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:	**CBrivGIG Macrophytes**
Information provided by:	Sebastian Birk and Nigel Willby with contributions from Annette Baattrup-Pedersen, Christian Chauvin, Luc Denys, Daniel Galoux, Martin McGarrigle, Maria Rita Minciardi, Karin Pall, Roelf Pot, Krzysztof Szoszkiewicz

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:**

Lead, DE – Sebastan Birk

UK – Nigel Willby, Geoff Phillips
AT – Karin Pall
BE (FL) – An Leyssen, Luc Denys
BE (WL) – Daniel Galoux
CZ – Libuse Opatrilova, Pavla Wildova
DE – Jochen Schaumburg, Christine Schranz
DK – Annette Baattrup-Pedersen, Lars K. Larsen
EE – Aive Kors
ES – Isabel Pardo
FR – Christian Chauvin
IE – Martin McGarrigle
IT – Maria Rita Minciardi
LT – Mindaugas Gudas
LU – Nora Welschbillig, Alain Dohet, Gérard Schmidt
LV – Solvita Strazdina
NL – Roelf Pot
PL – Krzysztof Szoszkiewicz

1.2. Participation

Indicate which countries are participating in your group. Are there any difficulties with the participation of specific Member States? If yes, please specify:

Austria, Belgium, Czech Republic, Denmark, Estonia, France, Germany, Ireland, Italy, Latvia, Lithuania, Luxembourg, Netherlands, Poland, Spain, United Kingdom

1.3. Meetings

List the meetings of the group:

Edinburgh (UK), March 13-14, 2008
Copenhagen (DK), April 29-30, 2009
Bordeaux (FR), May 27-28, 2010
Turin (IT), May 12-13, 2011

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

ID[1]	Country	Method	Abbr.	State
Finalized methods included in intercalibration (prime exercise, finished July 2011)				
69	Austria	Austrian Index Macrophytes for Rivers	AIM	1
125	Belgium (Flanders)	Flemish macrophyte assessment system[2]	MAFWAT	1
15	Belgium (Wallonia)	Biological Macrophyte Index for Rivers	IBMR-WL	2
154	France	Biological Macrophyte Index for Rivers	IBMR-FR	2
218	Germany	German assessment system for Macrophytes & Phytobenthos according to the EU WFD[3]	PHYLIB	1
244	Italy	Biological Macrophyte Index for Rivers	IBMR-IT	1
220	Poland	Macrophyte Index for Rivers	MIR	1
19	United Kingdom	Ecological Classification of Rivers using Macrophytes	LEAFPACS	1
Methods intercalibrated in a supplementary exercise (December 2011) – see Annex 3				
357	Denmark	Danish Stream Plant Index	DSPI	2
154	France	Biological Macrophyte Index for Rivers[4]	IBMR-FR	1
358	Ireland	Mean Trophic Ranking	MTR-IE	2
219	Luxembourg	Biological Macrophyte Index for Rivers	IBMR-LU	2
Methods under development and not included in intercalibration				
156	Netherlands	WFD-metrics for natural water types	KRW-maatlatten	3

Make sure that the **national method descriptions** meet the level of detail required to fill in the table 1 at the end of this document !

[1] http://www.wiser.eu/programme-and-results/data-and-guidelines/method-database/detail.php?id= *ID*

[2] Worst-case of the two sensitivity metrics (1) Type-specific index for water vegetation (TSw) and (2) Perturbation index (organic pollution, eutrophication) for water vegetation (Vw)

[3] only component "macrophytes" (component "diatoms" already intercalibrated in first IC round)

[4] including new classification (reference and boundary setting)

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).	Yes – all methods classify ecological status by one of five classes.
2. High, good and moderate ecological status are set in line with the WFD's **normative definitions** (**Boundary setting procedure**)	Yes – all methods set the high, good and moderate ecological status in line with the WFD's normative definitions (see *Annex 1*).
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.	"Macrophytes" is one of two components of the BQE "Macrophytes and Phytobenthos". Within the macrophyte methods all relevant parameters of the BQE (i.e. composition and abundance) are covered. See also *Annex 2* for further explanations.
4. Assessment is adapted to **intercalibration common types** that are defined in line with the typological requirements of the WFD Annex II and approved by WG ECOSTAT	Yes – all methods adapt their assessment to intercalibration common types.
5. The water body is assessed against **type-specific near-natural reference conditions**	Yes – all methods assess the water body against type-specific near-natural reference conditions.
6. Assessment results are expressed as **EQRs**	Yes – all methods express the assessment results as EQRs.
7. Sampling procedure allows for **representative** information about water body quality/ ecological status **in space and time**	Yes – the sampling procedures allow for representative information about the ecological status.
8. All data relevant for assessing the biological **parameters** specified in the WFD's normative definitions are covered by the **sampling procedure**	Yes – all relevant data are covered by the sampling procedures.
9. Selected taxonomic level achieves adequate **confidence and precision** in classification	Yes – the selected taxonomic level achieves adequate confidence and precision in classification.

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

Conclusion
All national methods using river macrophytes are WFD compliant.

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

Describe common intercalibration water body types and list the MS sharing each type

Common IC type	Type characteristics	MS sharing IC common type
Sandy lowland brooks (R-C1)	Abbreviation: R-C1 Catchment area: 10 - 100 km^2 Altitude: <200 m Geology: siliceous Channel substrate: Sand Alkalinity: >1 meq/l	Belgium (Flanders), Denmark, Estonia, Germany, Italy, Latvia, Lithuania, Netherlands, Poland, United Kingdom
Siliceous mountain brooks (R-C3)	Abbreviation: R-C3 Catchment area: 10 - 100 km^2 Altitude: 200 - 800 m Geology: siliceous Channel substrate: Boulders, cobble and gravel Alkalinity: <0.4 meq/l	Austria, Belgium (Wallonia), Czech Republic, France, Germany, Luxembourg, Poland, Spain, United Kingdom
Medium-sized lowland streams (R-C4)	Abbreviation: R-C4 Catchment area: 100 - 1,000 km^2 Altitude: <200 m Geology: mixed Channel substrate: Gravel and sand Alkalinity: >2 meq/l	Belgium (Flanders), Czech Republic, Denmark, Estonia, France, Germany, Ireland, Italy, Latvia, Lithuania, Luxembourg, Netherlands, Poland, Spain, United Kingdom

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
Austrian Index Macrophytes for Rivers (AT)	R-C3
Flemish macrophyte assessment system (BE-FL)	R-C1, R-C4
Biological Macrophytes Index for Rivers (BE-WL)	R-C3
Biological Macrophytes Index for Rivers (FR)	R-C3, R-C4
German assessment system for Macrophytes & Phytobenthos according to the EU WFD (DE)	R-C1, R-C3, R-C4
Danish Stream Plant Index (DK)	R-C1, R-C4
Mean Trophic Ranking (IE)	R-C4
Biological Macrophytes Index for Rivers (IT)	R-C1, R-C4
Biological Macrophytes Index for Rivers (LU)	R-C3, R-C4
WFD-metrics for natural water types (NL)	R-C1, R-C4
Macrophyte Index for Rivers (PL)	R-C1, R-C3, R-C4
Ecological Classification of Rivers using Macrophytes (UK)	R-C1, R-C3, R-C4
Conclusion Intercalibration is feasible in terms of typology	

4.2. Pressures

Describe the pressures addressed by the MS assessment methods

Table 4.1: Pressure-impact-relationships of national methods and selected pressure variables

Country	Pressure	Indicator tested	Sample size	r^2	p
Finalized methods included in intercalibration					
Austria	Eutrophication	Orthophosphate	55	0.42	<0.001
		Total Phosphorus	55	0.46	<0.001
		Nitrate	55	0.33	<0.001
		Ammonium	55	0.28	<0.001
Belgium (Flanders)	Eutrophication	Orthophosphate	260	0.17	<0.001
		Total Phosphorus	232	0.21	<0.001
		Nitrate	260	0.17	<0.001
		Ammonium	260	0.10	<0.001
	Others	BOD_5	232	0.08	<0.001
		Suspended solids	252	0.11	<0.001
Belgium (Wallonia)	Eutrophication	Orthophosphate	105	0.51	<0.001
France and Luxembourg	Catchment land use (R-C3)	% artificial	192	0.19	<0.001
		% intensive agriculture	192	0.28	<0.001
		% low agriculture	192	0.38	<0.001
		% (semi-)natural	192	0.48	<0.001
	Eutrophication (R-C3)	Ammonium	192	0.11	<0.001
		Nitrate	192	0.30	<0.001
		Total phosphorus	192	0.08	<0.001
		Orthophosphate	192	0.06	<0.001
	Catchment land use (R-C4)	% artificial	142	0.12	<0.001
	Eutrophication (R-C4)	Ammonium	142	0.20	<0.001
		Nitrite	142	0.22	<0.001
		Total phosphorus	142	0.25	<0.001
		Orthophosphate	142	0.21	<0.001
Germany	Eutrophication	Orthophosphate	19	0.18	<0.05
		Total Phosphorus	19	0.05	<0.05
Italy	Eutrophication	Orthophosphate	66	0.45	<0.001
		Nitrate	66	0.44	<0.001
	Others	Land use	66	0.37	<0.001
		Shading	66	0.47	<0.001
Poland	Eutrophication	Orthophosphate	153	0.27	0.0067
United Kingdom	Eutrophication	Orthophosphate	2040	0.48	<0.001
		Total Oxidized Nitrogen	2270	0.62	<0.001
Ireland	Eutrophication	Orthophosphate	279	0.04	<0.001
		Total Phosphorus	279	0.02	0.019
		Nitrate	279	0.11	<0.001
		Ammonium	279	0.07	<0.001
	Others	BOD_5	279	0.10	<0.001
		Irish Index for Macroinvertebrates	279	0.11	<0.001
Denmark	Eutrophication	Orthophosphate	400	NA[5]	<0.001
		WA Ellenberg N	1213		<0.001

[5] not applicable: discrete classes of national method tested against pressure indicators using ANOVA

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)

Conceptual differences between methods

There are differences in the concepts of the national methods that have effect on the feasibility of the intercalibration. For that reason some methods had to be (partly) excluded from the intercalibration process in this round.

Most assessment methods are designed according to the concept of positive, negative and indifferent indicator species to specific pressures or assign a continuous or ordinal scaled value to different taxa. The indication value has been derived from data that showed a correlation between the presence and abundance of the species and the impact value of the pressure. The calculation of EQR differs between the methods, but it is some way of weighted averaging of indicator species abundance. Methods of Austria, Belgium (Wallonia), France, Italy, Luxembourg, Poland and United Kingdom follow this concept completely, and most parts of the German method are of this concept type. The Flemish methods is partly of this concept type.

Other assessment methods are designed to the concept that any loss of species and structures from a naturally complex ecosystem is indicating loss of quality, independent of the pressure involved. The reference is the expected natural species composition and any derivation from that is considered as loss of quality. The calculation of EQR is based on the quotient of observed species over expected species, weighted for species abundance and vulnerability value of the species. The Dutch method on species composition is solely of this type, the Flemish method is partly of this concept type. The German method has additional metrics for species richness and evenness that are comparable to this concept. To fulfill the requirements of the assessment the Dutch method needs multiple samples of waterbodies; single site assessment results in underestimating the quality.

The national methods of Flanders, Germany, United Kingdom and the Netherlands add information on total abundance of specific growth forms or diversity of growth forms (Flanders) as an extra metric into the calculation. Since the data needed for most of these metrics are sometimes lacking not all these metrics could be included in the intercalibration. The Danish method (currently under development) is also conceptually different. It identifies assemblages of species in a multivariate multidimensional space; quality and degradation are defined as a gradient in this space and class boundaries are defined as contour shapes, perpendicular to the gradient. The calculation of EQR is based on fitting samples into this space and finding the distances between the nearest boundaries.

The main reason for the concept differences lies in the difference of pressure types that are considered as the most important to assess. Most of the national assessment methods are primarily designed to assess impact of eutrophication. The indicator species are related to trophic status and therefore the metric responds to nutrient enrichment. Though several pressures are additionally addressed by all national methods, it is assumed that they reflect their impact on the trophic status indirectly and are therefore accurately assessed by the method as well.

Countries with relatively high trophic status in (assumed) reference conditions, i.e. typically mesotrophic in slowly running lowland rivers, (e.g. Denmark, Flanders, Northern Germany, Netherlands) are dealing with loss of contemporary quality that is more related to hydromorphological degradation than eutrophication and is less readily assessable with metrics based on trophic status. Defining indicator species and their indicator values for hydromorphological degradation is much more complex and less obvious than for eutrophication.

Evaluation of "outlying" methods

Despite the differences in concepts, the species composition metrics in the national methods can be intercalibrated for most of the countries anyway. This is because the results of the metrics show that they are similar enough to compare. Correlation between the EQR of the national methods and the pseudo-common metric (see section on *IC common metrics*) is high enough for most of the methods and intercalibration is therefore regarded as feasible. The Dutch method is outlying but national improvements are to be expected.

The Dutch method did not correlate significantly and intercalibration was therefore regarded as not feasible in this phase. Since national evaluation of the Dutch metric shows several flaws within the method (2011, not yet published) and suggestions for improvements show that higher correlation with the Common Metric mICM is possible, it may be expected that the improved Dutch metric will meet the correlation criteria despite the different concept.

The metrics for total abundance of specific growth forms are excluded from all methods in this process. If it can be proven by the countries that these metrics will not result in systematically higher or lower EQR than of the species compositional metric, this omission will not have effect on the class-boundary setting process. In case the abundance metrics have systematically higher or lower EQR the national class-boundaries should be adjusted in accordance to these differences additional to the (possible) adjustments from the basic intercalibration process. This type of evaluation is excluded from the intercalibration process.

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

IC type	Country	Full dataset		Benchmark dataset	
		Number of sites	Number of surveys	Number of sites	Number of surveys
R-C1	Belgium (Flanders)	78	78	11	11
	Denmark	21	21	6	6
	Germany	98	98	4	4
	Italy	20	30	13	15
	Netherlands	14	14	4	4
	Poland	25	25	14	14
	United Kingdom	14	14	9	9
	Sum	***270***	***280***	***61***	***63***
R-C3	Austria	19	19	6	6
	Belgium (Wallonia)	33	35	11	11
	Czech Republic	12	13	7	7
	France	45	50	24	27
	Germany	79	81	4	5
	Luxembourg	11	11	0[6]	0[5]
	Poland	10	10	6	6
	United Kingdom	21	22	10	11
	Sum	***230***	***241***	***68***	***73***
R-C4	Germany	93	95	8	9
	Belgium (Flanders)	9	9	0[7]	0[6]
	Belgium (Wallonia)	21	21	4	4
	Denmark	14	14	9	9
	France	49	55	8	8
	Ireland	20	33	9	9
	Italy	4	4	3	3
	Latvia	36	36	20	20
	Lithuania	9	9	2	2
	Netherlands	8	8	0[6]	0[6]
	Poland	43	44	17	17
	United Kingdom	173	179	27	29
	Sum	***479***	***507***	***107***	***110***

[6] benchmarking based on French benchmark sites
[7] no benchmarking applied

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

Data acceptance criteria and data harmonization

All surveys showed at least three scoring taxa according to the Austrian method (only R-C3), an abundance share of > 75% of scoring taxa according to the German method and a minimum total (cubed) abundance of >15. Surveys where the variability among national EQRs was very high due to the variable performance of individual methods were removed.

Algal adjustment

A significant difference between national methods consisted in the recording and assessment of algae done by France, Italy, Luxembourg, Poland, United Kingdom and Wallonia. Other countries neither recorded algae within their macrophyte surveys nor included algal occurrence into their assessment. Thus, the EQRs of methods using algae were systematically biased when applied to data excluding algal records. For these cases we used a correction factor that adjusted the national EQR value for the lack of algal information.

Datasets combining surveys of various stream types at rivers in the above mentioned countries were used to establish relationships between the original national macrophyte metrics of these countries and metric versions that excluded all algal indicator taxa (except charophytes). Based on the linear regression equations (Table 5.1) we then inferred the original metric scores for surveys of countries not recording algae. These adjusted metric scores formed the basis of the EQRs for these particular surveys.

Table 5.1: Details of the regression analyses used for algal adjustment of selected national macrophyte metrics

Country	Metric name	Number of surveys	R-square	Regression equation
Belgium (Wallonia), France, Italy, Luxembourg	IBMR	757	0.85	y = 0.8425x + 1.4674
Poland	MIR	78	0.96	y = 0.9474x + 1.5638
United Kingdom	RMNI	2355	0.95	y = 0.9156x + 0.7145
United Kingdom	N-FG	2355	0.93	y = 1.0775x + 0.5879
United Kingdom	N-TAXA	2355	0.87	y = 1.1603x + 0.8294

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

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 Poland and the Netherlands, 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.

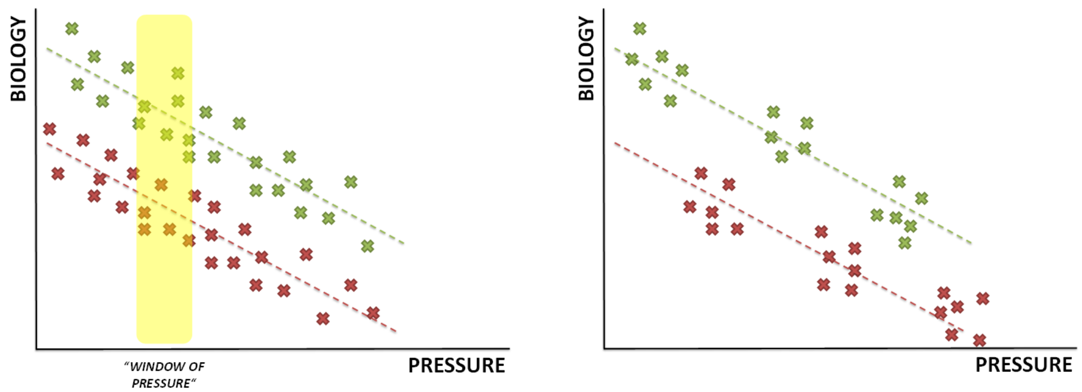


Figure 6.1: The two principal approaches of benchmarking in intercalibration
Left: "Benchmark sites" approach: Sites at similar (low) level of pressure ("window of pressure") are available from all countries (red and green).
Right: "Continuous benchmarking" approach: Countries' data (red and green) cover different levels of pressure.

The principle aim of benchmarking 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, the pressure effects on the assessment scores (i.e. national EQRs) have to be controlled to disclose any remaining discrepancies.

At reference or benchmark sites differences between countries due to pressure effects are minimised, because these sites do not cover a gradient of pressure. In continuous benchmarking the effects of pressure that differ between countries are removed by statistical modelling. Generalised linear models or comparable approaches (e.g. linear mixed effects models) can be applied for this purpose. The pressure variable(s) form the covariates in the model. The country of data origin is specified as a random factor. The term ‘country’ represents the surrogate for all systematic discrepancies among national EQRs that are unrelated to pressure. The model yields the intercept deviates of individual assessment methods which are then used in the benchmark standardisation of national EQRs (Figure 6.2).

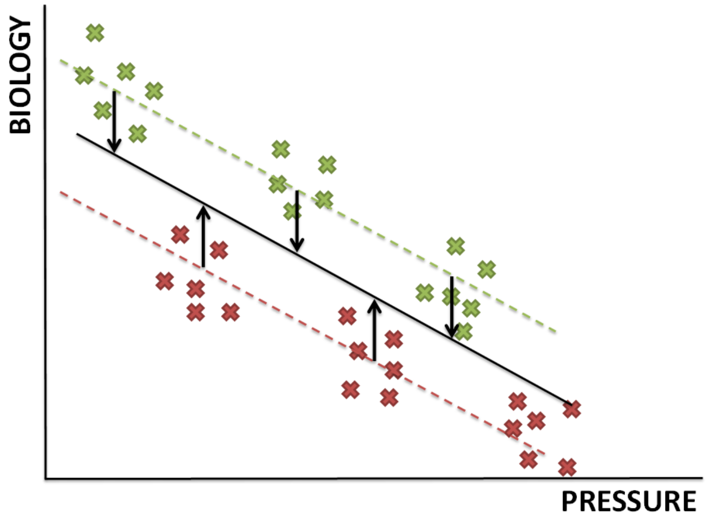


Figure 6.2: Adjustment of national EQRs (red and green) based on offset results from the statistical model

Benchmark standardisation

Since benchmarking aims at defining abiotic baselines that standardise the different national EQRs across their full range, their response pattern to human pressure is important. Do the national assessments only differ at (relatively) undisturbed conditions but converge at the more disturbed end of the gradient, or do these differences persist throughout the gradient? This distinction determines how to calculate the benchmark standardisation (Figure 6.3). Our standardisation was obtained either by directly subtracting the offsets yielded by the statistical model from the observed EQR values or by dividing the observed EQR values by a divisor calculated as 1-offset value[8].

Relying on pressure-impact-relationships of the national methods we referred to the national EQRs only and selected the appropriate calculation by testing if the average value of all national EQRs per survey in the full dataset was significantly correlated with its standard deviation. In case of a significant positive relationship, i.e. national EQRs converge towards the bad end of the quality gradient, division was used. An insignificant relationship, i.e. constant distances between EQRs across the full gradient, required subtraction (Table 6.1; Table 6.2).

[8] Although division was always used for benchmark standardisation in earlier exercises it was important to introduce this distinction here as the calculation affects the relative position of the class boundaries to be compared in intercalibration. In division, for instance, distances increase if the actual benchmark value is smaller than the national reference value; in subtraction all distances stay the same.

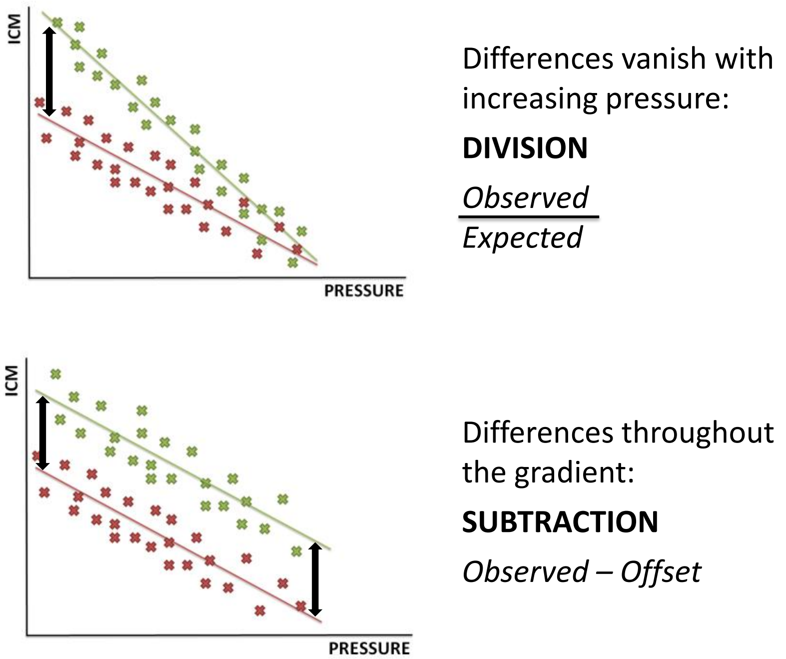


Figure 6.3: Types of pressure-impact-relationships and options of calculating the benchmark standardisation

Upper diagram: Standardisation is done by division if differences in the biological assessment results (ICM) between national methods (green and red crosses) vanish with increasing pressure.

Lower diagram: Standardisation is done by subtraction if differences remain throughout the entire pressure gradient.

Table 6.1: Results of correlation analysis of average national EQRs and their standard deviation used to identify the option for benchmark standardisation (CorrCoef: Pearson's correlation coefficient)

IC type	CorrCoef	P-level	Benchmark standardisation
R-C1	0.0101	0.871	Subtraction
R-C3	0.3854	<0.0001	Division
R-C4	0.4707	<0.0001	Division

Table 6.2: Country-specific offset values used in benchmark standardisation

Constant is based on the global mean of all country effects minus the specific country effect.

Divisor = the correction used to remove the country effect where the divisor = 1+(country effect mean - global mean of all country effects)

IC type	Data origin	National method							
		Austria	Belgium (Flanders)	Belgium (Wallonia)	France	Germany	Italy	Poland	United Kingdom
Sandy lowland brooks (R-C1)	***Benchmark standardisation***	***Constant***							
	Belgium (Flanders)		0.017			0.092	0.022	-0.024	-0.012
	Denmark		0.005			-0.187	-0.035	0.309	0.164
	Germany		-0.088			-0.017	-0.001	0.046	0.081
	Italy		-0.003			0.077	0.065	-0.041	-0.013
	Netherlands		0.004			0.106	-0.035	-0.050	-0.145
	Poland		-0.018			0.008	-0.020	0.070	0.051
	United Kingdom		0.083			-0.079	0.005	-0.311	-0.126
Siliceous mountain brooks (R-C3)	***Benchmark standardisation***	***Divisor***							
	Austria	0.971		0.978	0.977	0.951		1.013	0.999
	Belgium (Wallonia)	1.031		1.015	1.015	1.087		0.993	0.997
	Czech Republic	0.994		1.004	1.005	0.982		0.983	1.006
	France	0.987		0.856	0.851	0.937		0.923	0.909
	Germany	1.042		1.005	1.005	1.039		0.996	0.989
	Luxembourg	0.987		0.856	0.851	0.937		0.923	0.909
	Poland	0.956		1.006	1.007	0.960		1.027	0.933
	United Kingdom	1.020		1.136	1.140	1.044		1.065	1.167
Medium-sized lowland streams (R-C4)	***Benchmark standardisation***	***Divisor***							
	Belgium (Wallonia)				1.124	0.879	1.106	1.171	1.245
	Denmark				1.106	0.917	1.090	0.923	0.870
	France				0.925	0.954	0.936	0.934	0.923
	Germany				0.823	1.115	0.849	1.011	0.968
	Ireland				1.034	1.021	1.029	1.134	1.073
	Italy				1.010	1.028	1.009	0.995	0.997
	Latvia				1.035	0.907	1.030	1.017	0.993
	Lithuania				0.943	1.175	0.951	1.016	1.066
	Poland				1.068	0.997	1.058	0.969	0.901
	United Kingdom (Great Britain)				1.016	1.067	1.013	1.045	1.132
	United Kingdom (Northern Ireland)				0.945	1.058	0.953	0.887	1.086

- 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 refers to data from sites featuring various levels of disturbance. The effects of pressure that differ between countries are removed by statistical modelling. See also section above.

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

See section on the *collection of IC dataset*.

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

Continuous benchmarking yields country-specific benchmarks that differ in degree of pressure and (thus) biological communities. The validation of the benchmark is carried out *post-hoc* investigating sites at harmonised high status (i.e. their pressure-levels and biological communities). See section on *biological communities representing the "borderline" conditions*.

- 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

Continuous benchmarking yields country-specific benchmarks that differ in degree of pressure and (thus) biological communities. The validation of the benchmark is carried out *post-hoc* investigating sites at harmonised high status (i.e. their pressure-levels and biological communities). See section on *biological communities representing the "borderline" conditions*.

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

See section on *biological communities representing the "borderline" conditions*

7. Design and application of the IC procedure

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

Which IC option did you use?

- IC Options 3 - Similar data acquisition, but different numerical evaluation (BQE sampling and data processing generally similar, so that all national assessment methods can reasonably be applied to the data of other countries) → supported by the use of common metric(s)

Explanation for the choice of the IC option:

IC Options 3: Similar data acquisition but different numerical evaluation (BQE sampling and data processing generally similar so that all national assessment methods can reasonably be applied to the data of other countries)

In case of IC Option 2, please explain the differences in data acquisition

not applicable

7.2. IC common metrics (When IC Options 2 or 3 are used)

Describe the IC Common metric:

We performed an IC Option 3 (direct comparison) using pseudo-common metrics (PCM). After adjusting and standardising the national EQRs (algal adjustment, benchmark standardisation) the values were normalised to a range of 0 to 1. Based on the survey data commonly assessed by each method we related the national EQRs to a PCM constructed by averaging the EQR values of the remaining methods using ordinary least squares **regression**. Thus, in an exercise involving five national assessment methods, for method A we related the national EQR of method A (the 'test' method) to the average EQRs of methods B, C, D and E (the 'assessors'); this was repeated for method B versus the average of methods A, C, D and E etc.

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).

IC type	Country	r	p
Sandy lowland brooks (R-C1)	Belgium (Flanders)	0.736	<0.0001
	Germany	0.613	<0.0001
	Italy	0.769	<0.0001
	Poland	0.662	<0.0001
	United Kingdom	0.636	<0.0001
Siliceous mountain brooks (R-C3)	Austria	0.916	<0.0001
	Belgium (Wallonia)	0.959	<0.0001
	France	0.959	<0.0001
	Germany	0.861	<0.0001
	Poland	0.920	<0.0001
	United Kingdom	0.917	<0.0001
Medium-sized lowland streams (R-C4)	France	0.816	<0.0001
	Germany	0.518	<0.0001
	Italy	0.811	<0.0001
	Poland	0.803	<0.0001
	United Kingdom	0.700	<0.0001

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

The Flemish method had to be excluded from the analysis for stream type R-C4 due to its low correlation with the PCM (r=0.436)[9].

[9] Flemish method variant of river type "rk – small rivers" that showed the best correlation of all relevant Flemish river types.

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 applicable

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

not applicable

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

not applicable

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).

Sandy lowland rivers (R-C1)

In R-C1 rivers the major changes in the abundance of different functional groups are consistent with an overall increase in fertility coupled with a reduction in stream energy that results in increasing abundance of larger, more productive species (Figure 8.1). Thus, small amphibious species (e.g. *Myosotis scorpioides*) are progressively replaced by large emergent species (e.g. *Glyceria aquatica*). Several groups that are well represented in the least impacted sites (e.g. callitrichids, batrachids, broad-leaved pondweeds, such as *Potamogeton natans*, and medium sized emergents, such as *Berula erecta*) show only a slight contraction in abundance as ecological quality declines, but due to the overall increase in plant cover, are proportionally less abundant in the most impacted sites. The most pronounced changes as quality declines are the marked expansion in abundance of fine-leaved pondweed taxa (especially *Potamogeton trichoides*) and lemnids, an increasing abundance of the vallisnerid *Sparganium emersum*, and the progressive loss of various bryophytes. Such changes indicate an increasingly fertile and low energy environment where hard substrates are obscured through siltation, artificial disturbances such as dredging or weed cutting are common and the natural physical habitat heterogeneity and dynamism is reduced.

There are no threshold changes associated with the harmonised upper class boundaries. Compared to reference condition there is a slight expansion of fine-leaved pondweeds and lemnids, and a minor reduction in bryophytes at the HG boundary. By the GM boundary such changes are more advanced and begin to extend to a reduction in the cover of small amphibious taxa and an expansion of vallisnerids. However, the major attributes of reference condition assemblages remain recognisable, whereas by the MP boundary such features have been completely lost.

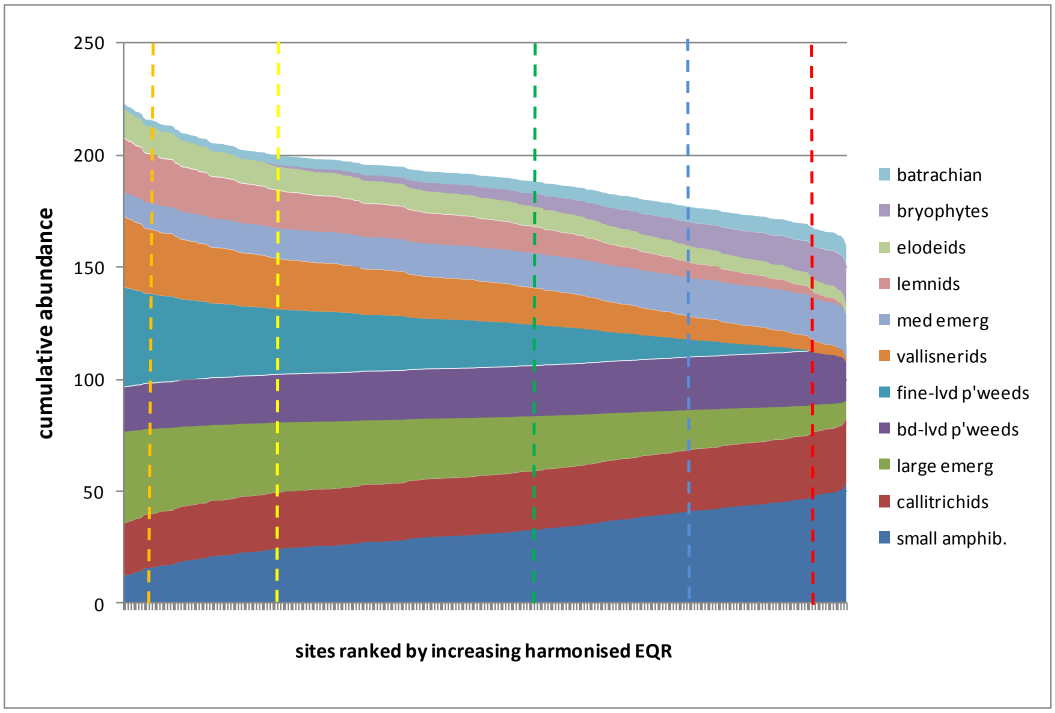


Figure 8.1: Graphical representation of modelled changes in major aquatic plant growth forms across harmonised EQR in R-C1 rivers

Lines from left to right indicate harmonised position of PB, MP, GM, HG boundaries and reference position respectively. Data is only shown for life forms that account for >1% of total recorded cover and life forms are stacked from the base in order of decreasing contribution to the total vegetation recorded in that river type.

Siliceous mountain brooks (R-C3)

The changes that occur in R-C3 rivers with declining ecological quality are consistent with enrichment from diffuse sources and increased fine sediment input. The 'core' vegetation of large pleurocarpous mosses (predominantly of the genera *Fontinalis*, *Hygrohypnum* and *Rhynchostegium*) is largely resistant to such impacts, but other groups, most notably the leafy liverworts, such as *Scapania* and *Chiloscyphus*, are much more sensitive (Figure 8.2). Acrocarpous mosses, such as *Racomitrium*, plus small red and blue green algae also decline, being replaced by larger green filamentous algae and batrachid species. Greater stability of channel margins and increased fine sediment content, possibly coupled with reduced tree cover, also leads to an expansion in the abundance of large emergent species such as *Phalaris arundinacea* or *Sparganium erectum*.

Again there are no clear threshold changes that can be associated with class boundaries. Batrachids and large emergent species are both very scarce at reference condition but are clearly established by the HG boundary. However, all the key elements of the reference condition flora are clearly still intact at this boundary. The changes observed at the HG boundary have become more advanced by the GM boundary and there has been a significant decline in the cover of leafy liverworts. However, groups including acrocarpous mosses, thallose liverworts and red and blue-green algae remain present, albeit in reduced quantities. Conversely by the MP boundary, leafy-liverworts, acrocarpous mosses and blue-green algae have disappeared.

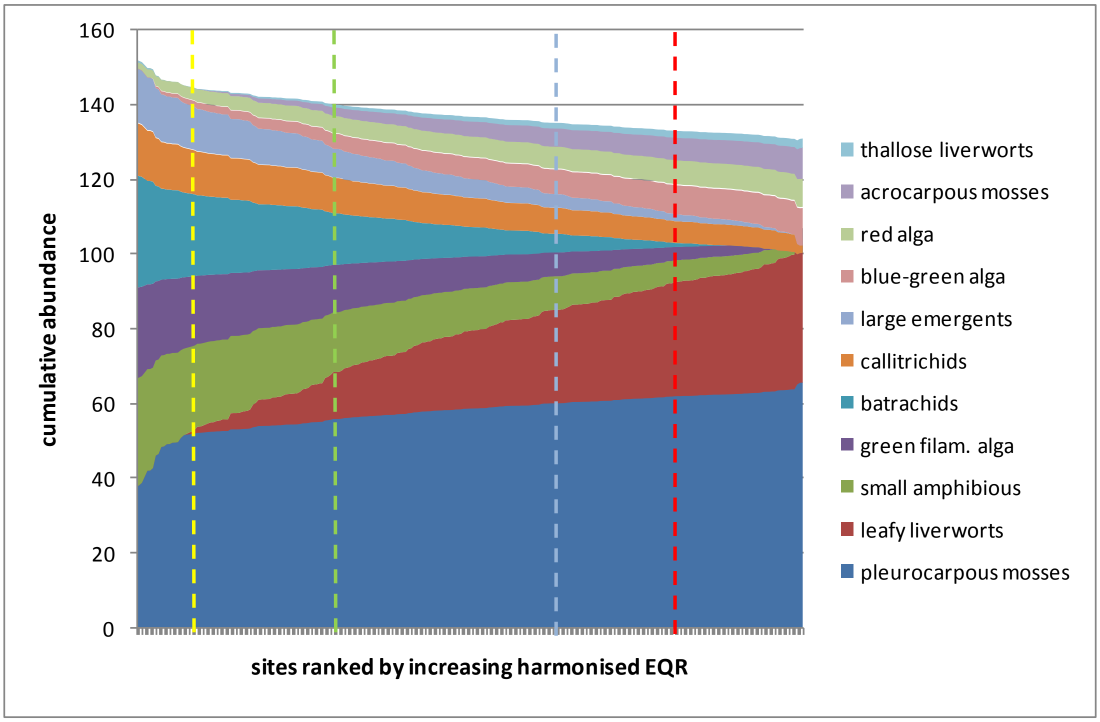


Figure 8.2: Graphical representation of modelled changes in major aquatic plant growth forms across harmonised EQR in R-C3 rivers

Medium-sized lowland streams (R-C4)

In R-C4 rivers the increase in cover and decrease in taxonomic richness as ecological quality declines is more pronounced than in the other two IC river types. This is clearly captured in Figure 8.3 which also highlights the reduction in evenness at a plant growth form level. The ecological signal is especially strong in R-C4 which is all the more remarkable since most Member States do not explicitly consider absolute abundance in their method and several specifically exclude algae.

The major pressure associated with decreasing quality is eutrophication from a variety of sources, but this is coupled with either a reduction in stream energy that supports a proliferation of fine-leaved pondweeds, or instability and flashiness, possibly associated with management or urbanisation, that promotes the expansion of opportunistic larger green filamentous algae. Conversely, bryophytes (mainly *Fontinalis antipyretica*, *Amblystegium riparium* or *Rhynchostegium riparioides*), red algae, such as *Hildenbrandia*, and to a lesser extent small amphibious taxa, show strong declines along a gradient of decreasing quality. Many growth forms (e.g. batrachids, callitrichids and broad-leaved pondweeds), show quite subtle reductions in abundance over the length of the quality gradient but the share of these groups in the most degraded sites is substantially lowered by the parallel expansion in cover of filamentous algae and fine-leaved pondweeds.

There is a continuum of change rather than abrupt thresholds that can be associated with class boundaries. The distinction between reference condition and the HG boundary is subtle and linked mainly to the onset of expansion of green filamentous algae. By the GM boundary the expansion of filamentous algae is more pronounced while fine-leaved pondweeds are also becoming more prolific but the quintessential features of reference condition sites are still present. A reduction in cover of bryophytes and red algae is a precursor to the changes at the MP boundary by which point both these groups are rare or absent.

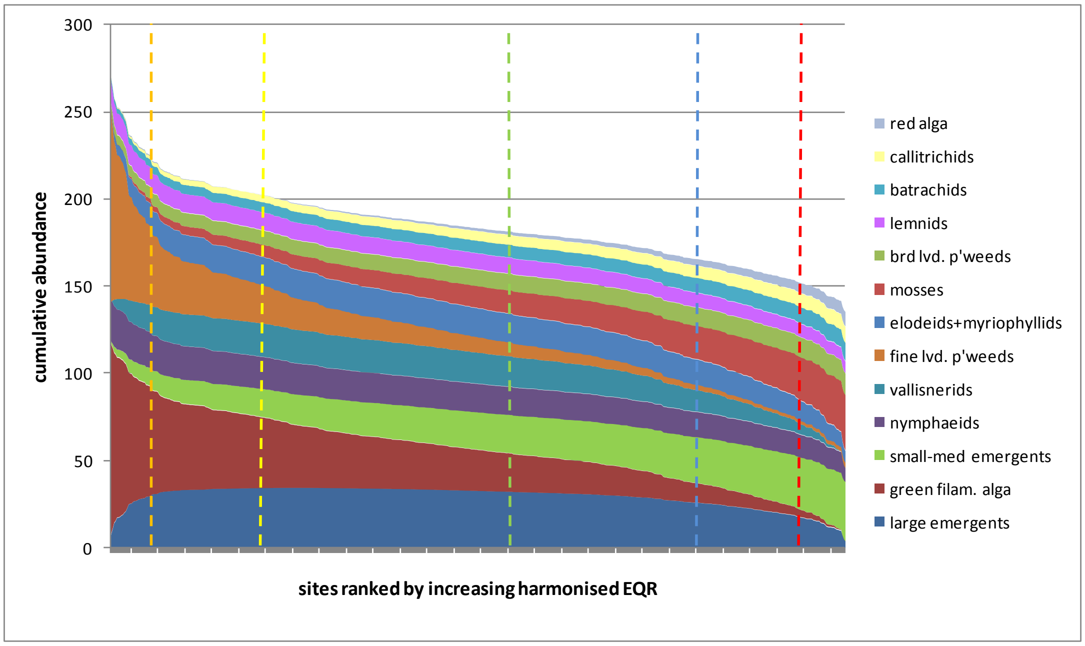


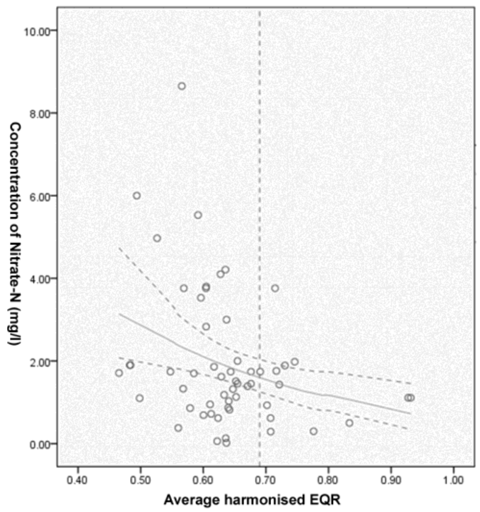
Figure 8.3: Graphical representation of modelled changes in major aquatic plant growth forms across harmonised EQR in R-C4 rivers

Post-hoc validation of the alternative benchmark

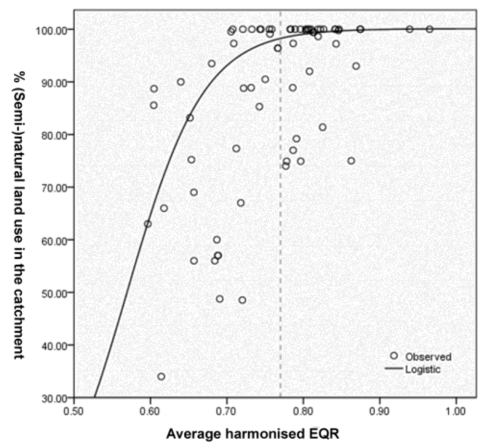
Sites at harmonised high status on average feature a low level of anthropogenic pressure as revealed by the parameters of water chemistry (Table 8.1). The catchment of all river types is dominated by (semi-) natural land use. While almost the complete catchment is natural for the mountain brooks the lowland rivers feature considerable shares of intensive agriculture and artificial land use. However, the concentrations of relevant nutrients such as phosphate and nitrate at the sampling site are low. Figure 8.4 describes various pressure models significantly relating the average harmonized EQR to selected environmental variables.

Table 8.1: Median values of environmental parameters at sites in harmonised high status

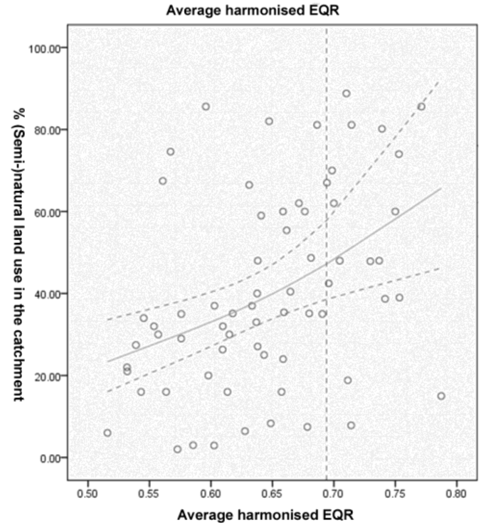
IC type		R-C1	R-C3	R-C4
Number of high status sites		47	62	67
Catchment land use (%)	Artificial	1.3	0.0	1.4
	Intensive agriculture	30.0	0.0	35.0
	Extensive agriculture	9.8	0.0	10
	(Semi-)natural	41.0	98.9	42.5
Water chemistry	Conductivity (µS/l)	335	76	390
	NH_4 (mg N/l)	0.070	0.026	0.090
	NO_3 (mg N/l)	1.45	0.85	2.71
	PO_4 (mg P/l)	0.060	0.020	0.070
	BOD_5 (mg O_2/l)	1.8	1.5	1.9



Generalized Linear Model of the average harmonized EQR of R-C1 against Nitrate-N (p=0.006)



Logistic regression model of the average harmonized EQR of R-C3 against (semi-)natural land use in the catchment (p<0.001)



Generalized Linear Model of the average harmonized EQR of R-C4 against (semi-)natural land use in the catchment (p=0.002)

Figure 8.4: Pressure models relating the average harmonized EQR to selected environmental variables (dashed vertical lines = harmonized reference position)

8.3. Boundary comparison and harmonization

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

Translating national boundaries

We took the model formula for each regression and determined the PCM value that equated to the upper class boundaries for each national method. Therefore, for method A, if y = mx +c where y = the PCM value, m = the regression slope, x = the EQR value of method A and c = the regression intercept, we derived the value on the PCM scale for values of x corresponding to the High-Good and Good-Moderate class boundaries. This was repeated for each national method's regression. All regression characteristics were checked, i.e. the relationship had to be significant ($p \leq 0.05$), the correlation between each method and the PCM should be $r \geq 0.5$ (Pearson's correlation coefficient), the slope of the regression was tested to be significantly different from 0.

Defining a harmonisation guideline

By fitting national class boundaries to each of the national EQR versus PCM relationships we established the predicted values on a PCM scale for each upper class boundary. This yielded a mid-point represented by the global average of the predicted values. This mid-point was 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 principal applied irrespective of the error associated with the projection of each national class boundary onto a common scale.

Evaluating the level of boundary bias

We defined the difference between the harmonisation guideline and the national boundary values on the PCM 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 PCM units. The boundary bias was then calculated by subtracting the boundary position on the PCM scale from the harmonisation guideline. Finally, we related this difference to the width of the respective national class intersected by the harmonisation guideline (Figure 8.4, Table 8.1).

Harmonising national class boundaries

National boundaries exceeding a bias of 0.25 class equivalents were adjusted to fall inside this permitted level of deviation. We raised or lowered the original boundary EQRs and observed the effects on the boundary bias. Thus, each adjusted boundary underwent the full analytical procedure as described above (benchmark standardisation, translation into PCM, evaluation of boundary bias) but was not allowed to change the position of the harmonisation guideline.

Analysis of class agreement

To check the class agreement after the boundary harmonization we calculated the unsigned (i.e. absolute) differences in classification between each method and every other method for all commonly assessed sites and defined the average absolute class difference (Table 8.2).

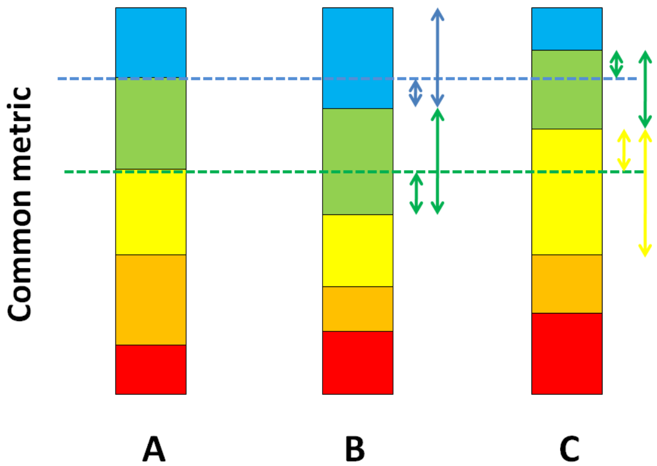


Figure 8.4: National classification schemes (A, B, C) intersected by the average boundary positions derived in intercalibration ("harmonisation guideline" – broken blue and green horizontal lines). Small arrows depict the distance of relevant national boundary to the harmonisation guideline. Large arrows define relevant national class width. The relation of small to large arrow-lengths specifies the boundary bias in class equivalents.

Table: 8.1: National class boundaries and boundary bias (adjusted boundaries, if bias >|0.25|)
Proposed adjustments: ↑ boundary to be raised, ↓ boundary can be lowered

IC type	Country		Original		Adjusted			
			HG	GM	HG		GM	
Sandy lowland brooks (R-C1)	Belgium (Flanders)	Boundary	0.800	0.600	0.810	↑	0.615	↑
		Bias	-0.26	-0.31	-0.22		-0.24	
	Germany	Boundary	0.745	0.495				
		Bias	0.12	0.24				
	Italy	Boundary	0.900	0.800	0.875	↓	0.770	↓
		Bias	0.51	0.37	0.25		0.21	
	Poland[10]	Boundary	0.900	0.650				
		Bias	-0.24	-0.23				
	United Kingdom	Boundary	0.800	0.600				
		Bias	-0.06	-0.18				
Siliceous mountain brooks (R-C3)	Austria	Boundary	0.875	0.625				
		Bias	0.09	-0.14				
	Belgium (Wallonia)	Boundary	0.925	0.607				
		Bias	-0.01	-0.23				
	France	Boundary	0.856	0.722			0.645	↓
		Bias	0.39	0.44			0.25	
	Germany	Boundary	0.745	0.545	0.800	↑		
		Bias	-0.41	-0.32	-0.24			
	Poland	Boundary	0.910	0.684				
		Bias	0.19	0.11				
	United Kingdom	Boundary	0.800	0.600				
		Bias	-0.14	0.02				
Medium-sized lowland streams (R-C4)	France	Boundary	0.890	0.765	0.855	↓		
		Bias	0.46	0.18	0.18			
	Germany	Boundary	0.575	0.395				
		Bias	0.03	0.07				
	Italy	Boundary	0.900	0.800	0.945	↑		
		Bias	-0.56	-0.36	-0.20			
	Poland	Boundary	0.900	0.650				
		Bias	0.05	-0.21				
	United Kingdom	Boundary	0.800	0.600				
		Bias	-0.06	0.04				

[10] based on adjusted Polish reference values for R-C1 (MIR=56)

Table 8.2: Average absolute class difference (AACD) of national methods after boundary harmonisation, including the number of pairwise comparisons (count)

IC type	Country	Count	AACD
Sandy lowland brooks (R-C1)	Belgium (Flanders)	1120	0.53
	Germany		0.52
	Italy		0.52
	Poland		0.56
	United Kingdom		0.55
Siliceous mountain brooks (R-C3)	Austria	1205	0.34
	Belgium (Wallonia)		0.31
	France		0.35
	Germany		0.38
	Poland		0.38
	United Kingdom		0.41
Medium-sized lowland streams (R-C4)	France	2028	0.38
	Germany		0.51
	Italy		0.37
	Poland		0.50
	United Kingdom		0.42

– Do all national methods comply with these criteria ? (N)
– If not. describe the adjustment process:

For certain types some national methods do not comply with the comparability criteria. Boundary bias is exceeded by the methods of

- Type R-C1 – Belgium (Flanders) – Boundaries HG and GM too relaxed [#]
- Type R-C1 – Italy – Boundaries HG and GM too stringent [Π]
- Type R-C1 – Poland – National reference value too relaxed [#]
- Type R-C3 – France – Boundaries HG and GM too stringent [Π]
- Type R-C3 – Germany – Boundaries HG and GM too relaxed [#11]
- Type R-C4 – France – Boundary HG too stringent [Π]
- Type R-C4 – Italy – Boundaries HG and GM too relaxed [#12]

The required boundary adjustments are specified in Table 8.1. The average absolute class difference after boundary harmonisation meets the comparability criteria for all national methods.

[#] Member States are requested to implement the adjusted boundaries to allow for completing the intercalibration exercise. Poland is asked to raise its national reference value of stream type R-C1 to a MIR value of 56.

[Π] Member States are not obliged to lower the boundaries that have been identified as being too stringent.

[11] Germany agreed to raise HG boundary to comply with the comparability criteria. This will also harmonise the position of the GM boundary (see Table 8.1).
[12] Italy agreed to raise the R-C4 reference value to comply with the comparability criteria as soon as the National Law will be revised.

Annex 1: National boundary setting

Table A1: CB GIG Member State national EQR boundary values (H/G and G/M) and rationales for defining Member State status class boundaries

	High / Good Boundary	Good / Moderate Boundary
AT	0.875 (R-C3)	0.625 (R-C3)
	At high status reference species are clearly dominant with indifferent species co-occurring. No or negligible occurrence of disturbance-indicating species.	At good status type-specific reference species and tolerant species are still dominant, pressure indicators are rare. At moderate status degradation indicators dominate over reference species.
	The class boundaries for each metric were defined according to the normative definitions and interpretations of the WFD as given in the REFCOND Guidance. Reference species have been defined following species-specific nutrient requirements and bio-geographical preferences.	
Be-F	0.800	0.600
	Equidistant division of the EQR gradient	Equidistant division of the EQR gradient
	EQR gradient is assumed to represent a continuous trend 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.	
Be-W	0.925 (R-C3)	0.607 (R-C3)
	Derived from metric variability at near-natural reference sites	
	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.	
DE	0.745 (R-C1) 0.745 (R-C3) 0.575 (R-C4)	0.495 (R-C1) 0.545 (R-C3) 0.395 (R-C4)
	High-good boundary derived from metric variability at near-natural reference sites Type-specific reference species and tolerant species are still dominant, pressure indicators are rare = slightly deviation from high status (normative definitions)	In good status reference and tolerant species are abundant, degradation indicators occur. In moderate status degradation indicators dominate over reference species.
	The boundaries were set at the zones of distinct changes of the biocoenosis (macrophytes and phytobenthos).	

<table>
<tr><th></th><th>High / Good Boundary</th><th>Good / Moderate Boundary</th></tr>
<tr><td rowspan="2">FR</td><td>0.856 (R-C3)
0.890 (R-C4)</td><td>0.722 (R-C3)
0.765 (R-C4)</td></tr>
<tr><td colspan="2">Boundaries were defined using a reference dataset collected on reference sites, meeting the "reference condition" criteria given form the EU Guidance. For each group of river national types, the base for defining boundaries is the 25th percentile of reference values for H/G boundary. The G/M boundary is derived from equidistant division of the rest. These values were adjusted from expert judgment to balance the lack of data (few reference data or no data for several river types).</td></tr>
<tr><td rowspan="2">IT</td><td>0.900 (R-C1)
0.900 (R-C4)</td><td>0.800 (R-C1)
0.800 (R-C4)</td></tr>
<tr><td colspan="2">Boundary setting has been identified using data from sites belonging to different quality level assessed using pressure data (land use, hydrological, morphological and chemical features), comparison with others BQE, expert judgement.</td></tr>
<tr><td rowspan="4">NL</td><td>0.800</td><td>0.600</td></tr>
<tr><td>Equidistant division of the EQR gradient</td><td>Equidistant division of the EQR gradient</td></tr>
<tr><td colspan="2">The reference score for the sum of the scores of the species is derived from frequency data in the national vegetation database on well developed plant communities in the Netherlands which is considered a good estimate for the probability of finding the species in a fixed amount of samples. The fraction of species at H/G and G/M are estimated with expert judgment and adjustment may be needed because of too low number of reference sites. Final adjustment of the reference scores are based on intercalibration results.</td></tr>
<tr><td colspan="2">Good status of small rivers is characterized by a variety of species, growing at several habitats. Pressure tolerant species are present but only in low abundances; total cover of vegetation is moderate and type-specific.</td></tr>
<tr><td rowspan="3">PL</td><td>0.900 (R-C1)
0.910 (R-C3)
0.900 (R-C4)</td><td>0.650 (R-C1)
0.684 (R-C3)
0.650 (R-C4)</td></tr>
<tr><td>Expert judgment based on relationship with nutrient content in the water.</td><td>Expert judgment based on relationship with nutrient content in the water.</td></tr>
<tr><td colspan="2">Class boundaries are generally equidistant in the EQR gradient, although slight H/G adjustment based on expert judgment was used.</td></tr>
<tr><td rowspan="3">UK</td><td>0.800</td><td>0.600</td></tr>
<tr><td>Derived from metric variability at near-natural reference sites using paired metrics that respond in different ways to the influence of the pressure</td><td></td></tr>
<tr><td colspan="2">The relative positions of High-Good and Poor-Bad boundaries are effectively symmetrical with sensitive species overwhelmingly dominant at one and tolerant species overwhelmingly dominant at the other. Using the same standard error from logistic regressions, a ratio of sensitive:tolerant species of 85:15 is used as the High-Good boundary, since this represents the upper error when tolerant species are predicted to be absent. These ratios are reversed at the Poor-Bad boundary with 15% sensitive species representing the lower error when sensitive species are predicted to be absent.</td></tr>
</table>

	High / Good Boundary	**Good / Moderate Boundary**
	Sensitive taxa dominate, highly sensitive taxa are scarcer and account for about half the contribution of sensitive taxa. Tolerant taxa are present, but remain subordinate. Highly tolerant taxa, if present are rare. Macrophyte functions at high status all remain intact, undesirable disturbances are rare and macrophyte cover is stable over time.	

Annex 2: Remarks on the record of macrophyte abundance

Besides taxonomic composition abundance is one of the main criteria used to define the ecological status according to the WFD normative definitions for the BQE "macrophytes and phytobenthos". Macrophyte taxa generally show a patched distribution according to the variation of the micro-habitats (substrate, velocity, depth etc.). Therefore, the assessment of macrophytes in rivers is mainly based on a complete record of the present taxa in an entire stretch of stream. This technique is fundamentally different from sampling benthic invertebrates or diatoms. Data from these groups are acquired on the basis of representative spot-checks and their abundance is often extrapolated. Most survey protocols of river macrophytes define the relative cover of each taxon in relation to the considered area as a proxy of abundance. The cover is estimated as the absolute percentage or as a class. Currently, national protocols compared in the intercalibration refer to a 5-class system of abundance for calculating indexes and metrics.

The absolute cover for aquatic vegetal species can be considered as mainly influenced by

- growing pattern of each species (e.g. large area covered by only one plant of *Nuphar lutea* compared to small area covered by any helophytic plant),
- light available on the water/substrate surface,
- mobility and hospitality of the substrate,
- habitat disturbances inducing unbalanced and proliferating population, like nutrient enrichment (e.g. high cover of *Potamogeton pectinatus* or filamentous algae), channelization (filamentous algae), impoundment etc.

The total cover is not a representative parameter of the ecological status even when considering individual river types, but rather a descriptor of the particular and local functioning of the aquatic system. Thus, the use of total abundance as an indication of anthropogenic disturbance needs to be taken with caution.

For most of national methods, the assessment of macrophytes provides an overall view of the trophic level including all parameters of primary productivity: nutrients, chemical macro-elements, temperature, light and flow velocity. This trophic level is reflected by two parameters: taxonomic composition (trophic preferences of each taxon) and population balance. Disturbances may be indicated by the development of tolerant species and the decrease of taxa evenness. In this line the relative abundance is more sensitive to assess a shift of the population balance excluding the natural constraints than the total cover can be.

Against this background we consider that:

- Abundance is taken into account by all assessment methods by weighting the ecological indication of each taxon.
- We recommend methods using only composition metrics with presence/absence approach to integrate the information given by the relative abundance of each taxon.
- Relative abundance is more relevant for evaluating the trophic level by assessing macrophyte populations.
- The assessment of absolute abundance is not essential if macrophytes are used as indicators of the trophic level.

Annex 3: Supplementary exercises

To allow for the intercalibration of assessment methods that did not participate in the prime exercise finishing in July 2011, national good status boundaries were compared against the average boundary positions (i.e. „harmonisation guideline“) set in the prime exercise.
The harmonised EQRs derived from the common dataset (see Section 5 of this report) acted as a common metric against which the national EQRs were related. The average boundary positions derived in the prime exercise were defined as follows:

- Sandy lowland brooks (R-C1): high-good=0.591; good-moderate=0.478
- Siliceous mountain brook (R-C3): high-good=0.670; good-moderate=0.452
- Medium-sized lowland streams (R-C4): high-good=0.610; good-moderate=0.509

A supplementary exercise was carried out individually for each national method/intercalibration type and comprised the following steps:

1. Calculating the national assessment scores for all surveys in the common dataset and transforming these scores into EQRs based on the national reference score relevant for the intercalibration type.
2. Applying benchmark standardisation (continuous benchmarking, see Section 6.2 of this report) to the national EQRs using benchmark site information in the common dataset.
3. Relating the benchmark standardised national EQRs against the harmonised EQRs and checking intercalibration feasibility (Pearson's correlation coefficient > 0.5).
4. Comparing the national boundary positions with the harmonisation guideline and testing comparability (boundary bias ≤ |0.25|, see Section 8.3 of this report).
5. Adjusting national class boundaries if boundary bias was exceeded.

The outcomes of these supplementary exercises including the harmonised national boundary values are summarised in Table A2 on the next page.

Table A2: Results of the supplementary intercalibration exercise
R_P – Pearson's correlation coefficient

Country	Method name	IC type	R_P	Boundary		Boundary bias	Comments
Denmark	Danish Stream Plant Index (DSPI)	R-C1	0.766	high-good	0.625[13]	0.25	Exercise based on reduced dataset (n=134) including benchmark standardisation of Danish, Italian, Dutch, Polish and British sites.
				good-moderate	0.500	0.16	
		R-C4	0.646	high-good	0.639[13]	0.25	Exercise based on reduced dataset (n=149) including benchmark standardisation of Danish, German, Latvian and Polish sites.
				good-moderate	0.500	0.06	
France	Biological Macrophyte Index for Rivers (IBMR)	R-C3	0.915	high-good	0.930	-0.05	-
				good-moderate	0.790	0.06	
		R-C4	0.848	high-good	0.905	0.23	-
				good-moderate	0.790	0.06	
Ireland	Mean Trophic Ranking	R-C4	0.557	high-good	0.740	0.15	MTR reference value: 50
				good-moderate	0.620	-0.25	
Luxembourg	Biological Macrophyte Index for Rivers (IBMR)	R-C3	0.915	high-good	0.890	-0.15	IBMR reference value: 13.43; benchmarking based on French benchmark sites
				good-moderate	0.790	0.07	
		R-C4	0.848	high-good	0.890	-0.05	IBMR reference value: 10.49; no sites from Luxembourg in the common database, benchmarking based on French benchmark sites
				good-moderate	0.790	-0.11	

[13] Denmark will keep its original high-good boundary of 0.7.