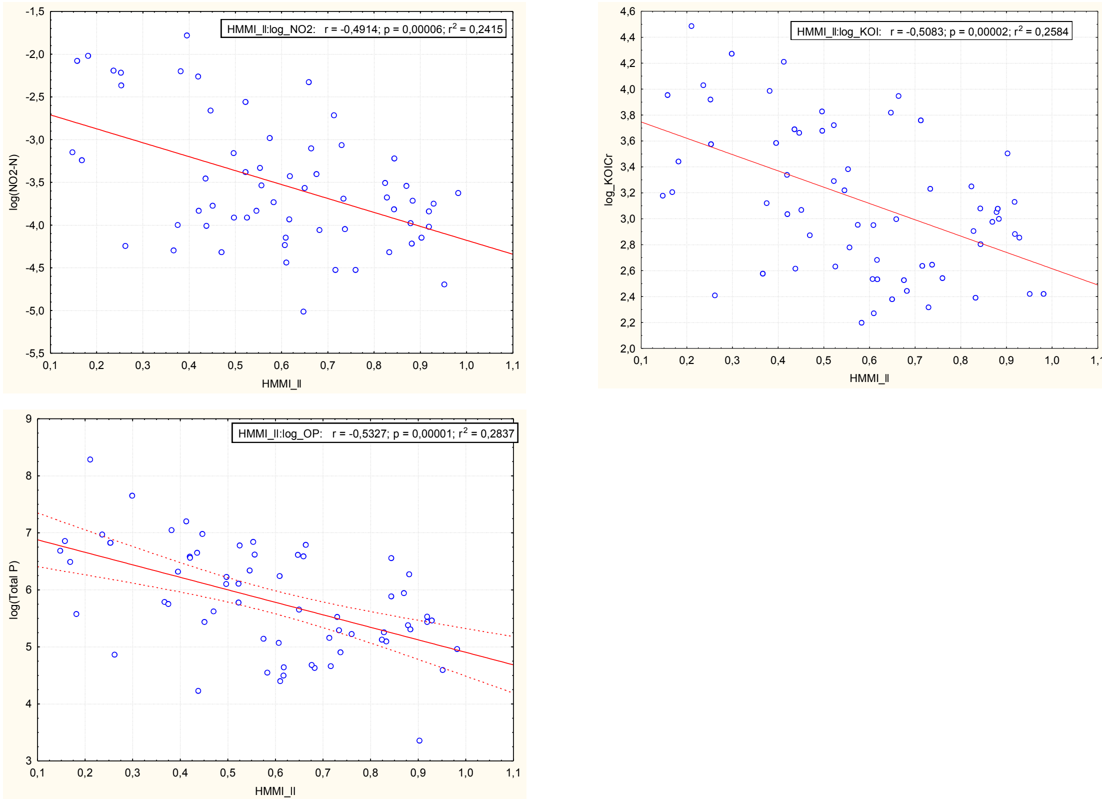


Figure: Pressure-impact relationships of selected chemical parameters (Nitrite, Chemical Oxygen Demand, total phosphorus) and the Hungarian Multimetric Index for Large Rivers (HMMI_II)

Lithuania

Log-linear analysis detected no relation between river types and Danish Stream Fauna Index (DSFI) using log-linear analysis (effect of adding two-way interaction, Chi-square <25, df=16, p>0.07). At least judging from the analysed data, river types thus do not systematically impact on the estimates of DSFI, and the DSFI is an appropriate measure to evaluate ecological status of benthic invertebrates in various running waters. Based on a dataset containing all Lithuanian monitoring sites (n=432) the DFSI showed highly significant relationships to chemical and hydromorphological pressures with an average correlation coefficient of r=0.24.

Netherlands

The Dutch National WFD metric for macroinvertebrates in rivers was shown to respond to pressures in small rivers. In the intercalibration process for small rivers the method was tested against the ICMi metric (Van Riel & Knoben 2006[7]). The national metric appeared to be better responding to hydromorphological pressures than chemical, which reflects the dominant situation in Dutch rivers. For the large rivers a small adaptation in the metric formula was made to give more weight to the presence of EPT-taxa in those rivers. However, in the Netherlands there is a lack of gradient for most pressures and consequently this adaptation could not be validated. But in the Netherlands hydromorphological alterations (shipping, bank

[7] Riel MC van, Knoben RAE (2006) The Dutch assessment of macroinvertebrates in international comparison. Analysis of the Dutch WFDi assessment method and comparison of ICM-metric scores of Dutch references with references from other Member States. Royal Haskoning, 's Hertogenbosch.

enforcement etc.) are considered the main pressure in large rivers and therefore it is assumed that these pressures are measured by the assessment method with sufficient precision.

Romania

The Romanian multimetric index for invertebrate-based assessment of large rivers was tested against selected pressure indicators based on 83 samples taken at nine large rivers in Romania (Arges, Danube, Ialomita, Jiu, Mures, Olt, Prut, Siret, Somes). Multiple regression analysis revealed significant relationships of the assessment method with percentage intensive land use in the catchment (beta = -0.35), orthophosphate concentration (beta = -0.25) and degree of channelization (beta = -0.22), resulting in 19 % explained variability of the assessment results.

Slovakia

Pressure-response of the Slovakian classification was guaranteed by the specific design of the multimetric index: Due to lack of reference sites in large rivers the individual metrics were selected based on the assumption that these metrics, the values of which change along an altitudinal gradient in running waters, are sensitive to anthropogenic stress. This assumption was supported by regression analysis of altitude and metrics from small streams (see figure below): metrics with a good ability to separate reference and monitoring sites did show a significant relationship to altitude (Sporka et al. 2009[8]).

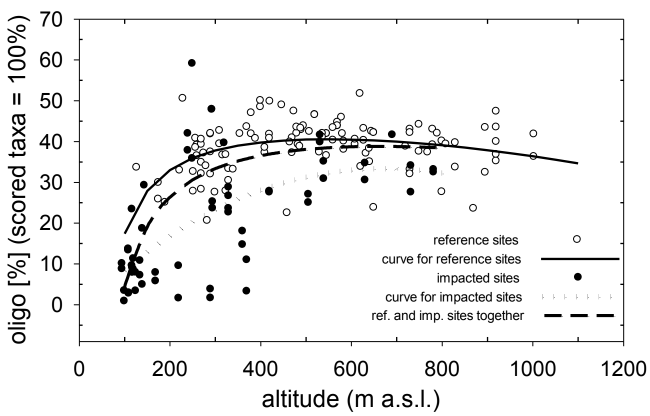
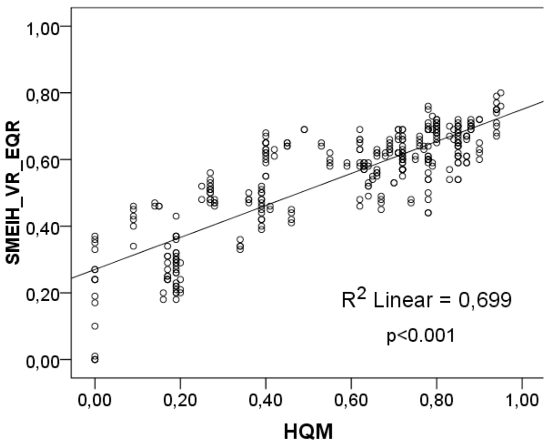


Figure: Relationship between altitude and the metric oligo (%) determined separately for reference, monitoring and mixed sites of small rivers (< 100 km²)

Slovenia

Large rivers in Slovenia are defined as rivers with a catchment area >2500 km^2 and/or mean annual discharge >50 m^3/s. Benthic invertebrates- as well as phytobenthos-based classification systems are modular with two modules. However, as benthic invertebrates are used to assess impact of two different pressure types (organic pollution and hydromorphological alterations) pressure-impact relationships were tested for each module separately (see Figures below).

[8] Šporka F, Pastuchová Z, Hamerlík L, Dobiašová M, Beracko P (2009) Assessment of running waters (Slovakia) using benthic macroinvertebrates — derivation of ecological quality classes with respect to altitudinal gradients. Biologia 64:1196-1205.



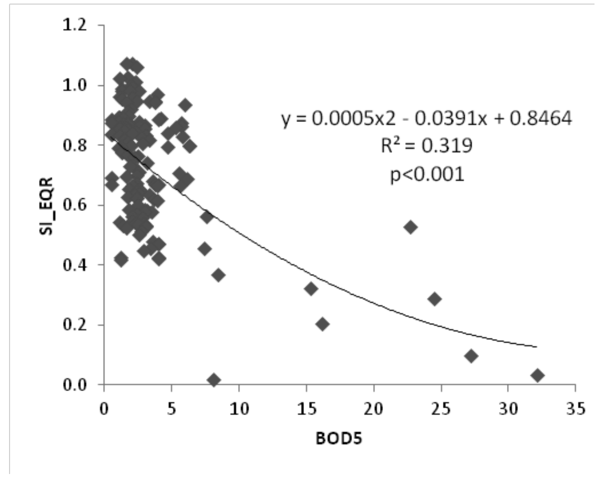


Figure (left): Regression plot of Habitat quality and modification index (HQM) against Slovenian multimetric index of hydromorphological alteration impact assessment in large rivers (SMEIH_VR_EQR) of Slovenia.

Figure (right): Regression plot of BOD5 against Saprobic Index (SI_EQR) in large rivers of Slovenia.

Spain

The response of the IBMWP to different types of anthropogenic pressures was validated by different authors (see Alba-Tercedor et al. 2002[9], Munné & Prat 2004, Sánchez-Montoya et al. 2010[10]). Although some of the studies carried out included in their databases very large rivers, a specific analysis for this river type was performed based on the XGIG Large River database: From a total of 25 anthropogenic pressures analyzed (including land use, hydromorphology and chemical water quality) 14 pressure-parameters showed significant correlations with the IBMWP (Spearman R, $p<0.05$). The 14 significant pressures included five hydromorphological parameters, three parameters covered types of land use, and six parameters addressed chemical water quality. The hydromorphological and chemical pressures included single parameters as well as combined pressure-parameters. Selected pressure-impact relationships are depicted in the figure below.

[9] Alba-Tercedor J, Jáimez-Cuéllar P, Álvarez M et al. (2004) Caracterización del estado ecológico de los ríos mediterráneos ibéricos mediante el índice IBMWP (antes BMWP'). Limnetica 21:175-185.

[10] Sánchez-Montoya MM, Vidal-Abarca MR, Suárez ML (2010) Comparing the sensitivity of diverse macroinvertebrate metrics to a multiple stressor gradient in Mediterranean streams and its influence on the assessment of ecological status. Ecological Indicators 10:896-904.

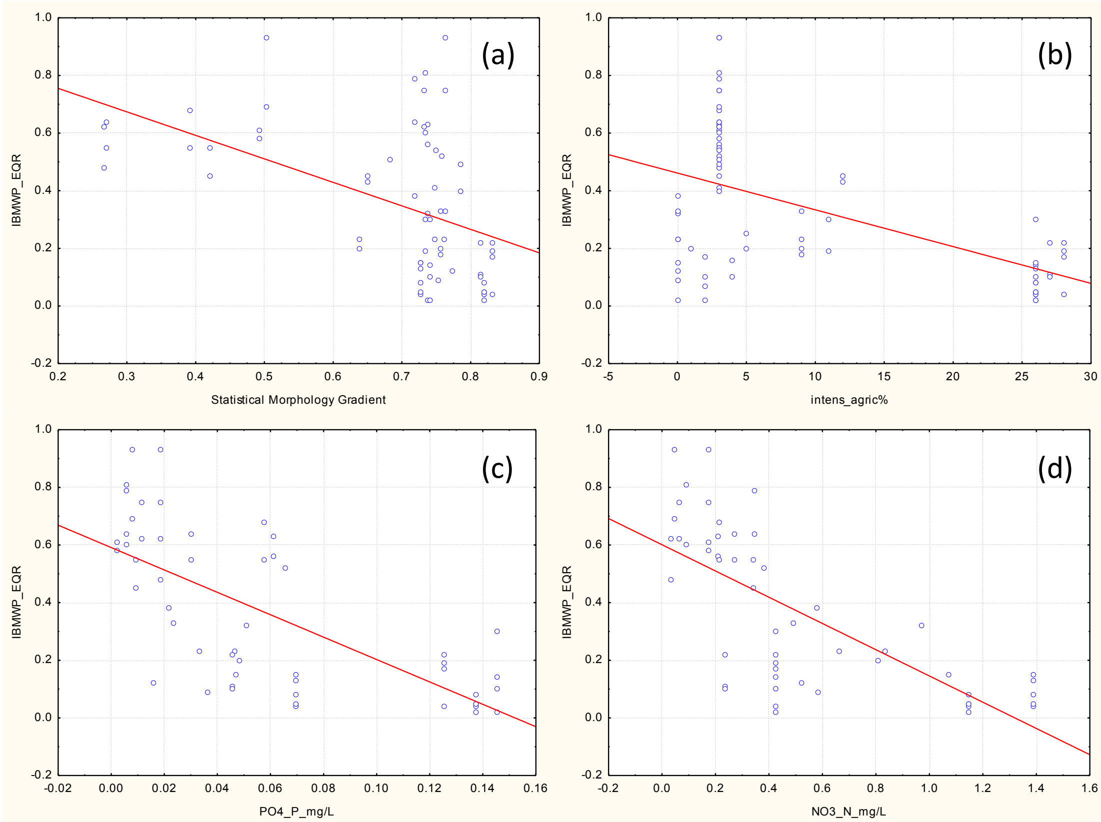


Figure: Response of IBMWP expressed as EQR to different pressure types: a) global hydromorphological pressure; b) land use pressure; c) and d) water quality pressures (phosphates and nitrates, respectively)

Phytobenthos (benthic diatoms)

Austria

The assessment method for phytobenthos is applied to all river sizes (including large rivers). Reference conditions - based on trophic reference status - have been defined for different river types. The method consists of the following stressor-specific modules: (i) Trophic Index (TI): nutrients, (ii) Saprobic Index (SI): organic pollution, (iii) Reference Species Index (RI): other stressors like acidification, salinity, toxicity etc.

The following graph shows the general relation between TI and Orthophosphate, large rivers are included in this gradient (although slope might be different for large rivers only).

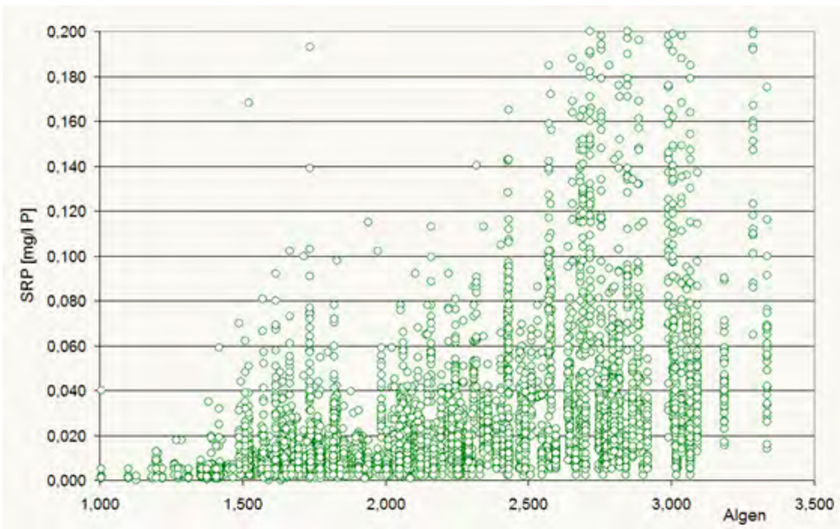


Figure: Relationship of Orthophosphate (SPR) and Trophic Index (Algen)

Belgium (Flanders)

At high status, the relative abundance of impact-sensitive diatoms is not reduced from what can be expected for the type in unimpacted conditions. Impact-sensitive taxa are listed for each water type, separately. Each list includes those diatom taxa that have been reported from water courses in the BE-FL region and for which the relative abundance decreases distinctly if at least one of the pressures affecting the respective water type increases. The minimum relative representation of these taxa corresponding to high status for each water type is set by expert judgment and examination of data from other Member States. Impact-sensitive and indifferent taxa dominate at high status, whilst the abundance of impact-associated taxa remains very limited. At good status, the relative abundance of impact-associated diatoms is not higher than what can be expected for the type with slight human impact. The maximum relative representation of these taxa corresponding to good status is estimated from its relation to pressure-related variables in smaller river types. The good/moderate boundary relates to occurrence of average BOD values above 4 mg l^{-1}, indicating that self-purification capacity is exceeded. Impact-sensitive and indifferent taxa dominate at good status and the abundance of impact-associated taxa remains limited. Due to poor data availability for very large rivers, current boundary values for the representation of indicative taxa at class boundaries are considered to be the same as those used for large rivers. The pressures addressed by the EQR for very large rivers would be mainly eutrophication, organic pollution, salinisation and reservoir construction but lack of data prevents examination of type-specific relationships.

Belgium (Wallonia)

The IPS applied to Walloon large rivers responds to physical-chemical pressure. Gosselain et al. (2004)[4] demonstrate good correlations between the IBGA and the parameters : nitrogen compounds excluding nitrate, oxidisable organic matter, total phosphorus, suspended matter ($R^2 = 0.70$).

Czech Republic

The response of the Czech (Saprobic-Trophic) Index to abiotic quality parameters indicating organic pollution or nutrient enrichment was tested based on approx. 750 samples, including 110 samples from rivers of Strahler stream order 8 and 9 (i.e. large rivers). The relationships were highly significant and showed correlation coefficients ranging from 0.6 to 0.75 (see figure).

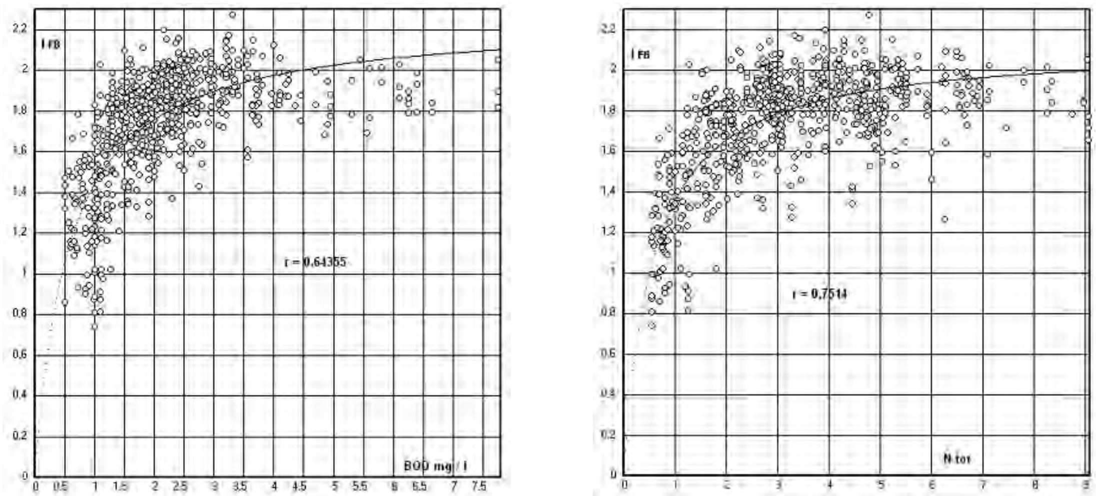


Figure: Regression plot of Biological Oxygen Demand (left) and Total Nitrogen (right) against the Czech (Saprobic-Trophic) Index (I_{FB}).

Estonia

The pressure-impact relationship of the Estonian diatom assessment method for large rivers was validated by expert judgement. As the method's pressure-response relationship was detected at the small and medium rivers, it was suggested that it works also for the large rivers: Ecological data from 139 river reaches were examined to establish pressure-impact relationship between diatom indices and eutrophication gradient. The relationship between three indices (IPS, TDI and Watanabe index) and TP (measured in summer during low water period) showed significant correlation (Spearman correlation ranging from 0.28 to 0.55; $p < 0.001$).

Finland

The method is not tested specifically for very large rivers, but the response of the method is expected to be similar as in smaller rivers. The method was developed so that the rivers studied were classified to five classes according to the degree of human impacts in the drainage basin in general or near the sampling station. Rivers with more or less natural state of very low degree of human impacts showed IPS-index values > 16, whereas those with slight human impact had the IPS from 14 to 16. The index values decreased markedly with increasing strength of human impact (Eloranta & Soininen 2002[11]).

[11] Eloranta P, Soininen J (2002) Ecological status of some Finnish rivers evaluated using benthic diatom communities. Journal of Applied Phycology:1-7.

France

The assessment method for phytobenthos is applied to all river sizes (including large rivers). A recent work on a data set comprising 960 sites was carried out by Villeneuve et al. (2011)[12], and provides the following conclusions regarding the IBD2007: The significant variables that have a negative effect (shown in red in the figure below) are herbicides, metals, BOD_5, ammonium, nitrite, nitrate, total phosphorus, orthophosphate, organic carbon, percentage of urbanization in the basin, intensive agriculture in the basin, low impact agriculture in the basin, the ratios of artificial agriculture in the area, the ratio of drainage, the rate of dikes in the riverbed, the sediment trapping rate in the flood plain and the population of the watershed. Note that IBD2007, however, is not known to react with toxic pressures, so it is likely that the significant effect of herbicides such as metals are artefacts relating with the percentage of agricultural areas. The significant variables that have a positive effect (in blue in the figure) are dissolved oxygen and oxygen saturation, percentage of natural areas in the basin, communication channels in the riparian buffer, and vegetation in the floodplain. The Partial-Least-Squares (PLS) regression of IBD2007 based on the complete set of pressures shows a determination coefficient of 45%. So all the pressures explain 45% of the variability of the index.

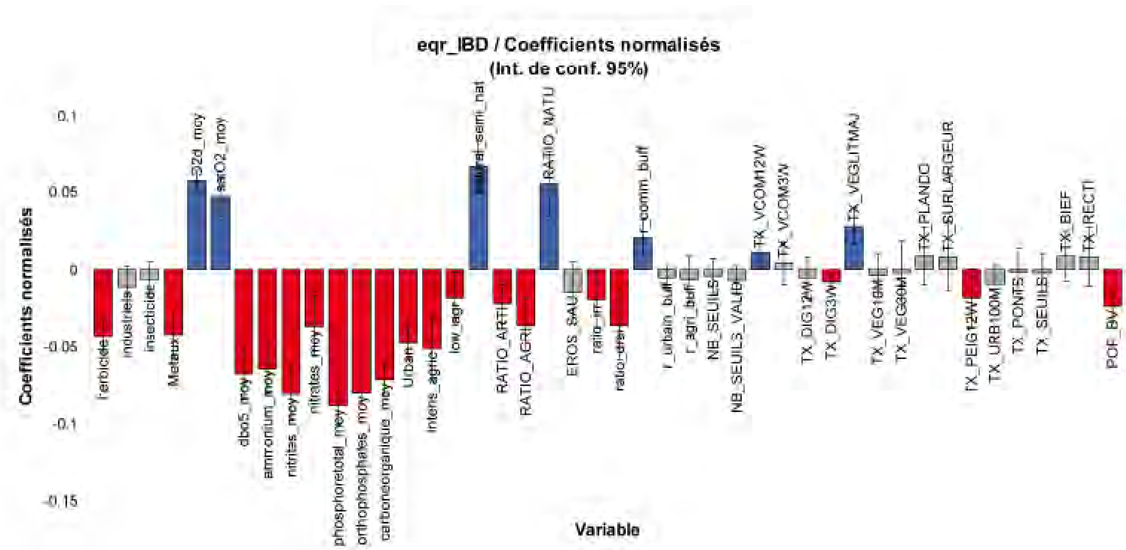


Figure: Normalized coefficients from the PLS – EQR-IBD2007 regression = f(predictive variables) (Villeneuve et al. 2011)

Germany

The German module for the diatom assessment of large rivers was tested against selected chemical parameters based on 25 samples taken at large rivers in Germany (Elbe, Danube, Mosel). Multiple regression analysis revealed strong relationships of the assessment module with orthophosphate (beta = -0.57), oxygen concentration (beta = -0.52) and ammonium (beta = -0.33) in the water, resulting in 51.5 % explained variability of the assessment module.

[12] Villeneuve B, Ferréol M, Valette L., Bougon N, Tornos T (2011) Extrapolation spatiale de l'état écologique des masses d'eau et modèles diagnostics. Rapport technique, version provisoire. Pôle hydroécologie des cours d'eau Cemagref-Irstea/Onema, 39 p.

Hungary

The pressure-impact relationship of the Hungarian diatom assessment metrics was tested against chemical parameters including large rivers (Várbíró et al. 2011[13]). Hungary distinguishes between large and very large, low altitude rivers with fine sediments (Group 5) and "Danube type" (Group 6). For group 5 rivers the diatom method was significantly correlated with conductivity (r=-0.26) and orthophosphate (r=-0.31). Group 6 rivers showed a correlation coefficient of r= -0.23 with orthophosphate.

Netherlands

The Dutch metric for large rivers is a copy of the one developed for the medium rivers (e.g. R-C4 or R-C5). The metric is based on the international common metric (IPS and TI) and the reference conditions are based on the values of other Member States in the GIG. In general, the metric shows a good relationship with nutrient pressure. In the Netherlands the metric has a moderate relationship with nutrients (high status and bad status sites are missing), e.g. R^2 of phytobenthos metric and total phosphorus range between 0.18 to 0.42. In the Netherlands we have only three large rivers and the history of measuring phytobenthos is too short for the development of a pressure-impact relationship. Such a relationship would need some spread in e.g. nutrient values. Because of the lack of data, we have decided to use the same metric of medium rivers in the large rivers as well. This decision was supported by the application of the metric to the large rivers. Recently, a project was started to evaluate the metrics, and decided to investigate the improvements of the phytobenthos metric for the application on large rivers, including tidal fresh water rivers. We expect to incorporate improvements, if needed, of the large rivers' phytobenthos metric before the next River Basin Management Plan (2015).

Slovakia

The pressure-impact relationship of the Slovakian diatom assessment metrics was tested against chemical, hydromorphological and land-use parameters based on a data covering various river types including large rivers[14]. The parameters responded significantly to the individual metrics in a multiple regression analysis with r-square values ranging from 0.37 to 0.44. Since the relationship was not individually validated for large rivers it is assumed that the results also apply to this river type.

[13] Várbíró G, Borics G, Csányi B, Fehér G, Grigorszky I, Kiss KT, Tóth A, Ács É (2011) Improvement of the ecological water qualification system of rivers based on first results of the Hungarian phytobenthos surveillance monitoring. Hydrobiologia (in press).

[14] Hlúbiková D, Hindáková A, Haviar M, Miettinen J (2007) Testing of diatom water quality indices in monitoring and reference sites of Slovakian rivers (Central Europe). Archiv für Hydrobiologie, Suppl. Large Rivers 17:443-464.

Slovenia

To demonstrate the pressure-impact relationship of the Slovenian phytobenthos method the phytobenthos-based EQR values were regressed against the pressure (see Figure below).

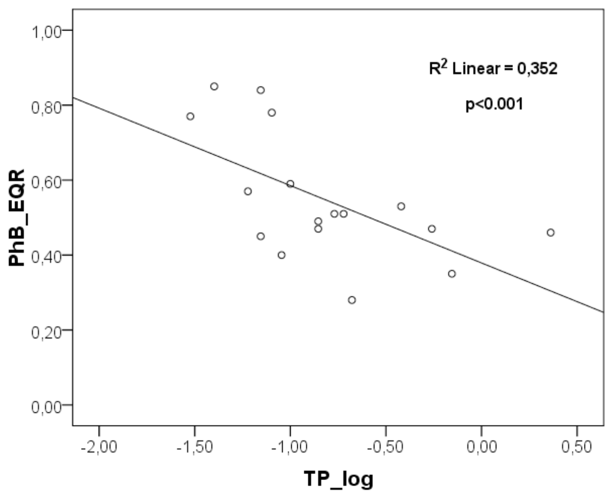


Figure: Regression plot of total phosphorus (log transformed data – TP_log) against phytobenthos-based Slovenian national EQR values (PhB_EQR) in large rivers of Slovenia.

Spain

The response of the anticipated Spanish diatom method for large rivers (IPS) to different types of anthropogenic pressures was assessed. IPS had a significant response to only two pressures taken into account in the Intercalibration exercise of the XGIG Large Rivers intercalibration exercise: percentage of intensive agriculture land use (Spearman R = -0.40, P = 0.04) and the synthetic index of water chemical quality (Spearman R = 0.53, P = 0.02). Moreover, in the case of the synthetic gradient, the response of IPS was not the expected (higher chemical contaminations should correspond to lower IPS, and vice versa). Therefore, **Spain will not intercalibrate the IPS applied to Spanish very large rivers**.

Sweden

It is assumed that the pressure-impact relationship of the Swedish diatom method for large rivers is comparable to the relationship observable at small and medium-sized rivers. In the figure below, for instance, the response of Diatom assessment metric IPS of large rivers fits the rest of the Swedish diatom database correlated with Total Phosphorus.

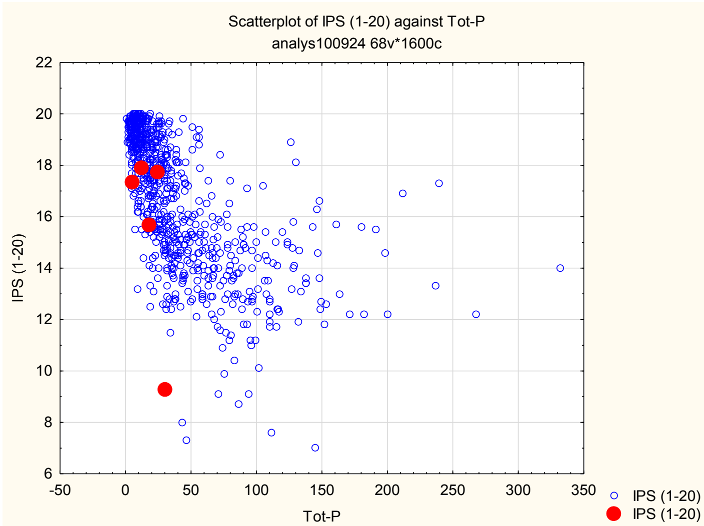


Figure: The response of Diatom IPS of large rivers (red dots) fit the rest of the Swedish diatom database (blue dots, n=1180 streams) when correlated against Total P (R^2 = 0.40)

Conclusion

The national methods using benthic invertebrates mainly indicate the effects of combined stressors (organic pollution/eutrophication, morphological degradation). This was demonstrated by several Member States using empirical pressure-impact analyses. Other methods are sensitive to general degradation, i.e. multiple pressures. The Combined Abiotic Pressure gradient (including parameters of physico-chemical and morphological pressure, see Section 6.2) used in the intercalibration analysis was significantly correlated with the common intercalibration metric (see Section 7.2). The correlation of national methods and the common metric can be regarded as an indirect empirical testing of their pressure-impact relationship (i.e. Estonia, Finland, Germany, Hungary, Slovenia and Spain).

All diatom-based methods for very large rivers indicate the effects of eutrophication and/or organic pollution.

4.3. Assessment concept

Do all national methods follow a similar assessment concept?

Examples of assessment concept:

- **Different community characteristics** - structural, functional or physiological - can be used in assessment methods which can render their comparison problematic. For example, sensitive taxa proportion indices vs species composition indices.
- Assessment systems may focus on **different lake zones** - profundal, littoral or sublittoral - and subsequently may not be comparable.
- Additional important issues may be the **assessed habitat type** (soft-bottom sediments versus rocky sediments for benthic fauna assessment methods) or **life forms** (emergent macrophytes versus submersed macrophytes for lake aquatic flora assessment methods)

The existing national assessment methods acquire their biological data from the main river channel and are based on concepts similar to the assessment of smaller rivers. Although the specific features of large rivers may require alternative, ecologically adapted classifications, the intercalibration exercise deals with the harmonization of the assessment methods that are currently used by the Member States.

5. Collection of IC dataset

Describe data collection within the GIG.
This description aims to safeguard that compiled data are generally similar, so that the IC options can reasonably be applied to the data of the Member States.

Make the following table for each IC common type

a) Benthic invertebrates

Country	Number of samples	Number of water bodies	Number of pressure data[a]
Austria	5	5[b]	5
Belgium (Flanders)	3	3	3
Belgium (Wallonia)	2	2	2
Germany	67	11	67
Estonia	8	3	8
Spain	75	25	57
Finland	10	5	10
Hungary	29	7	27
Lithuania	16	4	16
Netherlands	61	15	61
Romania	83	23	83
Slovenia	31	10	31
Slovakia	48	3	45
Total number	**438**	**116**	**415**

[a] Records of the Combined Abiotic Pressure gradient used in benchmark standardisation of common metrics
[b] Comprising the rivers Drava, Inn and Mura

b) Benthic diatoms

Subtype	Country	Number of samples	Number of water bodies	Number of pressure data[c]
Low alkalinity rivers (R-L1)	Finland	26	7	19
	Sweden	4	4	4
	Total number	**30**	**11**	**23**
Medium- to high alkalinity rivers (R-L2)	Austria	14	11	14
	Belgium (Flanders)	3	3	3
	Belgium (Wallonia)	4	2	4
	Czech Republic	24	7	24
	Estonia	5	2	5
	Germany	24	12	24
	Hungary	65	16	26
	Netherlands[d]	92	10	92
	Slovakia	40	3	37
	Slovenia	26	11	26
	Total number	**297**	**77**	**255**

[c] Records of average PO_4-P concentration used in benchmark standardisation of common metrics
[d] The Netherlands provided additional data after the completion of IC milestone version 1.0. Their boundaries thus did not contribute to the harmonisation guideline but were compared afterwards against the average boundary position established by the eight countries listed in Table 8.2c.

List the data acceptance criteria used for the data quality control and describe the data acceptance checking process and results

Data acceptance criteria	Data acceptance checking
Data requirements (obligatory and optional)	All data originated from national monitoring programmes of very large rivers that comply with pertinent QA/QC schemes. Furthermore, benchmark standardization allowed for the combination of the heterogeneous datasets.
The sampling and analytical methodology	
Level of taxonomic precision required and taxalists with codes	
The minimum number of sites / samples per intercalibration type	
Sufficient covering of all relevant quality classes per type	
Other aspects where applicable	

6. Benchmarking: Reference conditions or alternative benchmarking

In section 2 of the method description of the national methods above, an overview has to be included on the derivation of reference conditions for the national methods. In section 6 the checking procedure and derivation of reference conditions or the alternative benchmark at the scale of the common IC type has to be explained to ensure the comparability within the GIG.

Clarify if you have defined
- common reference conditions (N)
- or a common alternative benchmark for intercalibration (Y)

6.1. Reference conditions

Does the intercalibration dataset contain sites in near-natural conditions in a sufficient number to make a statistically reliable estimate? (Question 6 in the IC guidance)

- Summarize the common approach for setting reference conditions (true reference sites or indicative partial reference sites, see Annex III of the IC guidance):

Not applicable

- Give a detailed description of **reference criteria** for screening of sites in near-natural conditions (abiotic characterisation, pressure indicators):

Not applicable

- Identify the **reference sites** for each Member State in each common IC type. Is their number sufficient to make a statistically reliable estimate?

Not applicable

- Explain how you have screened the biological data for impacts caused by pressures not regarded in the reference criteria to make sure that true reference sites are selected:

Not applicable

- Give detailed description of **setting reference condition**s (summary statistics used)

Not applicable

6.2. Alternative benchmarking (only if common dataset does not contain reference sites in a sufficient number)

- Summarize the common approach for setting **alternative benchmark** conditions (describe argumentation of expert judgment, inclusion of modelling)

Continuous benchmarking: Theoretical background

Benchmarking of national assessment methods is an important precondition for the comparison and harmonisation of ecological status class boundaries in intercalibration. National boundaries are expressed as relative deviations from reference conditions at which the aquatic biota show no (or insignificant) impact from anthropogenic disturbance. These reference conditions are defined differently for individual assessment methods. Intercalibration thus requires the standardisation of assessment methods using common benchmarks. Data from reference sites meeting harmonised criteria, for instance, provide such benchmarks. Since these data are generally scarce for most European water types, alternative benchmarking based on sites impacted by similar (low) levels of anthropogenic pressure (i.e. benchmark sites) can be applied.

However, both options rely on the homogeneous distribution of undisturbed or similarly disturbed sites among countries within a common type. If certain countries feature more pronounced anthropogenic disturbance than others, the common benchmark is imbalanced or unachievable. Balancing the selection of benchmark sites will be a common problem if, for instance, countries featuring contrasting population densities or land use practices, like Austria and Bulgaria, are involved in the same exercise.

In such cases the approach of "**continuous benchmarking**" represents a solution (Figure 6.1). Continuous benchmarking refers to data from sites featuring various levels of disturbance and thus accounts for the actual data availability of individual countries. However, all countries need to provide the same (set of) pressure-variables for selected sites along with the biological data. Depending on the available data countries may cover different parts of the pressure gradient ranging from undisturbed to highly perturbed conditions. This makes continuous benchmarking different from the approaches described above: the standardisation of national assessment methods is carried out not using a preselected small range of the gradient but the full continuum.

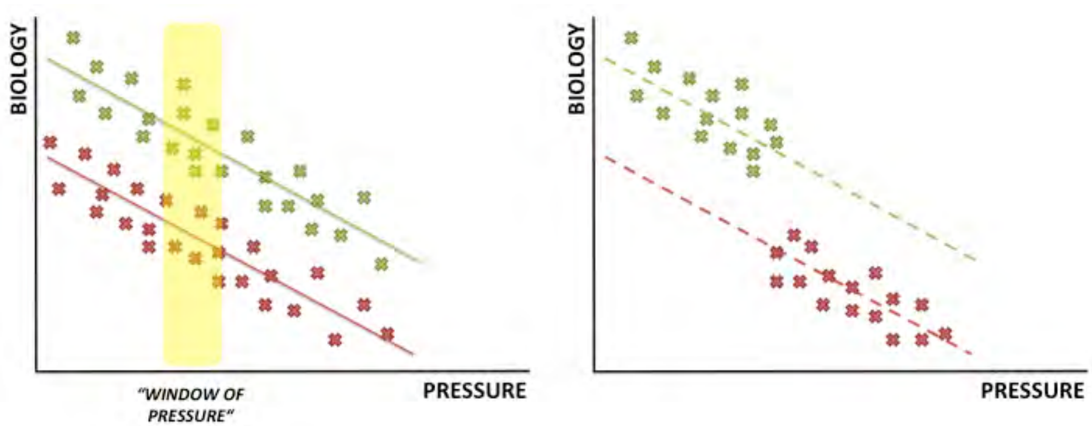


Figure 6.1: The two basic approaches of benchmarking in intercalibration

Left: "Benchmark sites" approach: Sites at similar (low) level of pressure ("window of pressure") are available from all countries (red and green).

Right: "Continuous benchmarking" approach: Countries' data (red and green) cover different levels of pressure.

The general aim of benchmark standardisation in intercalibration is to identify and remove differences among national assessment methods that are not caused by anthropogenic pressure, but by systematic discrepancies due to different methodology, biogeography, typology etc. If such differences are ignored they may have an overriding effect on the comparability exercise. Therefore, we controlled the pressure effects on the common metrics to disclose any remaining discrepancies. We used the Combined Abiotic Pressure gradient (see below) for the intercalibration of invertebrate-based methods, log-transformed PO_4-P values for diatom-based methods.

The standardisation applied in our exercise removed the effects of pressure that differed between countries by offset correction for most metrics and factor correction for the two taxa-number candidate metrics of the multi-metric invertebrate index. To obtain the correction values we used a linear mixed effects model in which the common metric was the dependent variable, the pressure variable(s) formed the covariates and the country (i.e. national dataset) was a fixed factor. The model yielded the offsets (i.e. intercepts) of the common metric values obtained from each national dataset, compared against the global relationship of pressure and the common metric (Figure 6.2). These offsets were then used to standardise the single common metrics individually for each national dataset, subtracting the country-specific offset value from the common metric calculated from the respective national dataset. For number of EPTCBO taxa and number of EPT taxa a factor was derived from the offset for standardisation. For the correction of invertebrate metrics we distinguished between "minor" large rivers (catchment size < 30,000 km^2, average annual discharge: ≤ 300 m^3/s) and "major" large rivers (catchment size ≥ 30,000 km^2, average annual discharge: > 300 m^3/s), as analyses revealed significant differences in metric responses between these subtypes (see Annex V).

In a second step of standardisation the same procedure was applied to the national EQRs: By statistical modelling we identified the country-specific offsets of the national EQRs against the ICM composed of benchmark-standardised single common metrics. Applying the offsets to the national EQRs did not effect the outcome of the usual boundary comparison procedures, but allowed for combining these data into a global regression of benchmark-standardised EQRs against the ICM.

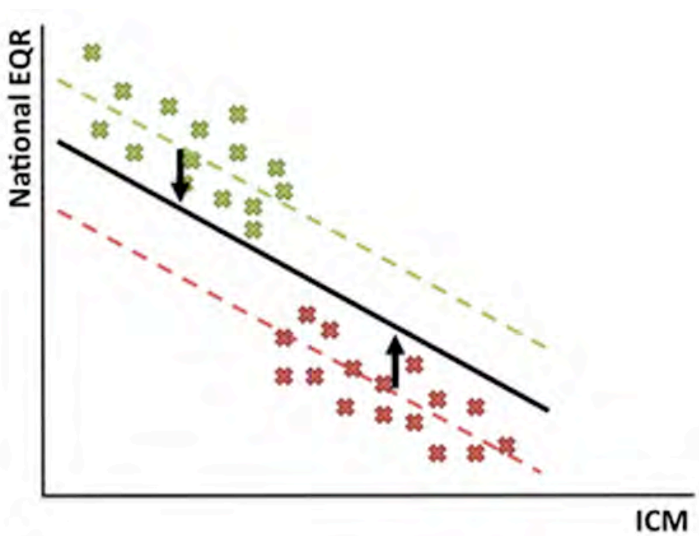


Figure 6.2: Adjustment of national EQRs (red and green) based on offset results from the statistical model

- Give a detailed description of **criteria** for screening of **alternative benchmark** sites (abiotic criteria/pressure indicators that represent a similar low level of impairment to screen for least disturbed conditions)

Continuous benchmarking was carried out based on the following anthropogenic pressure gradients: Orthophosphate (PO_4-P) measurements were available from the water bodies that were sampled for benthic diatoms. These data were aggregated to annual average concentrations, resulting in a total number of 410 PO_4-P data records matched to the biological samples. The invertebrate samples were matched with 415 data records of eight abiotic parameters acquired at water body level: navigation intensity, influence of damming, influence of impoundment, degree of water abstraction, degree of riparian habitat alteration, degree of channelization[15], average annual PO_4-P concentration and average annual nitrate (NO_3-N) concentration. The parameter values were standardised and averaged to gain the Combined Abiotic Pressure gradient (CAP).

- Identify the **alternative benchmark sites** for each Member State in each common IC type

See table in Section 5 *Collection of IC dataset*

- Describe how you validated the selection of the alternative benchmark with biological data

Not available

- Give detailed description how you identified the position of the alternative benchmark on the gradient of impact and how the deviation of the **alternative benchmark** from reference conditions has been derived

Not available

Describe the **biological communities** at reference sites or at the alternative benchmark, considering potential biogeographical differences:

Not available

[15] Navigation intensity of commercial transport (no / low / high)
Upstream dams' influence (yes / no)
Influence of impoundment causing flow velocity reduction (no / slight / strong, i.e. clear biological effects)
Degree of water abstraction (sites not affected by water flow alteration / sites slightly affected, i.e. less than 10% of the median annual flow and the median monthly flow during a critical period, e.g. low flow period / sites significantly affected, i.e. more than 10% of median annual flow and median monthly flow during a critical period, e.g. low flow period / sites strongly affected, i.e. more than 50% of median annual flow and median monthly flow during a critical period, e.g. low flow period
Degree of riparian habitat alteration = direct alteration of the riparian vegetation, i.e. adjacent natural vegetation appropriate to the type and geographical location of the river (no direct alteration, i.e. adjacent natural vegetation appropriate to type and geography / slight alteration / strong alteration / anthropogenic complete loss of riparian vegetation)
Degree of channelization = assessment of embankments, cross section alterations and flow velocity increase (no canalisation, no alteration of the "natural" cross section, i.e. no "hard works" affecting the whole river. No flow velocity increase / slight alteration, i.e. less than 10% of the segment affected by "hard works". No flow velocity increase / significant alteration, i.e. a main part of the segment is affected by "hard work". Flow velocity increase / strong alteration, i.e. straightened river, technical-U-profile section. Flow velocity increase)

7. Design and application of the IC procedure

7.1. Please describe the choice of the appropriate intercalibration option.

The ecological status boundaries of the national assessment methods were compared by means of biological common metrics that were widely applicable to the available data, highly correlated with the national assessment methods and showed good response to the anthropogenic pressures acting at the water bodies. The diatom intercalibration was divided into two separate exercises: (i) low alkalinity rivers shared by Finland and Sweden and (ii) medium- to high-alkalinity rivers shared by all other countries (see Annex VI). This distinction was made to ensure comparable patterns of pressure-impact relationships among water bodies.

To account for systematic differences between the national datasets the individual common metrics were standardised against the pressure gradient ("benchmark standardisation") and then combined into a multimetric index (ICM). A second standardisation comprised the national EQR values that enabled the combination of individual national assessment results into a global regression analysis. These analytical steps are outlined in Figure 7.1. The intercalibration analysis concluded with the actual comparison and harmonisation of national class boundaries of the good ecological status.

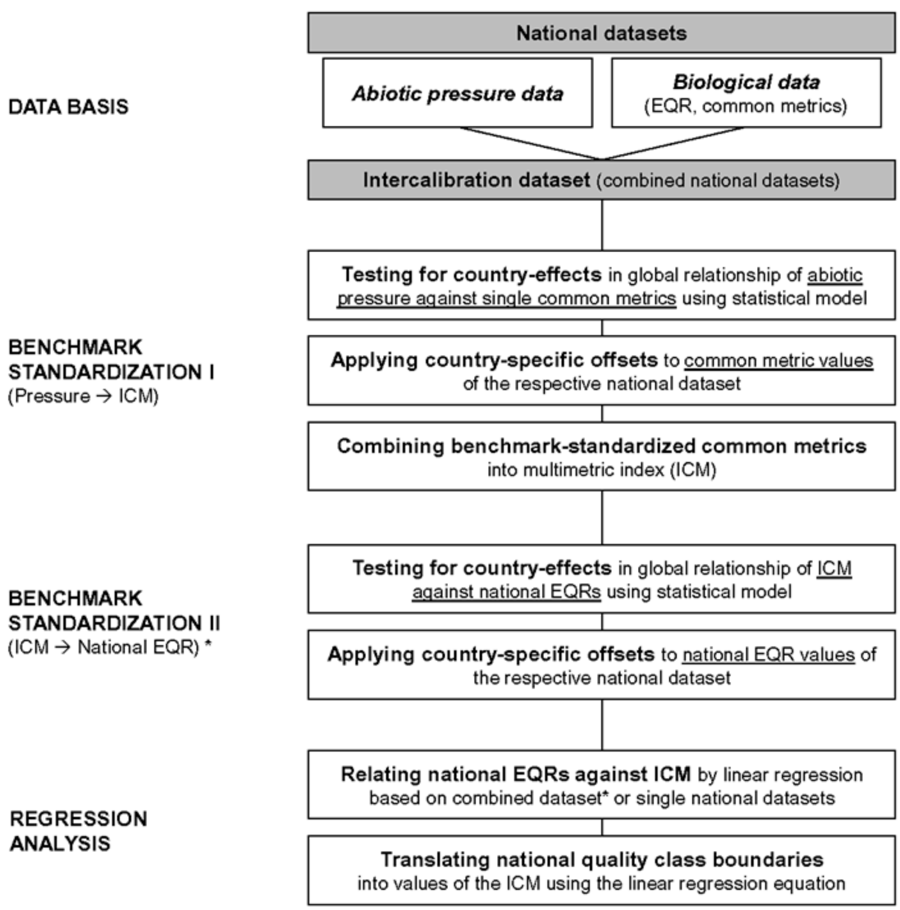


Figure 7.1: Overview of the analytical procedure to compare the national boundary positions on the common metric scale

* not done for selected national diatom datasets (see text for explanations)

7.2. IC common metrics

Describe the IC Common metric:

To identify suitable biological common metrics for the intercalibration of invertebrate-based assessment methods we selected eight core-metrics from an array of 117 candidate metrics based on correlation analyses with the national methods and the Combined Abiotic Pressure gradient (CAP). These core-metrics were benchmark-standardised according to the procedure described in Section 6.2, then normalised using the global 10th and 90th percentile values obtained from the combined national datasets. We calculated 27 variants of ICMs averaging four metrics at a time, and tested the ICMs' relationship to the national EQRs and the CAP. The best performing ICM was used in the intercalibration analysis (iICM = invertebrate ICM).

Since the national diatom assessment of very large rivers is very similar to small- and medium-sized streams we selected the dICM (diatom ICM) described by Kelly et al. (2009)[16]. This ICM averaged the benchmark standardised and normalised values of the metrics IPS and RT. The diatom metrics were also normalised using the global 10th and 90th percentile values obtained from the combined national datasets.

We used the following eight core-metrics to compose 27 invertebrate-based ICM variants based on benchmark-standardised values: (i) ASPT, (ii) Rhithron-Type-Index, (iii) % EPT individuals (abundance classes), (iv) number of EPTCBO taxa, (v) number of EPT taxa, (vi) Potamon-Type-Index, (vii) % individuals with active aerial dissemination strategy, (viii) % individuals of holoaquatic taxa.

The correlation of the ICM variants and CAP ranged from r=-0.37 to r=-0.49. The average relationship of the ICMs with the national EQRs showed a range from r=0.51 to r=0.60. The ICM composed of the core-metrics (i) to (iv) was among the best performing multimetric indices. Therefore, we used this index in the further analysis (i.e. iICM). The dICM calculated for the combined dataset showed a Pearson correlation coefficient of r=-0.67 with the log-transformed PO_4-P values. Its average correlation to the national EQRs was r=0.79 (Flemish dataset excluded). Several invertebrate-methods showed non-significant correlations with the iICM due to short gradients and/or low number of data values: *Belgium (Flanders)*, *Belgium* (*Wallonia*), *Lithuania*, *Slovakia*. *Austria* (*me2*) and *Belgium* (*Flanders*) were insignificantly correlated to the diatom-ICM (see table below). Among the significantly related invertebrate methods, *Belgium* (*Flanders*), *Romania*, *the Netherlands* and *Slovakia* showed r-values below 0.50. None of the diatom-correlations (except the Flemish) fell below r=0.50.

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

Are all methods reasonably related to the common metric(s)? (Question 5 in the IC guidance)

Please provide the correlation coefficient (r) and the probability (p) for the correlation of each method with the common metric (see Annex V of IC guidance).

a) Benthic invertebrates

National method	r	p
Austria	0.887	0.045
Belgium (Flanders)	-0.095	0.939
Belgium (Wallonia)	n.a.[a]	
Estonia	0.878	0.004
Finland	0.768	0.010
Germany	0.639	<0.001
Hungary	0.537	0.003
Lithuania	0.523	0.055
Netherlands	0.315	0.013
Romania	0.392	<0.001
Slovakia	0.290	0.051
Slovenia	0.857	<0.001
Spain	0.863	<0.001

[a] not available due to low number of data points in the analysis

b) Benthic diatoms

National method	r	p
Austria – me1	0.814	0.008
Austria – me2	0.877	0.051
Belgium (Flanders)	-0.625	0.570
Belgium (Wallonia)	0.986	0.014
Czech Republic	0.819	<0.001
Estonia	0.698	0.006
Finland	0.940	<0.001
Germany	0.546	0.005
Hungary	0.767	<0.001
Netherlands[b]	0.946	<0.001
Slovakia	0.796	<0.001
Slovenia	0.702	<0.001
Sweden	0.965	0.008

[b] The Netherlands provided additional data after the completion of IC milestone version 1.0. Their boundaries thus did not contribute to the harmonisation guideline but were compared afterwards against the average boundary position established by the eight countries listed in Table 8.2c.

Explain if any method had to be excluded due to its low correlation with the common metric:

See Table 8.1.

8. Boundary setting / comparison and harmonization in common IC type

Clarify if
- boundaries were set only at national level (Y)
- or if a common boundary setting procedure was worked out at the scale of the common IC type (N)

In section 2 of the method description of the national methods above, an overview has to be included on the boundary setting procedure for the national methods to check compliance with the WFD. In section 8.1 the results of a common boundary setting procedure at the scale of the common IC type should be explained where applicable.

8.1. Description of boundary setting procedure set for the common IC type

Summarize how boundaries were set following the framework of the BSP:

- Provide a description how you applied the full procedure (use of discontinuities, paired metrics, equidistant division of continuum)

Not relevant

- Provide pressure-response relationships (describe how the biological quality element changes as the impact of the pressure or pressures on supporting elements increases)

Not relevant

- Provide a comparison with WFD Annex V, normative definitions for each QE/ metrics and type

Not relevant

8.2. Description of IC type-specific biological communities representing the "borderline" conditions between good and moderate ecological status, considering possible biogeographical differences (as much as possible based on the common dataset and common metrics).

Not available

8.3. Boundary comparison and harmonization

Describe comparison of national boundaries, using comparability criteria (see Annex V of IC guidance).

Translating national boundaries
We generally applied a single linear regression analysis based on combined datasets to intercalibrate the invertebrate-based assessment methods. For the diatom-methods individual regression analyses (i.e. national EQR versus dICM based on national data) were used for countries providing at least 15 data samples. We took the model formula for the regression and determined the ICM value that equated to the upper class boundaries for each national method (i.e. high-good and good-moderate). Therefore, for method A, if y = mx +c where y = the ICM value, m = the regression slope, x = the EQR value of method A and c = the regression intercept, we derived the value on the ICM scale for values of x corresponding to the high-good and good-moderate class boundaries.

Evaluating sources of uncertainty
The intercalibration of very large rivers faced various sources of uncertainty. Some of these were attributed to the data collection itself and therefore relevant for all countries (see below and Annex III). Others depended on the relationship to the common metrics, the number of data samples and the distribution of data values, being very different for the participating countries. Two aspects mainly accounted for these country-/method-specific uncertainties: (1) the correlation of the national EQR and the ICM and (2) the data availability and range covered. Against this background we analysed the individual performance of the national methods in the exercise and evaluated the country-/method-specific level of confidence (Table 8.1).

In particular, the intercalibration exercise of invertebrate-based methods represented an "exceptional case" as outlined in the Intercalibration Guidance (p.73) due to the following:

- The exercise covered different types of large rivers and a broad geographical gradient.
- Multiple pressures were relevant for invertebrates in large rivers.
- Near-natural conditions were unavailable.
- National datasets often only covered a constraint gradient.
- Methodological differences existed among national classifications.

Above stated reasons caused general uncertainties in the intercalibration analysis that justified eased comparability criteria. We therefore opted for an acceptable boundary bias of ≤ 0.5 in invertebrate intercalibration[17].

Defining a harmonisation guideline
By fitting national class boundaries to each of the benchmark-standardised national EQR versus ICM relationships we established the predicted values on the ICM scale for each upper class boundary. This yielded a mid-point represented by the global average of the predicted values. **Only countries with a high**

[17] This definition implies that Member States joining the exercise later will also need to comply with this eased criterion.

level of confidence (see Table 8.1) were allowed to contribute to this mid-point that was then considered to represent the 'harmonisation guideline'. The more closely national class boundaries approached this guideline the greater the resulting level of harmonisation of their classifications and the lower the level of bias between methods. This principle applied irrespective of the error associated with the projection of each national class boundary onto a common scale.

Investigating the level of boundary bias

We defined the difference between the harmonisation guideline and the national boundary values on the ICM scale and converted these differences into class equivalents. First, the widths of national classes high, good and moderate were determined by subtracting the lower from the upper boundary value of each corresponding class translated into ICM units. The boundary bias was then calculated by subtracting the boundary position on the ICM scale from the harmonisation guideline. Finally, we related this difference to the width of the respective national class intersected by the harmonisation guideline (Figure 8.1).

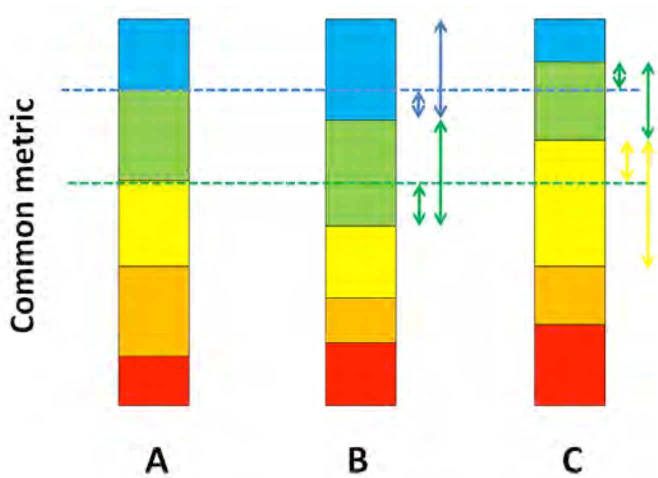


Figure 8.1: National classification schemes (A, B, C) intersected by the average boundary positions derived in intercalibration ('harmonisation guideline' – broken blue and green horizontal lines). Small arrows depict the distance of relevant national boundary to the harmonisation guideline. Large arrows define relevant national class width. The relation of small to large arrow-lengths specifies the boundary bias in class equivalents.

Figure 8.2 depicts the results of the linear regression analyses of the benchmark-standardised national EQRs against the ICMs. While the class boundaries of the invertebrate assessment methods were generally translated into iICM-units based on the combined national datasets, this data basis was used only for the national methods in the diatom exercise for which less than 15 samples were available. Table 8.2 informs about the regression features of countries with high confidence in the intercalibration analysis (see Table 8.1), including original and benchmark-standardised national boundary positions and these values translated into ICM-units based on the regression equation. Furthermore, the boundary bias that reflects the degree of deviation from the harmonisation guideline is specified.

The following national invertebrate methods exceeded a boundary bias of 0.50: *Estonia* (high-good boundary), *Finland* (good-moderate boundary) and *Hungary* (good-moderate boundary).

For the diatom methods a boundary bias of 0.25 was exceeded by *Sweden* (good-moderate boundary), *Austria-me1* (high-good boundary), *Belgium-Wallonia* (good-moderate boundary), *Germany* (good-moderate boundary), *Hungary* (high-good), *Slovakia* (high-good and good-moderate boundary), *Slovenia* (high-good boundary) and *the Netherlands* (high-good boundary). The required/possible boundary adjustments are listed in Tables 8.3a and 8.3b.

Presenting results of the boundary comparison for countries with low confidence

The national methods that showed low confidence in the intercalibration analysis (see Table 8.1) did not contribute to the harmonisation guideline. However, they were processed in the analysis to provide indication of a possible boundary bias (against the harmonisation guideline established by the classifications with high confidence). Due to the high level of uncertainty related to these outcomes we did not use them for boundary harmonisation. They can be the starting point of further national activities to complete the intercalibration exercise based on alternative analyses, if feasible[18] (Table 8.4).

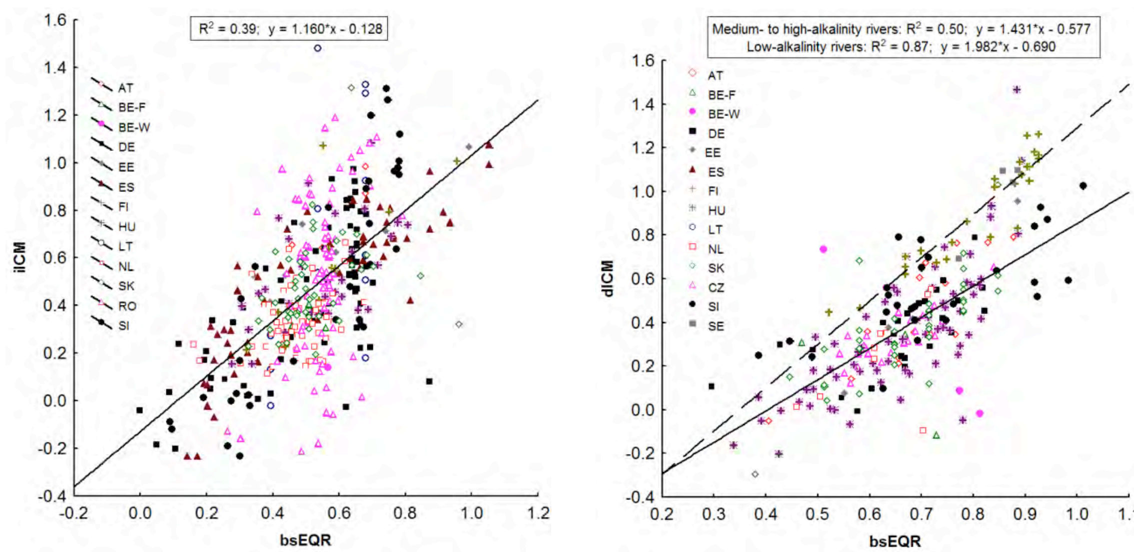


Figure 8.2: Linear regression plots of the benchmark-standardised (bs) national EQR values against the invertebrate-ICM (left) and the diatom-ICM (right) based on combined national datasets. The right figure includes the regressions for low-alkalinity rivers (broken line) and medium- to high-alkalinity rivers (solid line).

[18] Such alternative analyses may include bilateral comparisons at individual rivers using close-border monitoring stations (e.g. see Annex III, p. 43). If harmonised monitoring data are available covering various riparian countries (e.g. Joint Danube Survey) comparisons of national classifications against river-specific baselines can be carried out (see *Birk, van Kouwen & Willby. Harmonising the bioassessment of large rivers in the absence of near-natural reference conditions - a case study of the Danube River. Submitted to Freshwater Biology*). In any case the harmonisation guideline defined in this XGIG intercalibration exercise shall be regarded.
If such alternative analyses are not feasible (e.g. only very constrained quality gradient available, bilateral comparisons impossible, low number of water bodies existing etc.) the outcomes of this exercise may represent the best available intercalibration solution.

Table 8.1: Performance of the national methods regarding the two criteria of uncertainty and resulting level of confidence in the intercalibration analysis

1. Criterion: *Correlation*

Criterion: The correlation of the national method and the ICM has to be significant with a Pearson correlation coefficient of ≥ 0.5 ($p<0.05$).

Rationale: The ICM shall be well correlated to reliably reflect the national assessment results.

2. Criterion: *Data availability and range*

Criterion: Each country contributes with at least ten samples to the exercise.

OR

Samples cover good ecological status according to the national classification.

Rationale: A representative view on the national definitions of good ecological status in relation to pressures and common metrics requires a sufficient number of samples per country.

If the national datasets do not cover water bodies in (or close to) good ecological status, the good status boundaries can only be extrapolated or modelled for the boundary comparisons.

a) Benthic invertebrates

	1. Correlation	2. Data availability and range	Level of confidence
Austria	Passed	Failed	low
Belgium (Flanders)	Failed	Failed	low
Belgium (Wallonia)	Failed	Failed	low
Estonia	Passed	Passed	high
Finland	Passed	Passed	high
Germany	Passed	Passed	high
Hungary	Passed	Passed	high
Lithuania	Failed	Passed	low
Netherlands	Failed	Passed	low
Romania	Failed	Passed	low
Slovakia	Failed	Passed	low
Slovenia	Passed	Passed	high
Spain	Passed	Passed	high

b) Benthic diatoms

	1. Correlation	2. Data availability and range	Level of confidence
Austria – me1	Passed	Passed	high
Austria – me2	Failed	Passed	low
Belgium (Flanders)	Failed	Failed	low
Belgium (Wallonia)	Passed	Passed	high
Czech Republic	Passed	Passed	high
Estonia	Passed	Passed	high
Finland	Passed	Passed	high
Germany	Passed	Passed	high
Hungary	Passed	Passed	high
Netherlands	Passed	Passed	high
Slovakia	Passed	Passed	high
Slovenia	Passed	Passed	high
Sweden	Passed	Passed	high

Table 8.2: Original national high-good (HG) and good-moderate (GM) status class boundaries, benchmark-standardised (bs) boundaries, linear regression equations (including coefficient of determination), boundary values translated into ICM-scale and boundary bias.

Note: these tables only include methods with high confidence in the intercalibration analysis.

a) Benthic invertebrates

Country	Original		Original (bs)		Linear regression			ICM scale		Boundary bias[a]	
	HG	GM	HG	GM	Slope	Intercept	R^2	HG	GM	HG	GM
Estonia	0.900	0.700	0.940	0.740	1.160	0.128	0.39	0.963	0.731	0.62[#]	0.47
Finland	0.800	0.600	0.688	0.488				0.671	0.439	-0.35	-0.56[$]
Germany	0.800	0.600	0.787	0.587				0.786	0.554	-0.11	-0.06
Hungary	0.800	0.600	0.687	0.487				0.670	0.438	-0.35	-0.56[$]
Slovenia	0.800	0.600	0.888	0.688				0.903	0.671	0.36	0.45
Spain	0.780	0.480	0.904	0.604				0.921	0.573	0.29	0.02
Harmonisation guideline								***0.819***	***0.568***		

[a] acceptable boundary bias ≤ 0.50

[$] national boundaries requiring harmonisation; [#] national boundaries can be lowered

b) Benthic diatoms – low alkalinity rivers (R-L1)

Country	Original		Original (bs)		Linear regression			ICM scale		Boundary bias[b]	
	HG	GM	HG	GM	Slope	Intercept	R^2	HG	GM	HG	GM
Finland	0.800	0.600	0.828	0.628	1.947	-0.660	0.88	0.952	0.563	-0.08	-0.20
Sweden	0.890	0.740	0.862	0.712	1.982	-0.690	0.87	1.018	0.721	0.12	0.27[#]
Harmonisation guideline								***0.985***	***0.642***		

c) Benthic diatoms – medium- to high-alkalinity rivers (R-L2)

Country	Original		Original (bs)		Linear regression			ICM scale		Boundary bias[b]	
	HG	GM	HG	GM	Slope	Intercept	R^2	HG	GM	HG	GM
Austria – me1*	0.850	0.570	0.927	0.647	1.431	-0.577	0.50	0.750	0.349	0.99[#]	0.09
Belgium (Wallonia)*	0.970	0.730	0.812	0.572	1.431	-0.577	0.50	0.586	0.242	-0.13	-0.30[$]
Czech Republic	0.800	0.600	0.900	0.700	1.228	-0.468	0.67	0.638	0.392	0.07	0.19
Estonia*	0.830	0.640	0.874	0.684	1.431	-0.577	0.50	0.674	0.402	0.23	0.21
Germany	0.725	0.545	0.888	0.708	1.019	-0.313	0.45	0.591	0.408	-0.20	0.34[#]
Hungary	0.800	0.600	0.892	0.692	1.731	-0.786	0.59	0.758	0.412	0.69[#]	0.19
Slovakia	0.900	0.700	0.747	0.547	1.574	-0.683	0.63	0.493	0.178	-0.43[$]	-0.53[$]
Slovenia	0.800	0.600	0.716	0.516	0.971	-0.156	0.49	0.540	0.346	-0.46[$]	0.00
Harmonisation guideline								***0.629***	***0.345***		
Netherlands[c]	0.800	0.600	-	-	2.240	-0.993	0.90	0.799	0.351	0.38[#]	0.01
Slovakia (adjusted[d])	0.900	0.700	-	-	1.609	-0.855	0.80	0.593	0.271	-0.11	-0.23
Slovenia (adjusted[d])	0.800	0.600	-	-	1.492	-0.448	0.58	0.745	0.447	0.39	0.34

* Diatom methods analysed in global regression using a combined dataset of all available, benchmark-standardised national EQRs against the dICM.

[b] acceptable boundary bias ≤ 0.25

[c] The Netherlands provided additional data after the completion of IC milestone version 1.0. Their boundaries thus did not contribute to the harmonisation guideline but were compared afterwards against the average boundary position established by the eight countries listed in the table above.

[d] Results of the comparability analysis AFTER adjustment of the national classification (September 2012).

[$] national boundaries requiring harmonisation; [#] national boundaries can be lowered

Table 8.3a: Necessary boundary adjustments (current national classification too relaxed)

BQE	Country	Boundary	Original value	Recommended value[19]	Final value[20]
Benthic invertebrates	Finland	Good-moderate	0.600	0.623	*not intercalibrated*
	Hungary	Good-moderate	0.600	0.625	
Benthic diatoms	Belgium (Wallonia)	Good-moderate	0.730	0.740	0.740
	Slovakia[21]	High-good	0.900	0.939	0.900
		Good-moderate	0.700	0.745	0.700
	Slovenia[21]	High-good	0.800	0.834	0.800

Table 8.3b: Possible boundary adjustments (current national classification too stringent)

BQE	Country	Boundary	Original value	Recommended value[22]	Final value[20]
Benthic invertebrates	Estonia	High-good	0.900	0.851	*not intercalibrated*
Benthic diatoms	Sweden	Good-moderate	0.740	0.737	0.740
	Austria – me1	High-good	0.850	0.800	0.850
	Germany	Good-moderate	0.545	0.531	0.545
	Hungary	High-good	0.800	0.762	0.762
	Netherlands	High-good	0.800	0.780	0.800

Table 8.4: Indicators of possible boundary bias for the national methods with low confidence in the intercalibration analysis **– Note: these outcomes were not used for boundary harmonisation.**

These tables list the original national high-good (HG) and good-moderate (GM) status class boundaries, benchmark-standardised (bs) boundaries, linear regression equations (including coefficient of determination), boundary values translated into ICM-scale and boundary bias.

a) Benthic invertebrates

Country	Original		Original (bs)		Linear regression			ICM scale		Boundary bias	
	HG	GM	HG	GM	Slope	Intercept	R^2	HG	GM	HG	GM
Austria	0.800	0.600	0.928	0.728	1.160	0.128	0.39	0.950	0.718	0.56	0.65
Belgium (Flanders)	0.900	0.700	0.734	0.534				0.724	0.492	-0.26	-0.33
Belgium (Wallonia)	0.800	0.670	0.710	0.580				0.696	0.545	-0.31	-0.15
Lithuania	0.770	0.635	0.877	0.742				0.890	0.733	0.45	1.02
Netherlands	0.800	0.600	0.948	0.748				0.973	0.741	0.66	0.75
Romania	0.740	0.580	0.586	0.426				0.552	0.367	-0.49	-1.08
Slovakia	0.800	0.600	0.811	0.611				0.814	0.582	-0.02	0.06

b) Benthic diatoms – medium- to high-alkalinity rivers

Country	Original		Original (bs)		Linear regression			ICM scale		Boundary bias	
	HG	GM	HG	GM	Slope	Intercept	R^2	HG	GM	HG	GM
Austria – me2*	0.810	0.560	0.835	0.585	1.431	-0.577	0.50	0.618	0.261	-0.03	-0.33
Belgium (Flanders)*	0.800	0.600	0.699	0.499	1.431	-0.577	0.50	0.423	0.137	-0.72	-0.73

* Diatom methods analysed in global regression using a combined dataset of all available, benchmark-standardised national EQRs against the dICM

19 Recommended boundary value meeting the criterion: boundary bias ≥ -0.25
20 Boundary value finally adopted by the Member State
21 Slovakia and Slovenia adjusted their component metrics to meet the comparability criteria while keeping their original classification (see Table 8.2).
22 Boundary value meeting the criterion: boundary bias = 0.25

Annex I – Overview of national assessment methods

Table A1: National assessment methods for very large rivers using benthic invertebrates

Country	Method name	db_ID[23]	Sampling	Determination level	Metrics	Pressure	Literature reference
Austria	Assessment of the biological quality elements - part benthic invertebrates	49	Multi-Habitat-Sampling at banks (occasionally Airlift)	Species	Saprobic Index (*Danube, March, Thaya*); Degradation Index, Number of EPT taxa, Proportion of EPT taxa, Lithal-profundal preferring taxa (*Large Alpine rivers*)	Organic pollution, General degradation	Stubauer 2002, Ofenböck et al. 2005
Belgium (Flanders)	Multimetric Macroinvertebrate Index Flanders	123	Kick-Sampling at banks combined with hand-picking of animals from stones	Genus and Family	Number of taxa, Number of EPT taxa, Number of sensitive taxa, Shannon-Diversity, Mean Tolerance Score	Organic pollution, General degradation	Gabriels et al. 2010
Belgium (Wallonia)	Global biological index adapted to large watercourses and deep rivers	290	Artificial substrates	Genus and Family	IBGN (incl. richness and sensitivity metrics)	General degradation	Vanden Bossche & Usseglio-Polatera 2005
Estonia	Estimation of freshwater quality using macroinvertebrates	46	Multi-Habitat-Sampling at banks	Species	Number of Taxa, Number of EPT taxa, Shannon-Diversity, Average Score Per Taxon (ASPT), Danish Stream Fauna Index	Organic pollution, General degradation	Timm & Vilbaste 2010.
Finland	Finnish multimetric index	146	Multi-Habitat-Sampling at banks within riffle zones	Species	Occurrence of type-specific taxa, Occurrence of type-specific EPT families, PMA (Percent Model Affinity)	General degradation	Aroviita et al. 2008, Novak & Bode 1992
Germany	Assessment method for rivers using benthic invertebrates	275	Multi-Habitat-Sampling at banks, Grab	Species	Saprobic Index, Potamon-Typie-Index	Organic pollution, General degradation	Rolauffs et al. 2003, Schöll et al. 2005
Hungary	Hungarian Multimetric Macroinvertebrate Index for large and very large rivers	360	Multi-Habitat-Sampling at banks, Dredging	Species	Shannon diversity, Total number of taxa, ASPT, Number of EPTCBO taxa	Organic pollution, General degradation	Várbíró et al. 2011a
Lithuania	Assessment system for rivers using macrozoobenthos indicators (Danish Stream Fauna Index)	58	Multi-Habitat-Sampling at banks	Genus and Family	Danish Stream Fauna Index	General degradation	-
Netherlands	WFD-metrics for natural water types	288	Multi-Habitat-Sampling at banks, Sediment corer, Brushed stone samples	Species	Number of typical species, Proportion of negative species, Proportion of positive species	General degradation	-
Romania	Romanian assessment system for rivers using benthic invertebrates	n.a.	Multi-Habitat-Sampling at banks	Species	Saprobic Index, Number of EPT taxa, Shannon-Wiener Diversity Index, Number of families, Proportion of Chironomides and Oligocheates, Proportion of Functional Groups, Proportion of rhithral/potamal preferring taxa	Organic pollution, General degradation	-
Slovakia	Slovak assessment of benthic invertebrates in large rivers	165	Multi-Habitat-Sampling at banks	Species	Saprobic Index, Proportion of oligosaprobic taxa, Rhithron-Type-Index, Index of biocoenotic region, Proportion of akal-lithal-psammal preferring taxa, BMWP	Organic pollution, General degradation, Hydromorphological impairment	Šporka et al. 2009
Slovenia	Slovenian assessment system for rivers using benthic invertebrates	289	Multi-Habitat-Sampling at banks	Species	Saprobic Index, River Fauna Index, Proportion of akal-lithal-psammal preferring taxa	Organic pollution, Hydromorphological impairment	-
Spain	Iberian Biological Monitoring Working Party	9	Multi-Habitat-Sampling (Surber + Drowning Box)	Family	Iberian BMWP	Organic pollution	Munne & Prat 2009

[23] ID of the WISER methods' database; available online at *"http://www.wiser.eu/programme-and-results/data-and-guidelines/method-database/detail.php?id="* plus *"db_ID"*

Table A2: National assessment methods for very large rivers using phytobenthos (benthic diatoms)

Country	Method name	db_ID	Diatoms	Other PB[24]	Sampling	#Valves	Metrics	Pressure	Macrophyte combination	Literature reference
Austria	Assessment of the biological quality elements - part phytobenthos	168	yes	yes	Scraped from stones at river banks (5-10 replicates); full survey for other phytobenthos	500	Trophic Index, Saprobic Index, Reference-Taxa-Indices	Eutrophication, Organic pollution, General degradation	Worst-case	Pfister & Pipp 2010
Belgium (Flanders)	Proportions of Impact-Sensitive and Impact-Associated Diatoms	127	yes	no	Scraped from stones and plants at river banks (5 replicates)	500	IAD, ISD	Eutrophication, Organic pollution, General degradation	Worst-case	Hendrickx & Denys 2005, Leyssen et al. 2006
Belgium (Wallonia)	Pollution Sensitivity Index	14	yes	no	Scraped from stones at river banks	400	IPS	Eutrophication, Organic pollution, General degradation	-	-
Czech Republic	Assessment system for rivers using phytobenthos	76	yes	yes	Scraped from stones at river banks (5 replicates)	Abun-dance classes 1-7	Czech Saprobic-Trophic Index	Eutrophication, Organic pollution	-	Marvan & Heteša 2006, Marvan, Opatřilová & Heteša 2011
Estonia	Assessment system for rivers using phytobenthos in Estonia	111	yes	no	Scraped from stones at river banks (5 replicates)	400	IPS, Trophic Diatom Index, Watanabe Index	Eutrophication, Organic pollution	-	Vilbaste 2004
Finland	Pollution Sensitivity Index	247	yes	no	Scraped from stones at river banks (5-10 replicates)	400	IPS	Eutrophication, Organic pollution, General degradation	-	Eloranta & Soininen 2002
France	Indice Biologique Diatomées 2007 (IBD2007)	157	yes	no	Scraped from stones at river banks (5 replicates)	400	IBD2007	Eutrophication, Organic pollution, General degradation	Average	Coste et al. 2009
Germany	German Assessment system for Macrophytes and Phytobenthos according to the EU WFD	218	yes	yes	Scraped from stones at river banks (5-10 replicates); full survey for other phytobenthos	400	Trophic Index, Saprobic Index, Reference-Index	Eutrophication, Organic pollution, General degradation	Average	Schaumburg et al. 2004
Hungary	Improvement of the Hungarian ecological water qualification system - Phytobenthos in Rivers	221	yes	no	Scraped from stones at river banks (5-10 replicates)	400	IPS	Eutrophication, Organic pollution	-	Várbíró et al. 2011b, van Dam et al. 2007
Netherlands	WFD-metrics for natural water types	298	yes	no	Removal of 4-8 reed stems (several replicates)	200	IPS	Eutrophication, Organic pollution, General degradation	Average	van den Berg et al. 2007
Slovakia	Ecological status assessment system for rivers using phytobenthos	169	yes	no	Scraped from stones at river banks (5-10 replicates)	300-500	IPS, CEE, EPI-D, occurrence of filamentous bacteria	Eutrophication, Organic pollution, General degradation	Worst-case	Hlubikova et al. 2006
Slovenia	Ecological status assessment system for rivers using phytobenthos and macrophytes; Phytobenthos	37	yes	no	All available habitats	500	Saprobic Index, Trophic Index	Eutrophication, Organic pollution	Average (eutrophication module)	-
Sweden	Benthic algae in running water - diatom analysis	78	yes	no	Scraped from stones at river banks (5-10 replicates)	400	IPS, TDI, %PT, ACID	Acidification, Eutrophication, Organic pollution	-	-

[24] Phytobenthos

Table A3: Definition of reference conditions

BQE	Country	RefCond	Comments
Benthic Invertebrates	Austria	Expert knowledge, Least Disturbed Conditions	-
	Belgium (Flanders)	Expert knowledge	-
	Belgium (Wallonia)	Expert knowledge	For the river types where (i) no sites of high status were available but (ii) some good status sites were available, the high status was defined as the good status multiplied by 1.25 (see Vanden Bossche & Usseglio-Polatera 2005).
	Estonia	Existing near-natural reference sites	Reference values adopted from smaller rivers
	Finland	Existing near-natural reference sites	Reference values adopted from smaller rivers
	Germany	Expert knowledge, Least Disturbed Conditions	Module PTI: theoretical reference value (dominance of potamal species). Module "Saprobic Index": Best available sites (10th percentile of available large river data).
	Hungary	Expert knowledge, Least Disturbed Conditions	-
	Lithuania	Expert knowledge, Least Disturbed Conditions	Reference values adopted from smaller rivers
	Netherlands	Expert knowledge, Historical data	-
	Romania	Expert knowledge, Least Disturbed Conditions	90th percentile value of least-disturbed sites
	Slovakia	Expert knowledge, Least Disturbed Conditions	5th or 95th percentile (depending on decrease or increase of value by deterioration, pollution, degradation) of all selected metric values in monitoring localities due to absence of reference sites
	Slovenia	Expert knowledge, Least Disturbed Conditions, Modelling (extrapolating model results)	Module "Saprobic Index": Best available sites; Module "Hydromorphology Index": Maximum score of biological index plus 5%
	Spain	Least Disturbed Conditions	Chemistry - ammonium: < 0.2 mg/l (mean), < 1 mg/l (max); nitrate-N: < 10 mg/l (mean), < 20 mg/l (max); phosphate-P: < 0.1 mg/l (mean), < 1 mg/l (max); Hydrology - minimum flow: > 20 % of natural flow; near-natural flow regime variation; Morphology - good riparian conditions (QBR index > 75)
Benthic diatoms	Austria	Expert knowledge, Least Disturbed Conditions	Bioregion-specific reference definition (allocation of bioregional reference to large river stretch located in bioregion)
	Belgium (Flanders)	Expert knowledge	-
	Belgium (Wallonia)	Least Disturbed Conditions	No river size-specific reference values
	Czech Republic	Expert knowledge, Least Disturbed Conditions	The reference values were derived from the national dataset of best available sites (number of samples = 85) and by comparison with AT and DE methods. Subsequently, the equation for calculation of the reference value based on altitude, Strahler stream order and acid neutralisation capacity of the individual assessed site was derived.
	Estonia	Least Disturbed Conditions	No river size-specific reference values. References adopted from smaller rivers.
	Finland	Least Disturbed Conditions	No river size-specific reference values. Values adopted from smaller rivers (expert judgement)
	France	Least Disturbed Conditions	No river size-specific reference values. Values adopted from smaller rivers (expert judgement)
	Germany	Expert knowledge	Theoretical reference value (ideal type-specific species composition)
	Hungary	Expert knowledge, Least Disturbed Conditions	-
	Netherlands	Expert knowledge	-
	Slovakia	Modelling (extrapolating model results)	The regression model used for the index values: IPS = 15.821 + 0.005(altitude) - 0.491(HMQS) - 0.025(agricult.+urban) (R^2=0.366); CEE = 13.187 + 0.005(altitude) - 0.043(agricult.+urban) (R^2=0.412); EPI-D = 12.924 + 0.006(altitude) - 0.031(agricult.+urban) (R^2=0.444). Significant levels: p (altitude) = 0.000; p (HMQS) = 0.038; p (agricult.+urban) = 0.036. Catchment size is not statistically significant factor for either of the indices so it was not included in the models. HMQS was statistically significant only for IPS. Mid-point value for altitude category up to 200 m a. s. l. is 150. Hydromorphological impacts were set to 1 and land use parameters were set to zero impact levels, to derive the type specific reference values (HMQS = 1, agricultural and urban land combined = 'agricult.+urban'=0).
	Slovenia	Expert knowledge, Modelling (extrapolating model results)	Reference values of both indices (Trophic and Saprobic index) were defined using regression equations between log total phosphorous values and indices. Reference values were calculated as index values at log TP = -2 (TP = 10 μg/L).
	Sweden	Existing near-natural reference sites, Expert knowledge	No river size-specific reference values. Reference sites are mostly situated in northern part of country.

Table A4: Ecological class boundary setting (general schemes)

BQE	Country	Boundary setting	Boundary description	Boundary communities
Benthic Invertebrates	Austria	Calibrated against pre-classified sampling sites Equidistant division of the EQR gradient	Saprobic index: H/G boundary is 75th percentile of least-disturbed sites; MMI: Alpine rivers: reference value is 95th percentile of reference sites, Danube, March and Thaya: assessment currently based on saprobic index only (no MMI reference values); Other boundaries calculated by equidistant division	Expressed by specific metric values, no verbal description.
	Belgium (Flanders)	Equidistant division of the EQR gradient	EQR gradient is assumed to represent a continuous correlation with general degradation.	The EQR values at good status reflect metric values that are only slightly lower than at (expert-based) reference state, hence the community can be characterised as only slightly different from reference in terms of taxa richness, sensitivity and diversity.
	Belgium (Wallonia)	Equidistant division of the EQR gradient	Class boundaries are defined as deviations from maximum ecological potential using the values 0.75, 0.50 and 0.25.	-
	Estonia	Expert judgement	The scale of EQR values was established by expert judgement proposing appropriate intervals from high to bad ecological status.	Expressed by specific metric values, no verbal description.
	Finland	Equidistant division of the EQR gradient High-good boundary derived from metric variability at near-natural reference sites	Only EQRs are used to define ecological quality classes. High/Good status defined by the 25 percentage point from type-specific reference EQR distributions. Lower end of bad class is set to EQR=0 and class widths are equidistant.	Only EQRs are used to define ecological quality classes. Good status is defined by the 25 percentage point of reference site EQR in each type.
	Germany	Equidistant division of the EQR gradient	The national class boundary setting follows Option B of the REFCOND Guidance: The scale of EQR values was established by expert judgement proposing appropriate intervals from high to bad ecological status. The application of the scale to real datasets confirmed the proposed boundary setting. High status of the Module "General Degradation" corresponds to very low level of anthropogenic impact concerning land use und hydromorphological pressure.	-
	Hungary	Calibrated against pre-classified sampling sites High-good boundary derived from metric variability at sites in Least Disturbed Conditions	The high-good boundary was derived from metric variability at alternative benchmark sites (median). The good-moderate boundary was derived from metric variability at alternative benchmark sites (lower quartiles).	Good status communities feature high taxonomic diversity with at least 21 taxa in total, from which ten belong to the groups of Ephemeroptera, Plecoptera, Trichoptera, Coleoptera, Bivalvia or Odonata. Relevant Bivalvia taxa comprise species such as *Unio pictorum*, *Unio tumidus*, *Pisidium amnicum*, *Pisidium henslowanum* or *Pisidium supinum*.
	Lithuania	Calibrated against pre-classified sampling sites	The boundaries were derived from metric scores at adjacent status class sites	Good status is characterized by a high diversity of positive (disturbance sensitive) macroinvertebrate taxa. Diversity of negative (disturbance resistant) taxa is low.
	Netherlands	Calibrated against pre-classified sampling sites Equidistant division of the EQR gradient	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. 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 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.
	Romania	Calibrated against pre-classified sampling sites	The class boundary are set based on expert judgement. The application of the scale to real datasets confirmed the proposed boundary setting. Romanian large rivers are affected by hydromorphological and pollution pressures.	High status: presence of sensitive taxa such as: stoneflies, mayflies, caddis flies etc.; high diversity; absence or very low presence of oligochaets. Good status: slight deviation from high status regarding diversity, presence of sensitive taxa, low presence of tolerant taxa.
	Slovakia	Equidistant division of the EQR gradient	The reference value derived from the 95th (5th) percentile of the metric value range at monitoring sites corresponds to EQR = 0.8 and the rest of the range was equally distributed among the classes.	Expressed by specific metric values, no verbal description.

Table A4 (cont.): Ecological class boundary setting (general schemes)

BQE	Country	Boundary setting	Boundary description	Boundary communities
Benthic Invertebrates	Slovenia	*Module Organic pollution* - High/Good boundary was derived from metric variability at low disturbed sites. - Other boundaries (Good/Moderate, Moderate/Poor and Poor/Bad) by equidistant division. *Module Hydromorphological alteration/general degradation* - Using paired metrics that respond in different ways to the influence of pressure.	*Module Organic pollution* a) High/Good boundary was defined as 75th percentile value of best available sites. b) Other boundary values (Good/Moderate, Moderate/Poor and Poor/Bad) were defined using equidistant division of the remaining EQR gradient. c) Lower anchor was defined as highest possible index value based on the Slovenian operational taxa list; it is not maximum index value. *Module Hydromorphological alteration/general degradation* Boundary values between five ecological status classes were defined based on the changes in ratio between the sensitive and tolerant taxa using following criteria: a) High/Good boundary was defined where portion of tolerant taxa start to increase (tolerant < sensitive). b) Good/Moderate boundary was defined where portion of tolerant taxa reach the portion of sensitive taxa (tolerant ≈ sensitive). c) Moderate/Poor boundary was defined where portion of tolerant taxa exceed the protion of sensitive taxa (tolerant > sensitive). d) Poor/Bad boundary was defined where portion of tolerant taxa start to dominate (tolerant >> sensitive).	a) High/Good boundary community is dominated (>80%) by sensitive taxa and tolerant taxa are present in a small portion (<20%). In good status class portion of sensitive taxa is bigger than portion of tolerant taxa. b) Good/Moderate boundary community is still slightly dominated by the sensitive taxa (≈ 60%) but portion of tolerant taxa is approaching portion of sensitive taxa (≈ 40%). In moderate status class portion of sensitive and tolerant taxa is approximately equal.
	Spain	Division of the EQR gradient High-good boundary derived from metric variability at sites in Least Disturbed Conditions	The distribution of the IBMWP values expressed as EQR was assessed. The high/good boundary corresponds to the 25th percentile of the EQR in reference/non-disturbed sites. The rest of the boundary classes were set as percentages of the 25th percentile of reference/non-disturbed sites: 61% corresponds to the good/moderate boundary, 36% corresponds to the moderate/poor boundary and 15% corresponds to the poor/bad boundary.	The high status is characterized by a relatively high total number of families (approx. ≥ 23); the good status features 16 to 22 families. Although with a higher variability, the number of Plecoptera at high status is usually >2; the good class features a minimum of one Plecoptera family. The number of Ephemeroptera or Trichoptera families are >4 at high status sites and 2 to 4 at good status sites. The good status is characterized by a community composed of 5 to 9 EPT families, whereas at high status sites usually more than 8 EPT families occur.
Benthic diatoms	Austria	High-good boundary derived from metric variability at near-natural reference sites	H/G: 10th percentile of high class Trophic Index / Saprobic Index values (all values lying within the defined type specific trophic / saprobic reference class based on Trophic Index / Saprobic Index classes according to ROTT's trophic / saprobic indication system) G/M: Upper Trophic Index boundary of next worse trophic class (following the type specific trophic reference class)	For good status defined common reference species and/or river type-specific species must obtain a certain percentage of all occurring algae (percentage varying in different bioregions).
	Belgium (Flanders)	Equidistant division of the EQR gradient	EQR gradient is assumed to represent a continuous trend with general degradation.	The EQR values at good status are characterised by a relatively low IAD and a ISD that is slightly reduced in comparison to reference.
	Belgium (Wallonia)	Equidistant division of the EQR gradient	The presence of very sensitive organisms has been taken in account to define high status.	A good community is relevant of the local typology and the presence of sensitive organisms or families is taken in account.
	Czech Republic	Equidistant division of the EQR gradient	EQR gradient is assumed to represent a continuous correlation with general degradation.	-
	Estonia	Adopted from intercalibration of other river types	Boundary values adopted from the results of the intercalibration exercise for small- to medium-sized rivers.	At good ecological status sensitive reference species dominate.
	Finland	Using discontinuities in the relationship of anthropogenic pressure and the biological response.	The rivers studied were classified to five classes according to the degree of human impacts in the drainage basin in general or near the sampling station. Rivers with more or less natural state of very low degree of human impacts showed IPS values > 16, whereas those with slight human impact had the IPS from 14 to 16. The index values decreased markedly with increasing strength of human impact. Based on the results, the following limit values for IPS for evaluation of ecological water quality classes were proposed. High quality IPS>17; Good quality IPS 15-17; Moderate quality 12-15; Poor quality 9-12; Bad quality <9	No description of boundary communities.
	France	High-good boundary derived from metric variability at sites in Least Disturbed Conditions Equidistant division of the EQR gradient	a) High/Good boundary was defined as 75th percentile value of best available sites. b) For each type, the remaining range below the High/Good boundary and the IBD minimum value was split into four equal classes to derive a preliminary Good/Moderate boundary. This preliminary boundary was then increased by one point on the IBD scale for all national types.	The IBD values obtained were then checked to verify their compliance with normative definitions: Analysing the percentage of sensitive species ('oligotraphent' + 'mesotraphent' species: van Dam et al., 1994) in reference conditions and along the ecological status gradient revealed (i) no significant difference in sensitive species % between reference conditions and high status; (ii) a very slight but significant decrease of sensitive species between high and good status; (iii) a drop in the percentage of sensitive species between good and moderate status.
	Germany	Calibrated against pre-classified sampling sites	The boundaries were set at the zones of distinct changes of the biocoenosis (macrophytes and phytobenthos), and depending on indicator species lists derived from nutrient dependent Trophic Index (diatoms).	Type-specific reference species and tolerant species are still dominant, pressure indicators are rare, i.e. slight deviation from high status (WFD normative definitions).

Table A4 (cont.): Ecological class boundary setting (general schemes)

BQE	Country	Boundary setting	Boundary description	Boundary communities
Benthic diatoms	Hungary	High-good boundary derived from metric variability at sites in Least Disturbed Conditions Equidistant division of the EQR gradient	Reference conditions which could be applied across rivers in Hungary have not been established yet. Nevertheless, unimpacted stretches or sites with low pollution and with smaller hydromorphological alterations can be found in almost every river type (i.e. sites in Least Disturbed Conditions - LDC). On basis of the pressure data (TP, BOD, CODCr, Electrical Conductivity) the LDS were selected.10th percentiles of the index values of the selected LDS sites were considered as high/good (H/G) class boundaries and 75th percentiles as good/moderate (G/M) boundaries in every type. The rest of data was divided into 3 equal parts between the minimum value of the index in a given river groups and the G/M value in order to set the further boundaries.	At good status stands of sensitive taxa are well developed. They are dominant, but significantly decreasing at good-moderate boundary and replaced by tolerant taxa. The 10th percentiles of the index values of the selected LDS sites were considered as high/good (H/G) class boundaries.
	Netherlands	Boundaries taken over from the intercalibration exercise	By using similarities in the geographic conditions the score on the IPS-scale in accordance with the reference condition is deduced for the Dutch situation. Next the scores of ten variants of an IPS-based metric were calculated for samples of CB-GIG type R-C1 and R-C4 For each of the ten variants the boundary values H/G and G/M at the intercalibration metric and several other performance characteristics were calculated, including the 95% confidence intervals of the boundary values. Finally a metric with a reference value has been chosen with boundary values which deviates less than the required 0.05 units from the mean values of all Member States.	The Good-Moderate boundary is based on the Intercalibration Metric.
	Slovakia	Equidistant division of the EQR gradient	Benthic diatoms: two modules - benthic diatoms and filamentous bacteria. a) benthic diatoms modul-4 altitude categories, based on reference sites within 2004 . For 200-500, 500-800 and above 800 - boundary between H/G = 25. Percentile of average based on reference sites in 2004. For altitude below 200 linear model used - derived from type of altitude 200-500 by means of modelling (this procedure - applied for all 3 metrics). The other boundaries calculated using the range of metrics values within high status (best value) and minimal calculated value of metric from the data set. The whole range was equally subdivided and boundaries were stated accordingly. b) filamentous bacteria module - proportion of bacteria (%) in phytobenthos found during field survey, expressed in five level scale Result of both modules= the worse value classifies.	In Slovakia background taxa lists are not prescribed and not especially created for good status conditions as well as any other ecological status classes.
	Slovenia	High/Good and Good/Moderate boundaries were derived from metric variability at stressor class 2. Other boundaries (Moderate/Poor and Poor/Bad) by equidistant division.	a) High/Good boundary was defined as 10th percentile value of the pressure class 2 (out of three). b) Good/Moderate boundary was defined as 50th percentile value of the pressure class 2 (out of three). c) Other boundary values (Moderate/Poor and Poor/Bad) were defined using equidistant division of the remaining EQR gradient. d) Lower anchor was defined as highest possible index value based on the Slovenian operational taxa list; it is not maximum index value.	Expressed by specific metric values (Trophic index and Saprobic index), no verbal description.
	Sweden	High-good boundary derived from metric variability at near-natural reference sites Using paired metrics that respond in different ways to the influence of the pressure	High-good boundary derived from pre-classified reference state. Good/moderate boundary: IPS=14.5. The G/M boundary was set to the IPS value where the nutrient tolerant and pollution tolerant species exceed a relative abundance of approx. 30 % (and the amount of sensitive species falls below approx. 30 %).	A good status community is defined by the relative abundance of sensitive species of at least ± 30 %, and nutrient tolerant and pollution tolerant species not more than approx. 30 %.

Annex II – Conceptual paper on large river bioassessment

In this annex to the IC milestone report we highlight the main difficulties of large river bioassessment according to the European Water Framework Directive.

Typology

Large rivers often do not fulfil the doctrine of longitudinal zonation given in textbooks and have long been considered as **individuals** that do not allow for typification. The rivers Rhine or Danube, for instance, can change their character rather quickly from slow flowing, alluvial reaches dominated by fine sediments and various islands into fast flowing mountain stretches with coarse bed material. Moreover, large rivers feature a variety of different habitats at small spatial scales. While, for instance, the active channels of mountainous floodplain rivers are inhabited by rheophilous organisms typical of the potamal zone, characteristic stillwater communities can be found in the channel margin habitats and the often vast floodplain areas close by.

Although **floodplains** are a common feature of many large rivers and an important part of the functionality of these hydrosystems, they are rarely considered in the assessment of ecological status. This especially applies to the quality of lateral connectivity. The dynamic hydrology especially of large floodplain rivers – with specific patterns of flooding and low water, ground water influence, fluvial erosion and sedimentation etc. – creates a mosaic of aquatic habitats such as secondary river arms, oxbow lakes, permanent or temporal ponds, all of which feature different reference communities at natural conditions. Floodplains also feature a high diversity of terrestrial and semi-terrestrial habitats (e.g. softwood and hardwood floodplain forest, bare or vegetated gravel bars) which are closely related to the aquatic zone.

Implications for large river bioassessment

The complexity of habitats complicates ecological assessment of large river systems because, for example, the occurrence of stillwater species may indicate good conditions at floodplain lakes, but bad conditions at an anthropogenically impounded river section nearby. To account for this complexity integrated assessment of large rivers requires **modular approaches**, combining the status classification of the main river channel (intercalibrated in the current large river exercise) with quality evaluations of the diverse floodplain habitats. Here, organism groups such as dragonflies, caddis flies and amphibians represent suitable bioindicators as their life-cycles cover both terrestrial and aquatic stages. Terrestrial and semi-terrestrial habitats can be assessed using specific groups of insects (ground beetles, spiders) and plants (pioneer vegetation, willow scrubs and forests). Lateral connectivity is best indicated by the fish fauna.

Reference conditions

Compared to smaller streams **large rivers are relatively rare** and exposed to substantial human influence for centuries. This is why none of the large rivers, at least in most of Europe, meet near-natural reference conditions anymore. Due to intensive anthropogenic use (e.g., discharge of industrial and municipal waste water and/or cooling water, power generation, navigation, commercial fishery, water extraction, reclamation of agricultural land, flood protection works) biological reference communities cannot be described satisfactorily.

Referring to the **historical communities** the river once featured is possible for most of the economically relevant fish species in Rhine, Danube, Rhône, Seine or Elbe, but rather difficult for other biological elements. The historical invertebrate assemblages, for instance, can partly be reconstructed for the Rhine River due to scientific records dating from the early 20th century. But even then detailed accounts about individual abundances and locations of the species are missing. Without reference material the historical species identification is doubtful since the taxonomic assignment may have changed significantly over the years for certain groups.

Many of the once abundant and characteristic species of large rivers are **extinct** in most of Central Europe, their refugia have either completely disappeared or are too remote for natural recruitment (examples of macrozoobenthos genera extinct from the Rhine: *Prosopistoma*, *Marthamea* or *Palingenia*). Therefore, any re-establishment of the original biocoenosis remains unsuccessful. Since these taxa form part of the reference community, reaching high ecological status for large rivers by appropriate measures of revitalisation is basically impossible. However, the Water Framework Directive stipulates the achievement of good status or, in the case of most large rivers, good potential that depends on the definition of individual objectives in relation to the use of the rivers.

On the other hand many **non-native species** have immigrated, are now well-established and cannot be removed easily without harming the environment. Especially the biological communities of large, navigable watercourses were restructured several times due to the invasion of non-natives in the last 100 years. These synanthropic species are dispersed by shipping traffic and often occur in masses dominating the river coenosis. Strictly speaking, these biota are not part of the natural reference community, so how should they be considered in the assessment of ecological status? Here, European Member States have established different approaches.

Pressure-Impact Relationship

Another feature of large rivers is the cumulative effect of the various anthropogenic influences in the catchment. Abiotic differences are levelled out, and the biological assemblages are affected by a multitude of natural and anthropogenic factors. Moreover, the generally extensive use of large rivers implies a constantly high degree of pressure acting on the biocoenosis over the entire river course. A full pressure gradient ranging from undisturbed to highly degraded conditions is rarely present, therefore pressure-impact relationships are difficult to establish. In addition, the buffer capacity of large hydrosystems is higher than the one of small rivers.

The biological response to increasing or decreasing pressure seems to be less evident in large rivers compared to smaller watercourses. The recolonisation of certain invertebrate species, for instance, was temporally delayed at the Rhine and Elbe rivers (> 5 years) despite of the successful enhancement of water quality. Contaminated sediments, the "memory" of large rivers, are suspected to cause such effects. On the other hand, several large rivers are influenced by invasive species, which often dominate the benthic community. Therefore, their presence and invasiveness also can mask the positive effect of decreasing pressures and possible improvement of the benthic assemblages.

Data acquisition

Biological surveys at large rivers are elaborate, time-consuming and costly due to their size, their complexity, hydrological (e.g., water level fluctuations) and geomorphological (e.g., bedrock substrate) features. Thus, various methods of data acquisition are used in Europe, differing in sampling technique and frequency, number of replicates, sampled area and level of taxonomic identification. Regarding invertebrate sampling, results gained from various methods are difficult to compare (e.g., multi-habitat-sampling close to the river banks, dredging, air-lift-sampling, grab sampling, artificial substrates etc.). In general, sampling large hydrosystems is not only a matter of technique, but a matter of what the method targets to assess.

Annex III – Sources of uncertainty in the intercalibration exercise of invertebrate-based assessment methods

The following statements are derived from testing many different options for benchmarking (incl. no benchmarking) and analysing lots of graphs.

1) The two major sources of uncertainty in the outcome for a certain country are
   a) too few data, and
   b) the country covers only a very limited part of the pressure gradient.

   This becomes obvious in the figure below.

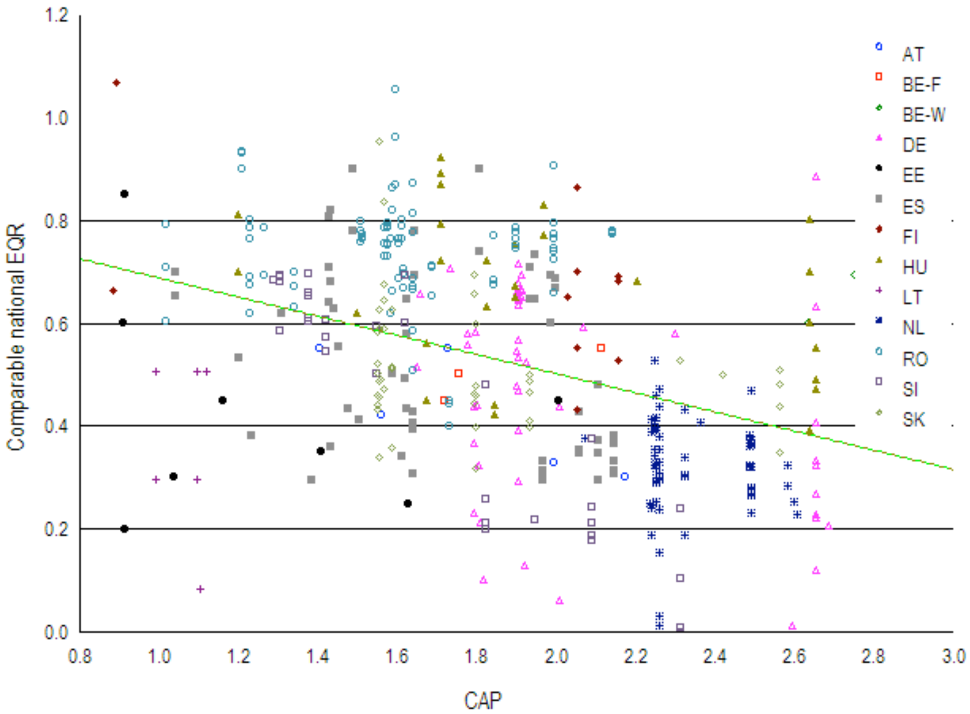


Figure: National assessment results in dependence of the pressure indicating the national origin of the sampling data (without any benchmarking and without involvement of common metrics). Assessment results are expressed as a continuous number with the class boundaries at 0.8, 0.6, 0.4 and 0.2 (= "Comparable national EQR")

In the graph above the assessment results the countries are plotted against the Combined Abiotic Pressure gradient (CAP) without any benchmarking or common metrics involved. To exemplify specific uncertainties of the exercise we briefly discuss the cases of Austria and the Netherlands: Both countries clearly have lower assessment results in comparison to the average (i.e. regression line). Since this stricter assessment of Austria is very clear and in accordance with the boundary comparison results via benchmark-standardised common metrics, the reason for the uncertainty is not in the intercalibration procedure itself. The uncertainty for Austria derives from the question if these five samples for Austria are representative for the Austrian large rivers. For the Netherlands the question is if the stricter assessment, which is very obvious at higher pressure levels, would also hold true for less impacted sites, for which the data are missing, or if the regression line would be steeper than for the other countries.

This question can only be answered with more data. The suggestion to compare its adjacent sites of different countries is interesting, but it would require really comparable sites and samples (season, morphology, year etc.). Independent of the outcome the comparisons would be based on even fewer samples which are also subject to a lot of individual variance. In the case of the Netherlands a direct comparison gave the same results as the whole exercise itself, maybe somewhat less pronounced: The closest sites of the River Rhine (two German and two Dutch sites) had a slightly stricter assessment for the Netherlands, but were all in poor status. The closest sites of the River Meuse which forms the boundary between the Netherlands and Belgium (three Flemish sites, one Walloon site and five Dutch sites) were assessed much stricter by the Netherlands (three samples moderate, 13 samples poor, three samples bad) than by Belgium (all four samples moderate). All in all, the problem remained that the comparison was mainly based on samples in poor status.

2) Another source of uncertainty is the combined pressure variable: It is composed of six very rough hydromorphological parameters based on expert judgement (navigation intensity, influence of damming, influence of impoundment, degree of water abstraction, degree of riparian habitat alteration, degree of channelization) plus two chemical parameters (orthophosphate concentration, nitrate concentration). Here, a) the parameter weighting and composition of the combined stressor metric might influence the final outcome, and b) the expert judgement might be biased in different countries and too imprecise.

Concerning a): We tried many different combined pressures (with or without chemistry, different hydromorphological parameters, different weighting etc.), and although the outcomes varied to some extent, the principal results were always the same (e.g. Finland and Hungary being more relaxed, and Slovenia being stricter with their assessments within the range of the data of each country). In conclusion, the composition and weighting of the pressure parameters is not that important, as long as several parameters are combined.

Concerning b): This might be an issue. In the cases of Austria and the Netherlands, however, the abiotic pressure index would have to become much worse in order to get the assessment results onto the average line (in the figure above Austria would have to be moved into the range of the Dutch data, and the Netherlands would have to be moved to the worse end of the scale, i.e. the best pressure index for the Netherlands would be worse than the worst for all other countries). In our opinion, this is unlikely. Getting more precise data for the hydromorphological indices might be desirable, but probably not achievable (e.g. how to describe impoundments more precise?).

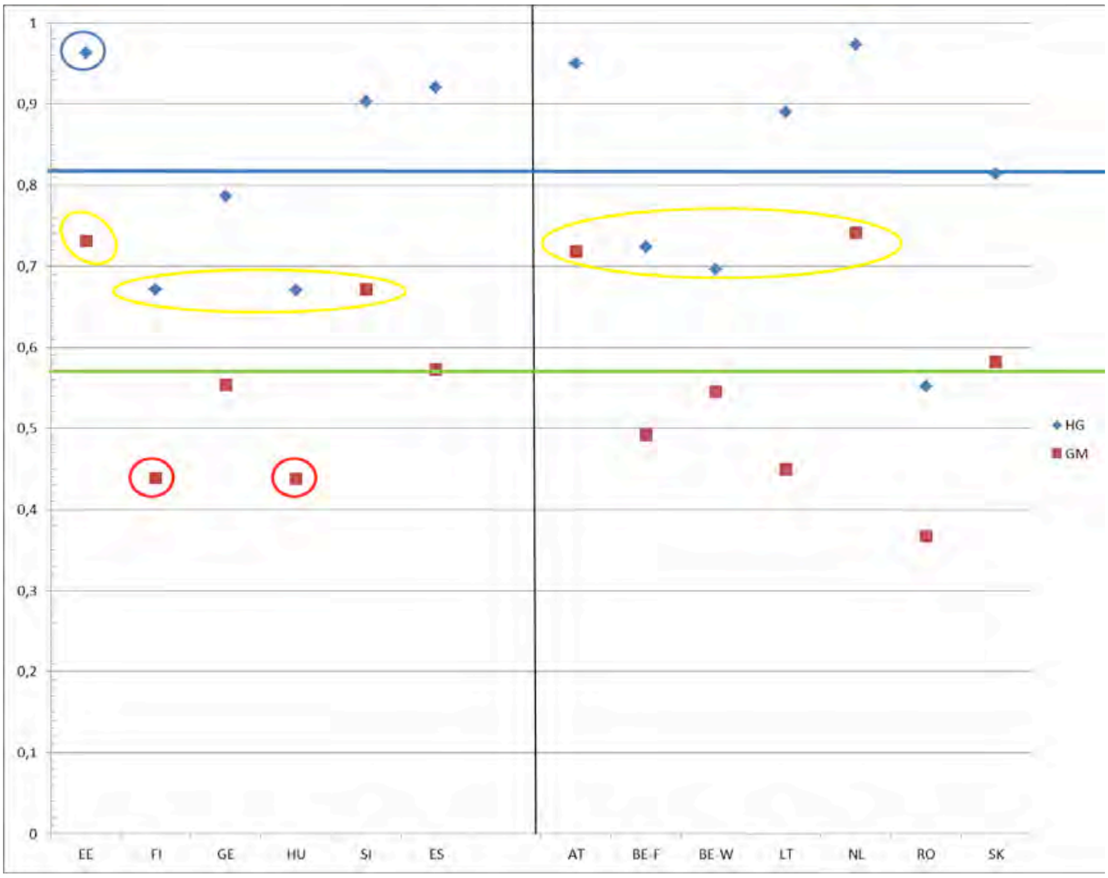
Annex IV – Austrian comments to the milestone report

We appreciate the work done under very difficult circumstances including scarce data availability and variability of typology and methods.

Obviously the variability is not eliminated with the very complex and sophisticated harmonisation procedures. Thus, the final result of the intercalibration exercise for macroinvertebrates should be interpreted very carefully.

The factors for the high degree of variability are clearly described in Chapter 8.3 of the report (Evaluating sources of uncertainty) and we fully agree with this interpretation - but we disagree with the proposed solution. We doubt that changing the acceptability criteria (boundary bias from 0.25 to 0.50 class width) is a good solution for this problem. It would mean that class boundary differences can be up to one entire status class.

Especially for benthic invertebrates the results show that for some MS HG boundary values are at the same level as the GM boundary values for other MS (or they are even lower!) – can this really be judged as being comparable? The following graph presents the results of IC results on ICM scale (Tables 8.2 to 8.4 of the milestone report):



red circles = necessary boundary adjustments, blue circles = possible boundary adjustments, yellow circles = overlapping HG/GM boundary

We think that it would be careless to change national boundaries on the basis of the intercalibration results with such a high degree of variability and sources of error, because this would endanger a reasonable national water management. At least in our opinion, it is not the priority objective to include our boundary values in the Commission Decision. The main goal should be achieving comparable classifications throughout Europe. And this goal has not been reached by this exercise.

From our point of view, instead of presenting a non-reasonable intercalibration it would be more straightforward to rather:

- present the difficulties of large river intercalibration;
- demonstrate with data that the intercalibration is not feasible at the moment with the given data set per type;
- compare national approaches and show differences within the intercalibration methods available now.

Annex V – Proposal for typifying European large rivers based on benthic invertebrates

Introduction

Large rivers are individuals characterised by their specific environmental conditions. However, the ecological status of European large rivers is assessed on a typological basis. To harmonise the national status classifications it is thus necessary to investigate which natural factors influence the biology at large rivers. This allows to evaluate how comparable the national assessments actually are. Such an analysis might reveal groups of rivers sharing similar environmental conditions and biological communities. But attempts to typify large rivers are hampered by the scarcity of near-natural river reaches lacking significant human influence. This document summarises methods and results of a typological study using benthic invertebrate fauna data from sites in least-disturbed conditions (LDC). Our findings suggest the distinction of four European large river types differing in size and climatic conditions.

Objectives

- Determining major factors influencing the distribution of the benthic invertebrate fauna in European large rivers;
- Defining criteria of LDC sites for this biological quality element;
- Allocating LDC sites at European large rivers;
- Identifying natural influences on the benthic fauna community at LDC sites;
- Defining types of large rivers inhabited by similar biological communities;
- Selecting biological metrics best indicating the faunistic differences among large river types.

Methods

1. We compiled data of **706 benthic fauna (BF) samples** from 305 sites at 48 large rivers located in 19 European countries (Figure 1).
2. Each site featured **supporting environmental information** covering natural variables (e.g. catchment size, discharge, altitude, water temperature) and pressure-variables (e.g. annual averages of physico-chemical parameters; degree of channelization, riparian habitat alteration, water abstraction; navigation activity).

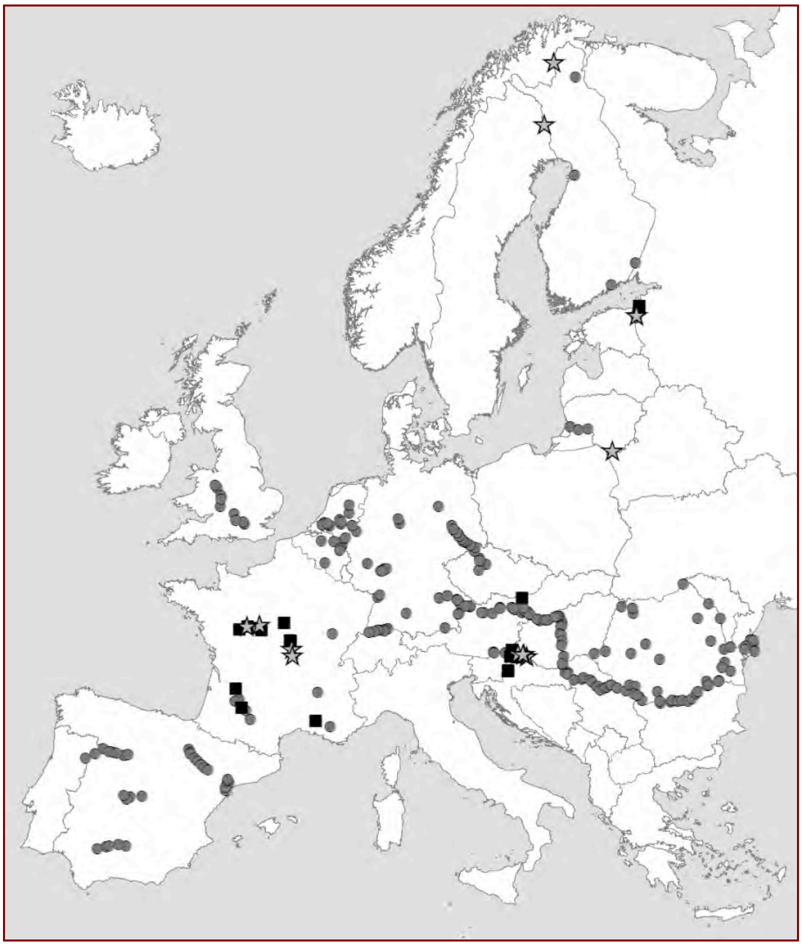


Figure 1: Location of BF-sampling sites used in this study (black squares: free-flowing sites without navigation but not in LDC, stars: LDC sites)

3. We used ordination analyses[25] to detect **environmental variables influencing the BF-community** (family-level, relative abundance).
4. **Indicator species analysis** (Dufrene & Legendre, 1997) was performed to identify the BF-families most indicative of the presence of navigation activity at 186 free-flowing[26] sampling sites.
5. We indentified **LDC sites** by selecting threshold values for pressure-variables that significantly influenced the BF-community.
6. BF-community data (family-level, abundance classes) from samples at LDC sites were processed by ordination analysis[27]. We related the main biological gradients to selected natural variables to discover (i) **groups of sites featuring similar BF-communities** under comparable environmental conditions, and (ii) **changes of the BF-**

[25] Detrended Correspondence Analysis, partial Canonical Correspondence Analysis (pCCA)
[26] no effects of damming or impoundment
[27] Non-metric Multidimensional Scaling (NMS)

communities along environmental gradients. Group differences were investigated by an analysis of similarity (ANOSIM; Clarke, 1993).

7. Selected **BF-metrics** and water temperature, catchment area and mean discharge were correlated based on 113 samples taken at free-flowing sites without navigation activity (including LDC sites; see Figure 1). Best correlated metrics were processed in a multiple regression analysis to determine the degree of metric-variance explained by the abiotic parameters.

8. We defined classes of water temperature and river-size relevant for the abiotic typification of European large rivers, and compared the distribution of selected BF-metrics between classes.

Results

1. **Navigation activity** had the strongest influence on the BF-community. BF-family richness of free-flowing sites, for instance, was significantly lower at sites influenced by navigation (Figure 2). Other significant pressure-variables were impoundment, water abstraction and various other morphological and physico-chemical parameters (Figure 3).

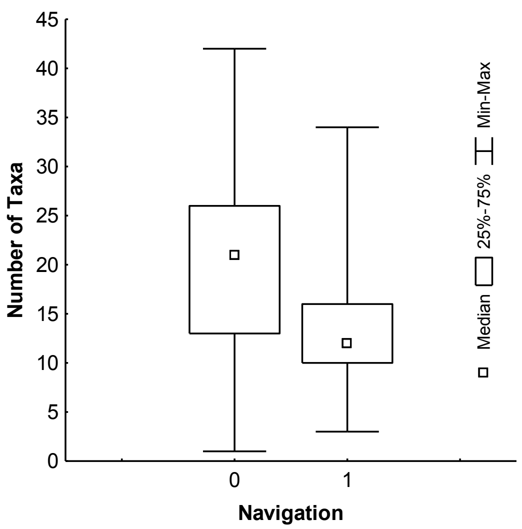


Figure 2: Taxa richness at sites in free-flowing reaches with (1) and without (0) navigation activity

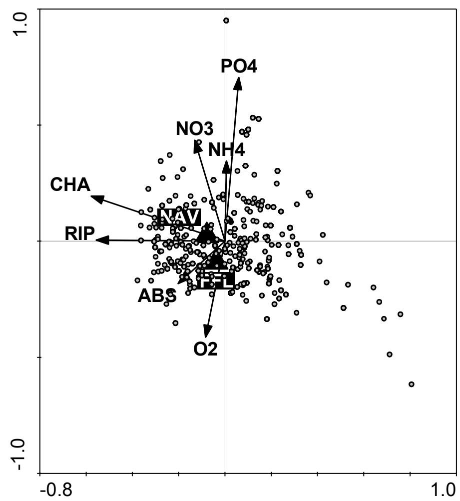


Figure 3: pCCA-biplot of samples (dots) and pressure-variables[28] (arrows and triangles) significantly explaining biological variability

2. Indicator species analysis revealed nine BF-families **indicative of** the presence of **navigation activity** (Table 1). Many of these families include neozoic species typical of

[28] NAV: navigation activity, FFL: site in free-flowing reach, ABS: degree of water abstraction, RIP: degree of riparian habitat alteration, CHA: degree of channelization, PO4: orthophosphate concentration, NH4: ammonium concentration, NO3: nitrate concentration, O2: minimum oxygen concentration

navigable waterways in Europe such as *Corophium curvispinum*, *Corbicula fluminea*, *Dreissena polymorpha*, *Dikerogammarus villosus* or *Jaera istri*.

Table 1: Indicator taxa of the presence of navigation activity. All Indicator Values (IV) are highly significant (p<0.001).

Taxon	Group	IV (%)
Janiridae	Crustacea	62.8
Hydrobiidae	Gastropoda	62.0
Gammaridae	Crustacea	61.3
Corophiidae	Crustacea	59.4
Dreissenidae	Bivalvia	52.8
Corbiculidae	Bivalvia	48.2
Planorbidae	Gastropoda	45.5
Ampharetidae	Polychaeta	37.6
Sphaeriidae	Bivalvia	35.3

3. We identified **13 sites at seven large rivers**[29] that met least-disturbed conditions (Figure 1, Table A1 and Figure A1 in the annex) based on the following **criteria**:
   - free-flowing;
   - no navigation activity;
   - at least good physico-chemical status[30];
   - no anthropogenic increase of water temperature;
   - channelisation: no more than slight alteration (less than 10% of the segment affected by "hard works"), no flow velocity increase;
   - no influence of upstream-dams;
   - no water abstraction;
   - riparian vegetation: no more than slight alteration of the adjacent natural vegetation;
   - near-natural vicinity evaluated by visual inspection of satellite images;
   - good overall river quality according to descriptions in Tockner et al. (2008).
4. The **BF-communities** differed between LDC sites with larger similarities among sites of the same region (river, country), and similar catchment areas/discharges and water temperatures (Figure 4). The taxa composition showed distinct changes along the gradient of average annual water temperature (Table 2).
5. Ordination analysis excluding all LDC sites ≥ 12 °C average water temperature resulted in NMS-axes 1 and 2 (explaining 85 % of community variance) highly **correlated to catchment area** (r=0.71 and 0.47) and **mean discharge** (r=0.57 and 0.68).
6. **% Rhithral-preferring taxa, % Active filter feeders** and the **Ratio of rhithral to potamal-preferring taxa** of samples at free-flowing sites without navigation were highly correlated to water temperature, catchment area and mean discharge. Multiple

[29] Allier, Loire, Mura, Narva, Nemunas, Teno, Torne
[30] minO_2 > 6 mg O_2/l, TP ≤ 0.2 mg P/l, PO_4 < 0.1 mg P/l, NH_4 < 0.3 mg N/l, NO_3 < 6 mg N/l, Cl^- < 200 mg/l

regression analysis revealed coefficients of determination up to 50% for these metrics (Table 3).

7. Based on the classes of water temperature[31] and river-size revealed from ordination analysis we identified **four European large river types** (Table 4) that showed significant differences in the BF-metric ranges for LDC and free-flowing sites without navigation (Figure A3 in the annex).

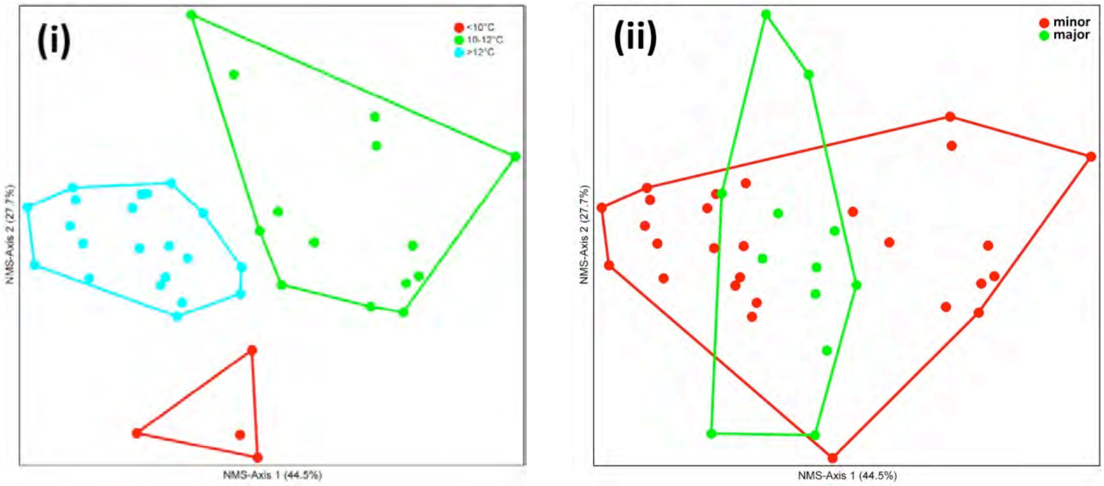


Figure 4: NMS-plots of the BF-communities of 39 samples at LDC sites (stress of 3-d solution: 13.2). Colour-codes represent samples taken at sites sharing similar ranges of (i) average annual water temperature[32] or (ii) catchment size and annual average discharge[33].

Only taxa occurring in >3 samples were included in the analysis.

ANOSIM R-Statistic:

0.801 between groups of countries;

0.762 between groups of rivers;

0.701 between groups of sites sharing similar average water temperature;

not significant between groups of sites sharing similar catchment size and discharge.

[31] We merged cold and temperate rivers to increase the data basis.

[32] < 10°C: cold (C), 10-12°C: temperate (T), >12°C: warm (W)

[33] minor: catchment area < 30,000 km^2, mean discharge ≤ 300 m^3/s; major: catchment area ≥ 30,000 km^2, mean discharge > 300 m^3/s – see Figure A2 in the annex for the classification rationale

Table 2: Spearman correlation coefficients (R_{Sp}; p<0.01) of average annual water temperature and BF-family abundance at LDC sites

Only taxa occurring in >3 samples were included in the analysis.

Taxon	Group	Representative Species/Genera	R_{Sp}
Rhyacophilidae	Trichoptera	*Rhyacophila nubila*	-0.57
Nemouridae	Plecoptera	*Amphinemura borealis* *Nemoura avicularis* *Nemoura cinerea* *Protonemura* sp.	-0.49
Taeniopterygidae	Plecoptera	*Brachyptera* sp. *Taeniopteryx nebulosa*	-0.48
Perlodidae	Plecoptera	*Arcynopteryx compacta* *Diura nanseni* *Isoperla obscura* *Perlodes* sp.	-0.45
Pediciidae	Diptera	*Dicranota* sp.	-0.41
Erpobdellidae	Hirudinea	*Dina punctata* *Erpobdella octoculata* *Erpobdella testacea*	-0.41
Dytiscidae	Coleoptera	*Nebrioporus depressus* *Platambus maculatus*	0.43
Corixidae	Heteroptera	*Micronecta griseola* *Sigara falleni*	0.45
Caenidae	Ephemeroptera	*Caenis horaria* *Caenis rivulorum*	0.63

Table 3: Summary of regression results based on 113 samples taken at free-flowing sites without navigation activity

All predictors are significant at p<0.001 except for * (p<0.05) and n.s. (not significant).

The beta coefficient specifies the relative contribution of each predictor to the overall prediction of the dependent variable.

Metric	Beta coefficient			Adj. R^2
	Catchment area	*Discharge*	*Water temperature*	
% Rhithral-preferring taxa	-0.81	0.42	-0.48	0.49
% Active filter feeders[34]	0.82	-0.23*	0.32	0.50
Ratio Rhithral- to Potamal-preferring taxa	-0.38	n.s.	-0.49	0.38

Table 4: Large river types in Europe based on the benthic fauna communities, grouped according to classes of catchment area (mean discharge) and average annual water temperature

Catchment area (mean discharge) / Average annual water temperature	< 30,000 km² (≤ 300 m³/s)	≥ 30,000 km² (> 300 m³/s)
< 12°C	**Temperate/cold minor large rivers**	**Temperate/cold major large rivers**
≥ 12°C	**Warm minor large rivers**	**Warm major large rivers**

34 $\log_{10}$-transformed

References of Annex V

Clarke KR (1993) Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology 18:117-143.

Dufrene M, Legendre P (1997) Species assemblages and indicator species: The need for a flexible asymmetrical approach. 67:345-366.

Tockner K, Uehlinger U, Robinson CT (Eds.) (2008) Rivers of Europe. Academic Press, Amsterdam. 728 pp.

Table A1: Descriptive statistics of selected environmental parameters at LDC sites

	Catchment area	Discharge	Altitude	Air temperature	Precipitation	Water temperature	Conductivity	pH
MEDIAN	13,676 km^2	280 m^3/s	110 m	11.5 °C	609 mm/a	11.1 °C	283 µS/cm	7.9
MIN	10,531 km^2	127 m^3/s	28 m	-2.4 °C	437 mm/a	5.5 °C	41 µS/cm	7.0
MAX	47,953 km^2	535 m^3/s	220 m	12.2 °C	735 mm/a	14.1 °C	431 µS/cm	8.3

	Avg. O_2	Min. O_2	TP-P	PO_4-P	NH_4-N	NO_3-N	Cl^{-1}
MEDIAN	10.6 mg/l	8.8 mg/l	0.083 mg/l	0.037 mg/l	0.045 mg/l	1.6 mg/l	12.4 mg/l
MIN	9.1 mg/l	6.1 mg/l	0.007 mg/l	0.002 mg/l	0.005 mg/l	0.2 mg/l	1.5 mg/l
MAX	13.7 mg/l	12.9 mg/l	0.200 mg/l	0.054 mg/l	0.100 mg/l	2.6 mg/l	17.0 mg/l

Figure A1: Sites in least-disturbed conditions – Narva 1 (top left), Narva 2 (top right), Teno (bottom left), Torne (bottom right) (all photos: Google Earth)

Figure A1 (cont.): Sites in least-disturbed conditions – Nemunas (top left), Mura 1 (top right), Mura 2 (bottom left), Mura 3 (bottom right) (all photos: Google Earth)

Figure A1 (cont.): Sites in least-disturbed conditions – Mura 4 (top left), Allier 1 (top right), Allier 2 (bottom left), Loire 1 (bottom right) (all photos: Google Earth)

Figure A1 (cont.): Sites in least-disturbed conditions – Loire 2 (photo: Google Earth)

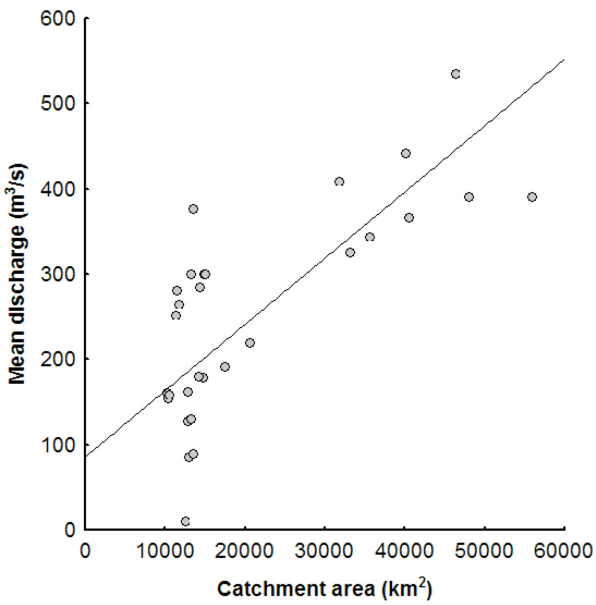


Figure A2: Relationship of catchment area and mean discharge for the 32 free-flowing sites without navigation activity. Two clusters of sites are discernable separated by a catchment area of about 30,000 km^2. This value corresponds to a mean discharge of 300 m^3/s.

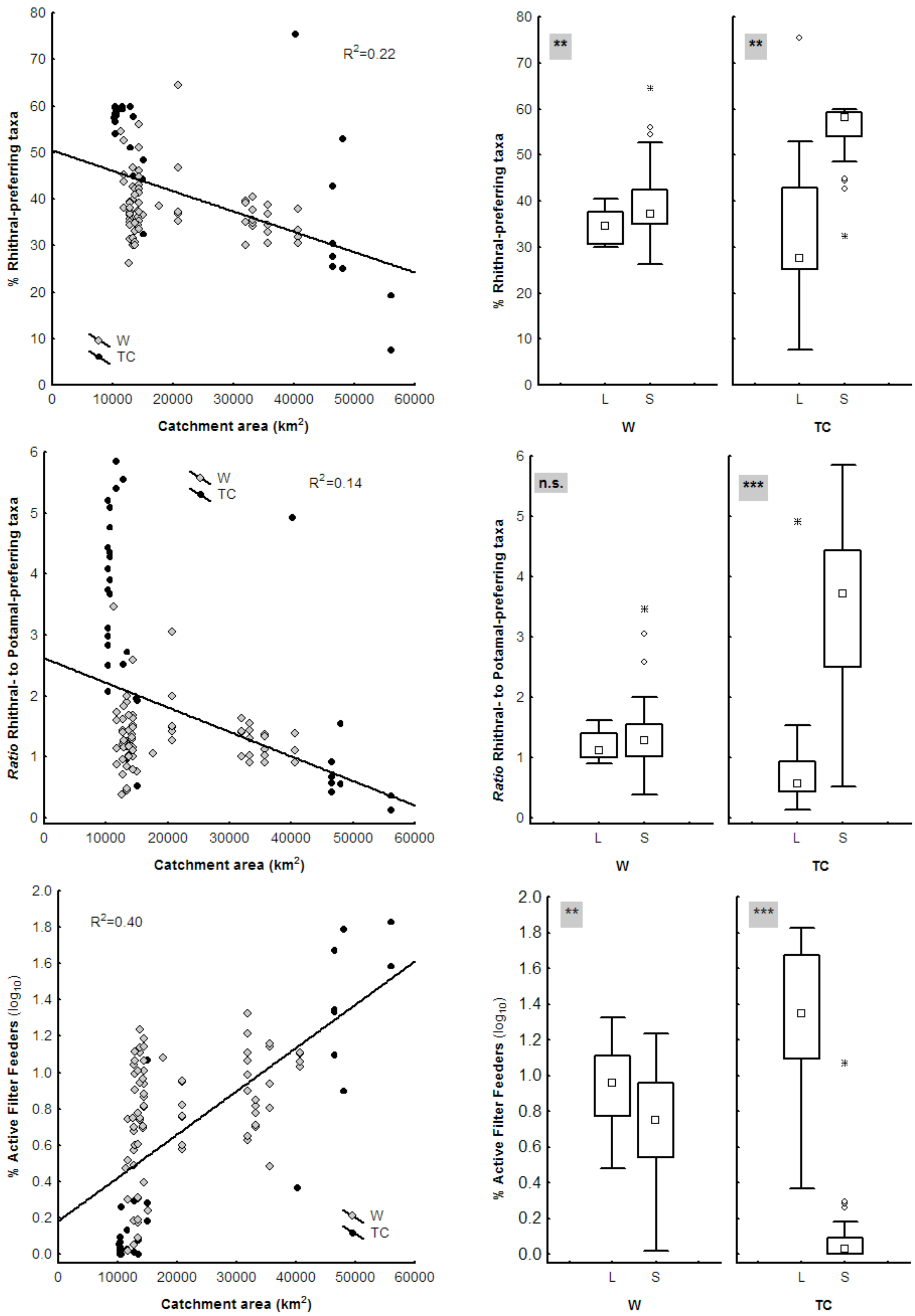


Figure A3: Relationships between catchment area and selected BF-metrics of 113 samples taken at free-flowing sites without navigation, belonging to either temperate/cold (TC) or warm (W) rivers. R^2-values specify coefficients of determination gained from regression analysis. Asterisks specify significance-levels of the Mann-Whitney U-test between size-categories (L-major, S-minor) within temperature-groups (*** - p<0.001, ** - p<0.01, n.s. - not significant).

Annex VI – Two proposals for typifying European large rivers based on benthic diatoms

Introduction

Do the benthic diatom communities naturally differ among European large rivers? Can we assign groups of large rivers (i.e. types) that feature similar communities? Would these differences be relevant in bioassessment and, hence, in the intercalibration of national assessment methods?

If such differences exist reference conditions and degradation pathways would differ among river types. This needs to be considered when the national assessment methods are compared and harmonised. A certain level of nutrient enrichment, for example, would impact differently on the diverse river types.

This summary presents the methods and results of an analysis to define large river types based on benthic diatoms. We propose **two alternative typologies** based on different interpretations of the results. The outcomes of this work will be discussed at the next XGIG Large River IC meeting in Koblenz in September.

Objectives

- Defining criteria for diatom sampling sites in least-disturbed conditions (LDC);
- Allocating LDC-sites at European large rivers;
- Identifying natural influences on the diatom community at LDC-sites;
- Defining types of large rivers inhabited by similar diatom communities;
- Deriving diatom metric values representing biological LDC;
- Identifying indicator species characterising the river types.

Methods

1. We compiled **465 diatom samples** from 271 sites at 46 large rivers located in 19 European countries (Figure 1).
2. Each site featured **supporting environmental information** covering natural variables (e.g. catchment size, discharge, altitude, water temperature, alkalinity) and pressure-variables (e.g. annual averages of nutrient concentrations, chloride, water abstraction, navigation activity).

3. The **biological metrics** *Indice de Polluosensibilité Spécifique* (IPS; Coste in CEMAGREF 1982) and *Rott's Trophic Index* (RT; Rott et al. 1999) were calculated for each sample based on an adjusted taxonomic indicator list (Kelly & Ector, in press).

4. Correlation analysis and multiple regression analysis were performed to indentify the **influences of the environmental variables** on the biological metrics.

5. We indentified **LDC-sites** by selecting threshold values for pressure-variables that significantly influenced the biological metrics.

6. Diatom community data from samples at LDC-sites were processed by **ordination analysis**[35]. We related the main biological gradient to selected non-biological variables to discover groups of sites featuring similar diatom communities under comparable environmental conditions.

Methods of 1st Proposal: Three river types

7. For each group we described the most characteristic **environmental features** and determined the **biological metric ranges**.

8. **Indicator species analysis** (Dufrene & Legendre 1997) was performed to identify the diatom taxa most indicative for each group.

Methods of 2nd Proposal: Two river types

7. For each group we tested if the concentrations of **orthophosphate** significantly influenced the biological metrics. If not, we defined metric values representing **least-disturbed biological conditions** from all samples belonging to the group. If there was a significant influence, we **predicted** the LDC-metric values at an orthophosphate concentration of 16 µg P/l[36] by regression analysis.

8. **Indicator species analysis** (Dufrene & Legendre 1997) was performed to identify the diatom taxa most indicative for each group. This analysis only included sites with average orthophosphate concentrations of ≤ 16 µg P/l.

[35] Non-metric Multidimensional Scaling

[36] Behrendt et al. (2003) modelled geogenic background concentrations of TP= 30 µg/l for the large rivers Danube, Elbe, Rhine, Weser. This concentration corresponds to PO_4 = 16 µg P/l based on regression analysis using the available data (PO_4-P = 0.4468*TP + 0.0026, R^2=0.63, n=419).

Figure 1: Location of the 271 diatom sampling sites including 44 sites in least-disturbed conditions (star-symbol)

Results

1. **Spearman correlation** identified the following parameters correlated to the biological metrics in the full dataset with R_{Sp}>0.3:
   - **IPS:** water temperature, chloride, orthophosphate (R_{Sp}>0.4); mean annual air temperature (R_{Sp}>0.3);
   - **RT:** chloride, orthophosphate (R_{Sp}>0.4); ammonium, water temperature, minimum oxygen concentration (R_{Sp}>0.3).
2. **Multiple regression of all environmental variables** gained the following parameters significantly explaining the metric variability with beta >|0.2|:
   - **IPS:** mean annual air temperature, water temperature, navigation activity, orthophosphate, mean annual precipitation
   - **RT:** navigation activity, mean annual air temperature, altitude, orthophosphate

3. **Multiple regression of pressure-variables** (after removing the effect of natural variables) resulted in the following parameters with significant explanatory power (beta >|0.2|):
   - **IPS:** orthophosphate, water abstraction;
   - **RT:** orthophosphate, navigation activity.
4. We identified 44 **LDC-sites** (see Figure 1) after screening according to the criteria:
   - mean annual average orthophosphate concentration ≤ 40 µg P/l[37],
   - mean annual minimum oxygen concentration ≥ 6 mg O_2/l,
   - mean annual average ammonium concentration ≤ 100 µg N/l[3],
   - no navigation activity,
   - sites not more than slightly affected by water abstraction (less than 10% of the median annual flow and the median monthly flow during a critical period, e.g. low flow period).

Results of 1st Proposal: Three river types

5. **Ordination analysis** revealed three distinct diatom community types differing between ecoregions (Figure 2).
6. **Nordic rivers** were represented by 8 sampling sites located at the rivers Dalälven, Kymijoki, Lule älv, Oulujoen, Paatsjoki, Teno and Torne. This river type is characterised by low water temperatures (mean value: 6.8 °C), low water alkalinity (mean value: 0.27 meq $CaCO_3$/l) and slightly acidic pH (mean value: 7.0) (Figure 3). Average PO_4-concentration amounts to 4 µg P/l[38].

   The average (median) values ± standard deviation of the diatom metrics (Figure 4) are:
   - **IPS:** 17.4 (17.8) ± 1.2,
   - **RT:** 1.53 (1.44) ± 0.27.

[37] reference threshold value for large lowland rivers (1,000-10,000 km^2 catchment area) according to Bennett et al. (2011)
[38] average TP: 14 µg P/l

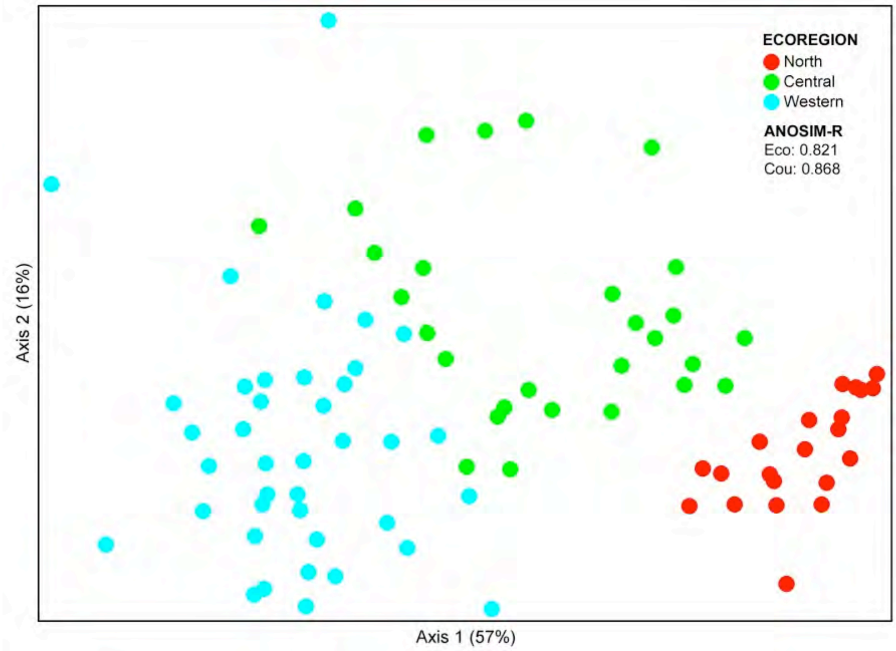


Figure 2: NMS-ordination plot of 88 diatom samples taken at 44 LDC-sites. Colour-codes depict samples belonging to the same ecoregion.

7. **Central rivers** comprised 14 sampling sites at the Drava, Inn, Mura, Rhone and Motala Ström. Except for the latter all rivers have Alpine influence. Moderate water temperatures (mean value: 11.1 °C) and an average PO_4-concentration of 14 µg P/l[39] characterise this river type (Figure 3).

   The average (median) values ± standard deviation of the diatom metrics (Figure 4) were:

   - **IPS:** 15.3 (15.4) ± 1.2,
   - **RT:** 2.71 (2.77) ± 0.38.

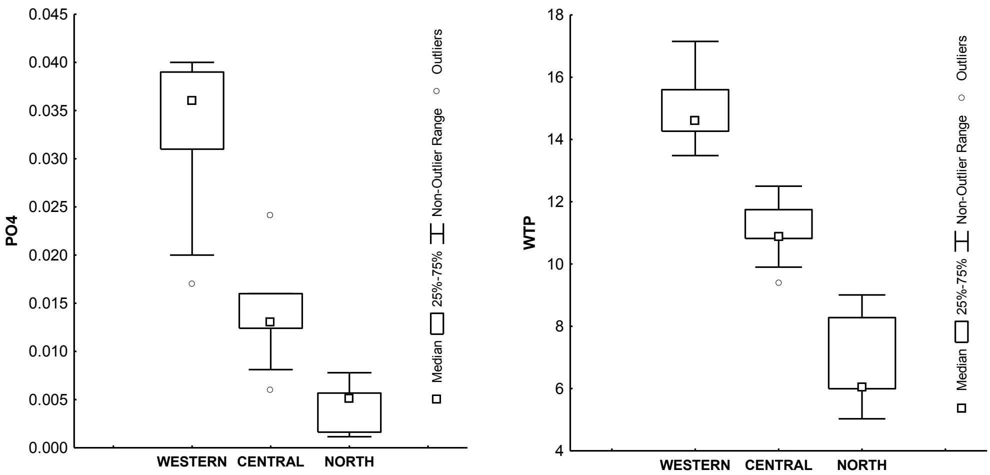


Figure 3: Distribution of orthophosphate concentration (PO4) and water temperature (WTP) among sites of the three large river types.

[39] average TP: 43 µg P/l

8. **Western rivers** were represented by 22 sampling sites at the Cher, Dordogne, Garonne, Loire, Lot, Tarn and Vienne. This type features higher water temperatures (mean value: 14.9 °C) and PO_4-concentrations (mean value: 33 µg P/l[40]) (Figure 3). The average (median) values ± standard deviation of the diatom metrics (Figure 4) were:
   - **IPS:** 11.8 (11.6) ± 1.9,
   - **RT:** 3.11 (3.14) ± 0.28.

9. **Indicator species analysis** resulted in 28 diatom taxa indicative of the three large river types (Table 1).

10. **Metric ranges** showed significant differences between river types (KW-Test: $p<0.01$) (Figure 4).

[40] average TP: 68 µg P/l

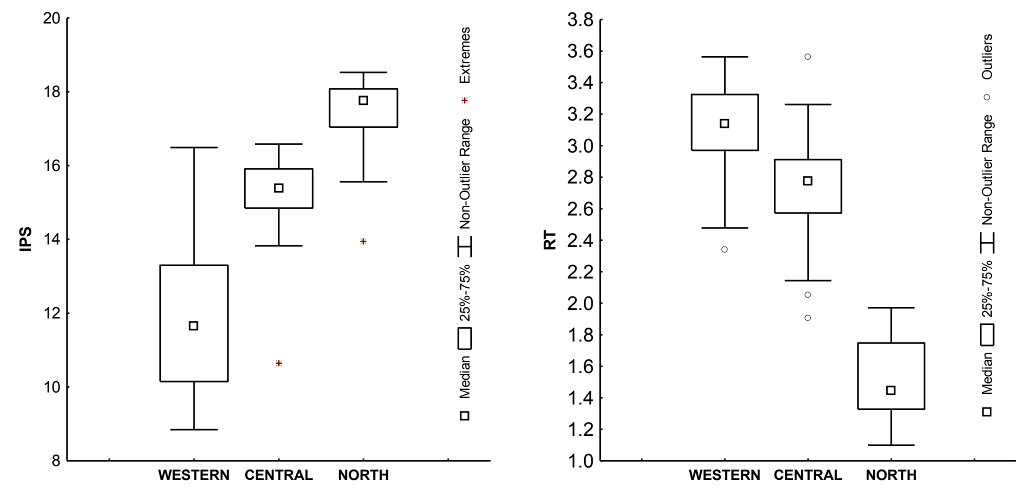


Figure 4: Distribution of diatom metric values among sites of the three large river types.

Table 1: Diatom taxa resulting from indicator species analysis of LDC-sites representing three European river types (only Indicator Values ≥ 40 %; all values are significant at p ≤ 0.05)

River type	Taxon code	Taxon name	Indicator Value (%)
Nordic	RPUS	*Achnanthidium pusillum* (Grun.in Cl. & Grun) Czarnecki	90.9
	BNEO	*Brachysira neoexilis* Lange-Bertalot	81.8
	SYNT	*Fragilaria tenera* (W.Smith) Lange-Bertalot	81.8
	SPIN	*Fragilaria pinnata* Ehrenberg var. *pinnata*	66.8
	PSAT	*Achnanthes subatomoides* (Hustedt) Lange-Bertalot et Archibald	63.2
	EINC	*Eunotia incisa* Gregory var.incisa	59.1
	EIMP	*Eunotia implicata* Nörpel. Lange-Bertalot & Alles	54.5
	GACU	*Gomphonema acuminatum* Ehrenberg	54.5
	EDES	*Cymbella descripta*	45.5
	NIPM	*Nitzschia perminuta* (Grunow) M.Peragallo	45.5
	NRAD	*Navicula radiosa* Kützing	43.8
	DITE	*Diatoma tenuis* Agardh	42.2
	CROS	*Cyclotella rossii* Hakansson	40.9
	PGIB	*Pinnularia gibba* Ehrenberg	40.9
	PLVD	*Achnanthes levanderi* Hustedt	40.9
Central	GOLI	*Gomphonema olivaceum* (Hornemann) Brébisson var. *olivaceum*	85.2
	RABB	*Rhoicosphenia abbreviata* (C.Agardh) Lange-Bertalot	73.4
	DVUL	*Diatoma vulgaris* Bory	52.0
	NMEN	*Navicula menisculus* Schumann var. *menisculus*	44.4
	NSIO	*Nitzschia sigmoidea* (Nitzsch)W. Smith	44.4
	SBRE	*Surirella brebissonii* Krammer & Lange-Bertalot var.*brebissonii*	43.8
	FVAU	*Fragilaria capucina* Desmazieres var.*capitellata* (Grunow) Lange-Bertalot	41.6
	SOVI	*Surirella ovalis* Brebisson	40.7
	EPRO	*Cymbella prostrata* (Berkeley) Grunow (*Encyonema*)	40.6
Western	NANT	*Navicula antonii* Lange-Bertalot	51.6
	NAMP	*Nitzschia amphibia* Grunow f.*amphibia*	48.6
	NPAE	*Nitzschia paleacea* (Grunow) Grunow in van Heurck	48.3
	GPAR	*Gomphonema parvulum* (Kützing) Kützing var. *parvulum* f. *parvulum*	44.8

Results of 2[nd] Proposal: Two river types

5. **Ordination analysis** revealed distinct diatom communities differing between low-alkalinity, slightly acidic (< 0.5 meq $CaCO_3$/l; pH < 7.5) and medium- to high-alkalinity, neutral rivers (> 0.5 meq $CaCO_3$ /l; pH > 7.5) (Figure 5).

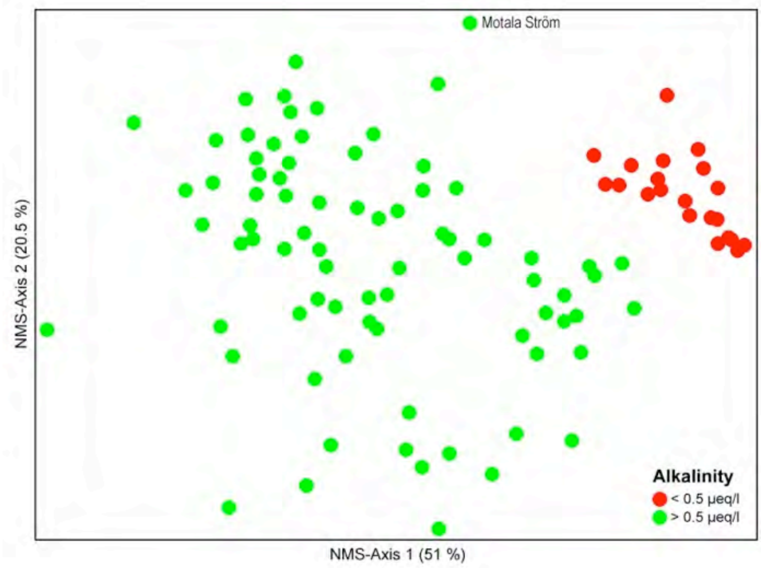


Figure 5: NMS-ordination plot of diatom samples taken at 44 LDC-sites. Colour-codes depict samples belonging to the same alkalinity and pH classes.
Motala Ström showed low alkalinity (0.86 meq $CaCO_3$/l) but a diatom community and metric values different from the low-alkalinity, slightly acidic rivers.

6. **Low-alkalinity, slightly acidic rivers** showed an alkalinity range of 0.15 to 0.32 meq $CaCO_3$/l (pH: 6.8 - 7.2) and were represented by 8 sampling sites located at the Scandinavian rivers Dalälven, Kymijoki, Lule älv, Oulujoen, Paatsjoki, Teno and Torne. Average concentrations of PO_4 amounted to 4 µg P/l[41]. Diatom metrics were uncorrelated to the orthophosphate gradient.

   The average (median) values ± standard deviation of the diatom metrics were:

   - **IPS:** 17.4 (17.8) ± 1.2,
   - **RT:** 1.53 (1.44) ± 0.27.

7. **Medium- to high-alkalinity, neutral rivers** had a range of 0.86 to 2.92 meq $CaCO_3$/l alkalinity (pH: 7.6 - 8.4). The 36 LDC-sites of this group covered a PO_4-gradient with higher concentrations at larger lowland rivers (Figure 6). The diatom metrics were significantly correlated with the orthophosphate concentrations.

[41] average TP: 14 μg P/l

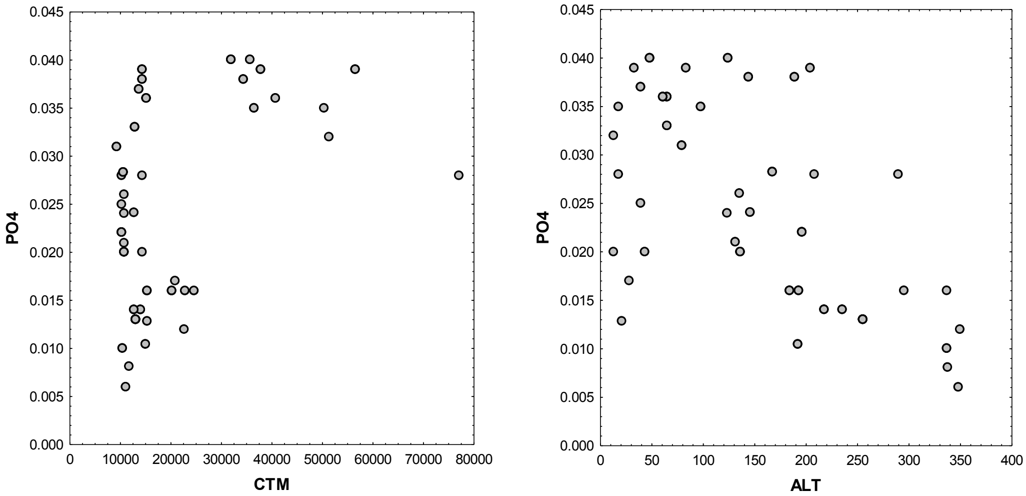


Figure 6: Orthophosphate concentrations of samples taken at different catchment sizes (CTM) and altitudes (ALT) of the LDC-sites.

8. By regression analysis we **predicted** the following metric values at an orthophosphate concentration of 16 µg P/l:
   - **IPS:** 14.7 ± 0.5 (95% confidence limit) at an adjusted R^2 of 0.56,
   - **RT:** 2.74 ± 0.09 (95% confidence limit) at an adjusted R^2 of 0.47.
9. **Indicator species analysis** resulted in 20 diatom taxa highly indicative of the two large river types (Table 2).
10. **Metric ranges** showed significant differences between river types (U-test: p<0.01) (Figure 7).

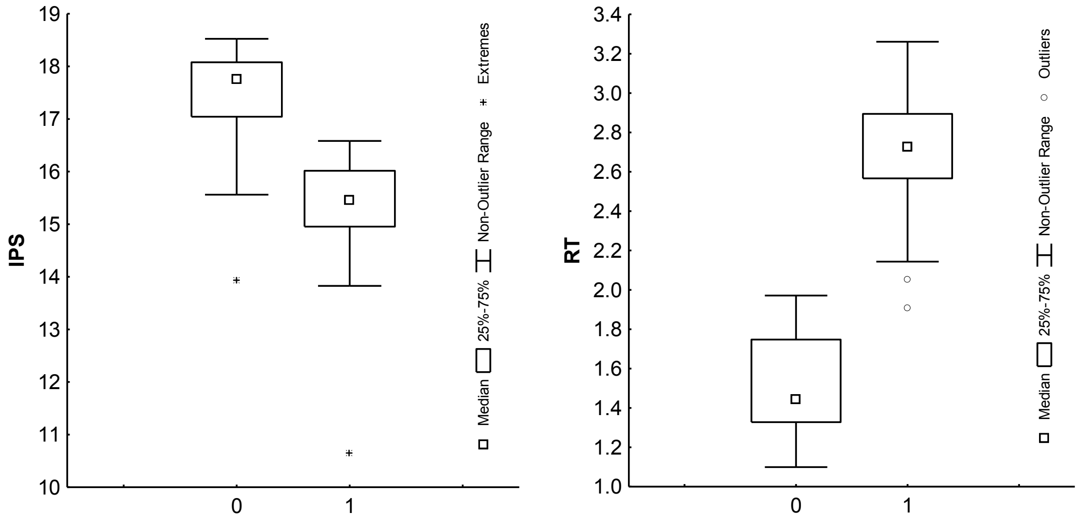


Figure 7: Ranges of diatom metrics at LDC-sites of low-alkalinity, slightly acidic rivers (0) and medium- to high-alkalinity, neutral rivers (1). The latter type only included samples with PO_4-concentrations ≤ 16 µg P/l.

Table 2: Diatom taxa resulting from indicator species analysis of LDC-sites representing two European river types (only Indicator Values ≥ 50 %; all values are significant at $p \leq 0.01$)

Low-alkalinity, slightly acidic rivers - alkalinity range 0.15-0.32 meq $CaCO_3$/l, average PO_4: 4 µg P/l, rivers: Dalälven, Kymijoki, Lule älv, Oulujoen, Paatsjoki, Teno, Torne

Medium to high alkalinity, neutral rivers - alkalinity range: 0.86-2.92 meq $CaCO_3$/l, average PO_4: 12 µg P/l, rivers: Drava, Rhone, Inn, Motala Ström. The type only included samples with PO_4-concentrations ≤ 16 µg P/l.

River group	Taxon code	Taxon name	Indicator Value (%)
Low-alkalinity, slightly acidic rivers (< 0.5 meq $CaCO_3$/l; pH < 7.5)	SYNT	*Fragilaria tenera* (W.Smith) Lange-Bertalot	85.7
	BNEO	*Brachysira neoexilis* Lange-Bertalot	80.4
	PSAT	*Achnanthes subatomoides* (Hustedt) Lange-Bertalot et Archibald	66.7
	SPIN	*Fragilaria pinnata* Ehrenberg var. *pinnata*	62.2
	EINC	*Eunotia incisa* Gregory var. *incisa*	61.9
	EIMP	*Eunotia implicata* Nörpel. Lange-Bertalot & Alles	57.1
	GACU	*Gomphonema acuminatum* Ehrenberg	57.1
Medium- to high-alkalinity, neutral rivers (> 0.5 meq $CaCO_3$/l; pH > 7.5)	RABB	*Rhoicosphenia abbreviata* (C.Agardh) Lange-Bertalot	90.0
	DVUL	*Diatoma vulgaris* Bory	90.0
	NLAN	*Navicula lanceolata* (Agardh) Ehrenberg	90.0
	GOLI	*Gomphonema olivaceum* (Hornemann) Brébisson var. *olivaceum*	85.0
	NTPT	*Navicula gracilis* Ehrenberg	73.6
	EPRO	*Cymbella prostrata* (Berkeley) Grunow (*Encyonema*)	70.0
	NFON	*Nitzschia fonticola* Grunow in Cleve et Möller	67.1
	MVAR	*Melosira varians* Agardh	60.0
	RSIN	*Cymbella sinuata* Gregory	59.1
	NLIN	*Nitzschia linearis* (Agardh) W.M.Smith var.*linearis*	52.0
	NMEN	*Navicula menisculus* Schumann var. *menisculus*	50.0
	NSIO	*Nitzschia sigmoidea* (Nitzsch) W. Smith	50.0
	SOVI	*Surirella ovalis* Brebisson	50.0

References of Annex VI

Behrendt H, Bach M, Kunkel R, Opitz D, Pagenkopf WG, Scholz G, Wendland F (2003) Nutrient Emissions into River Basins of Germany on the Basis of a Harmonized Procedure. UBA-Texte 82/03. Umweltbundesamt, Berlin. 191 pp.

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

CIS (Common Implementation Strategy) (2010) Guidance document on the intercalibration process 2008-2011. Guidance Document No. 14. Implementation Strategy for the Water Framework Directive (2000/60/EC). Brussels. 102 pp.

CEMAGREF (1982) Etude de Méthodes Biologiques Quantitatives d'Appreciation de la Qualité des Eaux. Rapport Q.E. Lyon-A.F.B. Rhône-Mediterranneé-Corse. 218 pp.

Dufrene M, Legendre P (1997) Species assemblages and indicator species: The need for a flexible asymmetrical approach. 67:345-366.

Kelly MG, Ector L (2012) Effect of streamlining taxa lists on diatom-based indices: implications for intercalibrating ecological status. Hydrobiologia in press.

Rott E, Pfister P, Dam H van, Pipp E, Pall K, Binder N, Ortler K (1999) Indikationslisten für Aufwuchsalgen in Österreichischen Fließgewässern, Teil 2: Trophieindikation und autökologische Anmerkungen. 248 pp.

References

Aroviita J, Koskenniemi E, Kotanen J, Hämäläinen H (2008) A priori typology-based prediction of benthic macroinvertebrate fauna for ecological classification of rivers. Environmental Management 42:894-906.

Berg MS, Coops HC, Pot R, Altenburg W, Nijboer R, Broek Tv, Fagel M, Arts G, Bijkerk R, Dam Hv, Ietswaart T, Molen Jv, Wolfstein K, Jong Dd, Hartholt H (2004) Achtergronddocument referenties en maatlatten waterflora. RIZA, Lelystad.

Coste M, Boutry S, Tison-Rosebery J, Delmas F (2009) Improvements of the Biological Diatom Index (BDI): Description and efficiency of the new version (BDI-2006). Ecological Indicators 9:621-650.

Eloranta P, Soininen J (2002) Ecological status of some Finnish rivers evaluated using benthic diatom communities. Journal of Applied Phycology 14:1-7.

Gabriels W, Lock K, De Pauw N, Goethals PL (2010) Multimetric Macroinvertebrate Index Flanders (MMIF) for biological assessment of rivers and lakes in Flanders (Belgium). Limnologica - Ecology and Management of Inland Waters 40:199-207.

Hendrickx A, Denys L (2005) Toepassing van verschillende biologische beoordelingssystemen op Vlaamse potentiële interkalibratielocaties overeenkomstig de Europese Kaderrichtlijn Water - Partim "Fytobenthos". Rapport Instituut voor Natuurbehoud, Brussels.

Hlubikova D, Miettinen J, Hindakova A (2006) Testing of diatom water quality indices in monitoring and reference sites of Slovakian rivers (Central Europe). In: Ács É, Kiss KT, Padisák J, Szabó S 6th International Symposium on the Use of Algae for Monitoring Rivers. Magyar Algológiai Társaság, Göd, pp. 62-67.

Leyssen A, Denys L, Schneiders A, Van Looy K, Packet J, Vanhecke L (2006) Afstemmen van referentiecondities en evaluatiesystemen voor de biologische kwaliteitselementen macrofyten en fytobenthos en uitwerken van een meetstrategie in functie van de Kaderrichtlijn Water. Rapport Instituut voor Natuurbehoud, Brussels.

Marvan P, Heteša J (2006) Metodika odběru a zpracování vzorků fytobentosu tekoucích vod. Standardní operační postup. Oddělení experimentální fykologie a ekotoxikologie BÚ AV ČR.www.mzp.cz/cz/prehled_akceptovanych_metodik_tekoucich_vod.

Marvan P, Opatřilová L, Heteša J (2011) Metodika hodnocení ekologického stavu útvarů povr-chových vod tekoucích pomocí biologické složky fytobentos.Závěrečná zpráva. (Final report) Výzkumný ústav vodohospodářský T. G. Masaryka, v.v.i., Praha,15 pp.

Munne A, Prat N (2009) Use of macroinvertebrate-based multimetric indices for water quality evaluation in Spanish Mediterranean rivers : an intercalibration approach with the IBMWP index. Hydrobiologia 10:203-225.

Novak, MA, Bode, RW (1992) Percent Model Affinity - A New Measure of Macroinvertebrate Com-munity Composition. Journal of the North American Benthological Society 11:80-85.

Ofenböck T, Moog O, Stubauer I, Huber T, Leitner P (2005) Entwicklung eines flächendeckend anwendbaren Systems zur Beurteilung des ökologischen Zustandes auf Basis des Makrozoobenthos - Endbericht. Universität für Bodenkultur, Wien:91.

Pfister P, Pipp E (2010) Leitfaden zur Erhebung der biologischen Qualitätselemente - Teil A3 - Phytobenthos. Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft, Wien:96.

Rolauffs P, Hering D, Sommerhäuser M, Rödiger S, Jähnig S (2003) Entwicklung eines leitbildorientierten Saprobienindexes für die biologische Fließgewässerbewertung. Umweltbundesamt, Berlin.

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

Schöll F, Haybach A, König B (2005) Das erweiterte Potamontypieverfahren zur ökologischen Bewertung von Bundeswasserstraßen (Fließgewässertypen 10 und 20: kies- und sandgeprägte Ströme, Qualitätskomponente Makrozoobenthos) nach Maßgabe der EU-Wasserrahmenrichtlinie. Hydrologie & Wasserbewirtschaftung 49:234-247.

Šporka F, Pastuchová Z, Hamerlík L, Dobiašová M, Beracko P (2009) Assessment of running wa-ters (Slovakia) using benthic macroinvertebrates — derivation of ecological quality classes with respect to altitudinal gradients. Biologia 64:1196-1205.

Stubauer I (2002) Ausweisung saprobieller Grundzustände österreichischer Fließgewässer. Dissertation zur Erlangung des Doktorgrades an der Universität für Bodenkultur. Wien.

Timm H, Vilbaste S (2010) Estonian methods of the ecological status assessment of the surface water bodies - the benthic diatom ecosystem in rivers - the benthic macroinvertebrate ecosystem in rivers and lakes. Contract 4-1.1/166 [in Estonian]. Estonian University of Life Sciences, Tartu. 112 pp.

van Dam H, Stenger-Kovács C, Ács E, Borics G, Buczkó K, Hajnal É, Soróczki-Pintér É, Várbíró G, Tóthmérész B, Padisák J (2007) Implementation of the European Water Framework Directive: Development of a system for water quality assessment of Hungarian running waters with dia-toms. Archiv für Hydrobiologie, Suppl. Large Rivers 17:339-364.

Vanden Bossche JV, Usseglio-Polatera P (2005) Characterization, Ecological Status and Type-specific Reference Conditions of Surface Water Bodies in Wallonia (Belgium) using Biocenotic Metrics based on Benthic Invertebrate Communities. Hydrobiologia 551:253-271.

Várbíró G, Borics G, Csányi B, Fehér G, Grigorszky I, Kiss KT, Tóth A, Àcs E (2011b) Improvement of the ecological water qualification system of rivers based on first results of the Hungarian phytobenthos surveillance monitoring. Hydrobiologia (in print).

Várbíró G, Fekete O, Ortmann-Ajkai A, Ficsor M, Cser B, Kovács K, Kiss G, Czirok A, Horvai V, Deák CS (2011a) Developing a multimetric macroinvertebrate index on mountainous, small and medium sized water bodies. Acta Biol. Debr. Oecol. Hung. 26:211-220.

Vilbaste S (2004) Application of diatom indices in evaluation of water quality in Estonian running waters. Proceedings of the Estonian Academy of Sciences, Biology and Ecology 53:37-51.