JRC Scientific and Technical Reports

Water Framework Directive intercalibration technical report

Part 1: Rivers

Edited by Wouter van de Bund

EUR xxxxx EN - 200*

Draft 28 August 2008

JRC EUROPEAN COMMISSION

ies Institute for Environment and Sustainability

The mission of the Institute for Environment and Sustainability is to provide scientific-technical support to the European Union's Policies for the protection and sustainable development of the European and global environment.

European Commission
Joint Research Centre
Institute for Environment and Sustainability

Contact information
E-mail: wouter.van-de-bund@jrc.it

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

Legal Notice
Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this publication.

Europe Direct is a service to help you find answers
to your questions about the European Union

Freephone number (*):

00 800 6 7 8 9 10 11

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server http://europa.eu/

JRC [PUBSY request]

EUR XXXXX LL
ISBN X-XXXX-XXXX-X
ISSN 1018-5593
DOI XXXXX

Luxembourg: Office for Official Publications of the European Communities

© European Communities, 20YY

Reproduction is authorised provided the source is acknowledged

Printed in Country

Contents

Contents 3

Section 1 - Introduction 5
1. Preface 5
2. Background 5
3. Geographical Intercalibration Groups (GIGs) and common Intercalibration types . 9
3.1. Geographical Intercalibration Groups 9
3.2. Common Intercalibration types 11
4. References 13

Section 2 – Benthic Macroinvertebrates 14
1 Introduction 15
2 Methodology and results 15
2.1 Central-Baltic Geographical Intercalibration Group 15
2.1.1 Intercalibration approach 15
2.1.2 National methods that were intercalibrated 26
2.1.3 Reference conditions and class boundary setting 27
2.1.4 Results of the comparison 28
2.1.5 Harmonisation 31
2.2 Northern Geographical Intercalibration Group 39
2.2.1 Intercalibration approach 39
2.2.2 National methods that were intercalibrated 41
2.2.3 Reference conditions and class boundary setting 41
2.2.4 Results of the comparison 42
2.2.5 Results of the harmonisation – Boundary EQR values 45
2.3 Alpine GIG 48
2.3.1 Intercalibration approach 48
2.3.2 National methods that were intercalibrated 49
2.3.3 Reference conditions and class boundary setting 50
2.3.4 Results of the comparison 50
2.3.5 Results of the harmonisation – Boundary EQR values 51
2.4 Mediterranean GIG 52
2.4.1 Intercalibration approach 52
2.4.2 National methods that were intercalibrated 58
2.4.3 Reference conditions and class boundary setting 58
2.4.4 Results of the comparison 61
2.4.5 Results of the harmonization 62
2.5 Eastern Continental GIG 64
2.5.1 Intercalibration approach 64
2.5.2 National methods that were intercalibrated 67
2.5.3 Reference conditions and class boundary setting 68
2.5.4 Results of the comparison 70
2.5.5 Results of the harmonisation – Boundary EQR values 70
3 Discussion 71
3.1 Comparability between GIGs 71
3.2 Open issues and need for further work for river macroinvertebrates 71
3.2.1 Central-Baltic GIG 71
3.2.2 Northern GIG 73

3.2.3 Alpine GIG 74
3.2.4 Mediterranean GIUG 74
3.2.5 Eastern Continental GIG 74
4 References 76

Section 3 - Phytobenthos 79
1 Introduction 80
2 Methodology and results 80
2.1 Northern GIG 80
2.1.1 Introduction 80
2.1.2 National approaches to assessing ecological status using phytobenthos 82
2.1.3 Test datasets 87
2.1.4 Standardisation of reference conditions 88
2.1.5 Development of Common Metric 93
2.1.6 Comparison of boundaries and harmonisation 97
2.1.7 Conclusions/Recommendations 100
2.1.8 Appendix 102
2.2 Central-Baltic GIG 104
2.2.1 Intercalibration approach 104
2.2.2 National methods that were intercalibrated 105
2.2.3 Reference conditions and class boundary setting 111
2.2.4 Results of the comparison 115
2.2.5 Results of the harmonisation – Boundary EQR values 136
2.2.6 Open issues and need for further work 147
2.2.7 Acknowledgements 149
2.3 Alpine GIG 150
2.3.1 Intercalibration approach 150
2.3.2 National methods that were intercalibrated 151
2.3.3 Results of the comparison 151
2.3.4 Results of the harmonisation – Boundary EQR values 153
2.3.5 Open issues and need for further work 155
2.4 Mediterranean GIG 156
2.4.1 Intercalibration approach 156
2.4.2 National methods that were intercalibrated 156
2.4.3 Reference conditions and class boundary setting 158
2.4.4 Results of the comparison 160
2.4.5 Results of the harmonisation – Boundary EQR values 164
2.4.6 Open issues and need for further work 169
3 Discussion 170
3.1 Comparability between GIGs 170
3.2 Open issues and need for further work 171
4 References 173
5 Glossary 175

Section 1 - Introduction

1. Preface

To be completed (should highlight that this is the result of a large collective effort of many expert groups)

2. Background

The **Water Framework Directive** (WFD) establishes a framework for the protection of all waters (including inland surface waters, transitional waters, coastal waters and groundwater). The environmental objectives of the WFD set out that good ecological status[1] of natural water bodies and good ecological potential[2] of heavily modified and artificial water bodies should be reached by 2015.

One of the key actions identified by the WFD is to carry out a European benchmarking or intercalibration (IC) exercise to ensure that good ecological status represents the same level of ecological quality everywhere in Europe (Annex V WFD). It is designed to ensure that the values assigned by each Member State (MS) to the good ecological class boundaries are consistent with the Directive's generic description of these boundaries and comparable to the boundaries proposed by other MS. The intercalibration of surface water ecological quality status assessment systems is a legal obligation.

Intercalibration is carried out under the umbrella of Common Implementation Strategy (CIS) Working Group A - Ecological Status (ECOSTAT), which is responsible for evaluating the results of the IC exercise and making recommendations to the Strategic Co-ordination Group or WFD Committee. The IC exercise aims at consistency and comparability in the classification results of the monitoring systems operated by each MS for biological quality elements (CIS WFD Guidance Document No. 14; EC, 2005). In order to achieve this, each MS is required to establish Ecological Quality Ratios (EQRs) for the boundaries between high (H) and good (G) status and for the boundary between good (G) and moderate (M) status, which are consistent with the WFD normative definitions of those class boundaries given in Annex V of the WFD.

All 27 MS of the European Union are involved in this process, along with Norway, who has joined the process on a voluntary basis. Expert groups have been established for lakes, rivers and coastal/transitional waters, subdivided into 14 Geographical

---

[1] 'Ecological status' is an expression of the quality of the structure and functioning of aquatic ecosystems associated with surface waters, classified in accordance with Annex V WFD; 'Good ecological status' is the status of a body of surface water so classified in accordance with Annex V.

[2] 'Good ecological potential' is the status of a heavily modified or artificial body of water, so classified in accordance with the relevant provision of Annex V.

Intercalibration Groups (GIGs -groups of MSs that share the same water body types in different sub-regions or ecoregions).

The IC exercise aims to ensure that the H/G and the G/M boundaries in all MS's assessment methods for biological quality elements correspond to comparable levels of ecosystem alteration (EC, 2005). Intercalibration guidance produced by CIS (WFD Guidance Document No. 14) warns that the process will only work if common EQR boundary values are agreed for very similar assessment methods or where the results for different assessment methods are normalised using appropriate transformation factors (EC, 2005). Different assessment methods (e.g. using different parameters indicative of a biological element) may show different response curves to pressures and therefore produce different EQRs when measuring the same degree of impact (EC, 2005).

In each GIG, the IC exercise will be completed for those MS that already have data and (WFD compliant) assessment methods to set boundary EQR values for some of the biological quality elements. Countries that do not have data or assessment methods already available, or do not actively participate in the current IC exercise, need to agree with the outcome of the IC exercise and harmonise their assessment methods, taking into account the results of the current exercise, when their data/methods becomes available.

The WFD refers to an 'intercalibration network', comprising sites selected from a range of surface water body types present within each ecoregion, as the basis for intercalibration (Annex V; 1.4.1). For each surface water body type selected, the WFD specifies that at least two sites corresponding to the boundary between high and good status, and between good and moderate status should be submitted by each Member State for intercalibration. However, as the IC exercise evolved, this network has become redundant, as these datasets were too small to permit robust intercalibration.

This Technical Report provides a detailed description of the work that was carried out in the framework of the EU Water Framework Directive intercalibration exercise. harmonising the classification scales of national methods for ecological classification scales for rivers across the European Union. The technical work was carried from 2004 to 2007 by groups of experts from all EU Member States, within the framework of the Common Implementation Strategy working group (2)A on Ecological Status, facilitated by a steering group lead by the European Commission Joint Research Centre (JRC) (Figure 1.1).

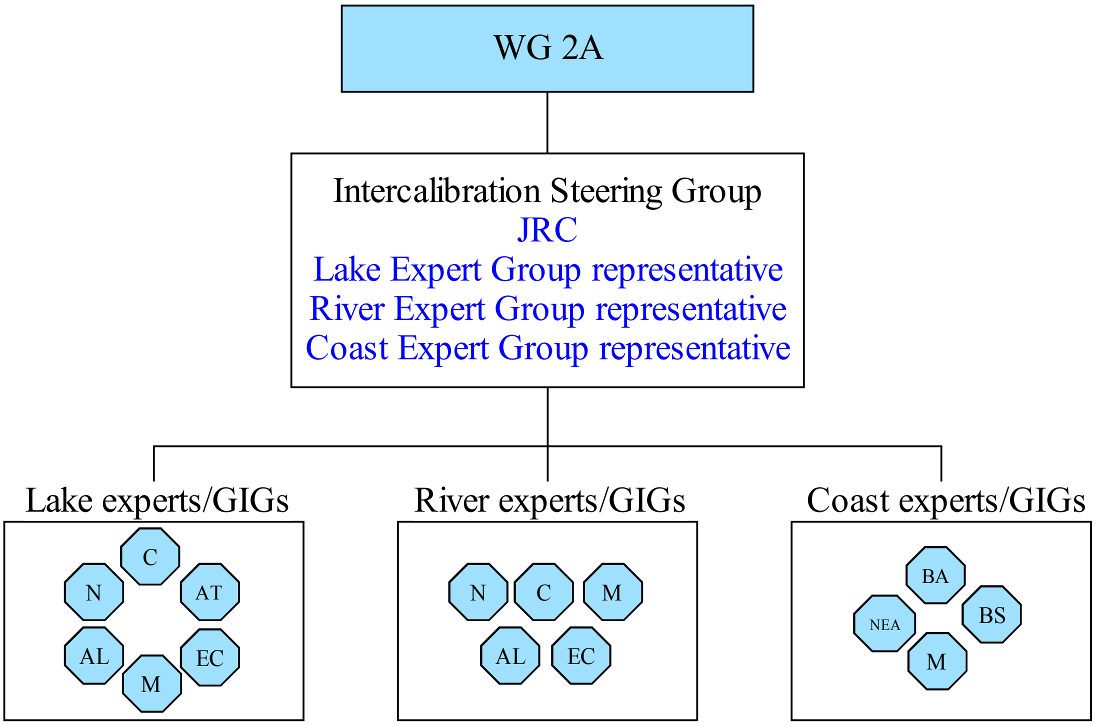


Figure 2.1 : Overview of the organisational structure of the intercalibration process (from EC 2005)

Before the start of the intercalibration exercise a guidance document (EC 2005) was agreed describing the key principles and process options for the intercalibration exercise. The key principles of the intercalibration process as described in the guidance document are reproduced below.

Key principles of the intercalibration process (from Guidance on the Intercalibration Process, EC 2005)

1. The intercalibration process is aimed at consistency and comparability of the classification results of the monitoring systems[3] operated by each Member State for the biological quality elements[4]. The intercalibration exercise must establish values for the boundary between the classes of high and good status, and for the boundary between good and moderate status, which are consistent with the normative definitions of those class boundaries given in Annex V of the WFD[5].

2. The essence of intercalibration is to ensure that the high-good and the good-moderate boundaries in all Member State's assessment methods for biological quality elements correspond to comparable levels of ecosystem alteration. Intercalibration is not necessarily about agreeing common ecological quality ratio (EQR) values for the good status class boundaries as measured by different assessment methods. Common EQR values only make sense, and are only possible, where very similar assessment methods are being used or where the results for different assessment methods are normalised using appropriate transformation factors. This is because different assessment methods (e.g. using different parameters indicative of a biological element) may show different response curves to pressures and therefore produce different EQRs when measuring the same degree of impact.

[3] The term 'monitoring system' in the way it is commonly used includes the whole process from sampling, measurement and assessment including all quality elements (biological and other). In the context of WFD Annex V, 1.4.1, the term 'monitoring system' only refers to a biological assessment method, applied as a classification tool, the results of which can be expressed as ecological quality ratios. This guidance uses the term 'WFD assessment method' in place of the term 'monitoring system' that may be misleading in this context.
[4] The WFD intercalibration as described in Annex V, 1.4.1 does not concern the monitoring systems themselves, nor the biological methods, but the classification results
[5] WFD Annex V, 1.4.1 (ii), (iii), (iv), (vi)

3. The first phase of the process is the establishment of an intercalibration network for a limited number of water body types consisting of sites representing boundaries between the quality classes High-Good and Good-Moderate, based on the WFD normative definitions. The WFD requires that selection of these sites is carried out "using expert judgement based on joint inspections and all available information[6]".

4. The Intercalibration Guidance states that "some artificial or heavily modified water bodies could be considered to be included in the intercalibration network, if they fit in one of the natural water body types selected for the intercalibration network. Artificial and heavily modified water bodies that are not comparable with any natural water bodies should only be included in the intercalibration network, if they are dominant within a water category in one or more Member States; in that case they should be treated as one or several separate water body types". An artificial or heavily modified water body is considered to fit in a natural water type if the maximum ecological potential of the artificial or heavily modified water body is comparable to the reference conditions of the natural type for those quality elements considered in the intercalibration exercise[7].

5. In the second phase of the process, each Member State's assessment method must be applied to those sites on the register that are both in the ecoregion (or, as pointed out in section 2.8, in the Geographical Intercalibration Group (GIG)) and of a surface water body type to which the system will be applied. The results of the second phase must be used to set the EQR values for the relevant class boundaries for each Member States' biological assessment system. The results of the exercise will be published by the Commission by 22 December 2006 at the latest.

6. Intercalibration sites are selected by the Member States, and represent their interpretation of the WFD normative definitions of high, good and moderate status. There is no guarantee that different Member States will have the same views on how the normative definitions should be interpreted. Differences in interpretation are reflected in the intercalibration network[8]. A common interpretation of the normative definitions should be the main outcome of the intercalibration exercise. At the end of the intercalibration exercise the intercalibration network may need to be revised according to this common interpretation.

7. The Intercalibration Exercise is focused on specific type/biological quality element/pressure combinations[9]. The selection of these combinations is based on the availability of adequate data within the time constraints of the exercise. This means that the exercise will not identify good status boundary EQR values for all the type/biological quality element/pressure combinations relevant for the implementation of the WFD. However, the Intercalibration Exercise will identify, and test the use of, a procedure and criteria for setting boundaries in relation to any such combinations[10].

8. The intercalibration process described in this guidance is aimed at identifying and resolving:

(a) Any major/significant inconsistencies between the values for the good ecological status class boundaries established by Member States and the values for those boundaries indicated by the normative definitions set out in Section 1.2 of Annex V of the WFD; and,

(b) Any major/significant incomparability between the values established for the good status class boundaries by different Member States.

9. The process will identify appropriate values for the boundaries of the good ecological status class applicable to the ecological quality ratio EQR scales produced by the Member States' assessment methods.

10. The Intercalibration Exercise will be undertaken within GIGs rather than the ecoregions defined in Annex XI of the WFD. This is to enable intercalibration between a maximum number of Member States.

11. The Intercalibration Exercise assumes that all Member States will have developed their national WFD assessment methods to a sufficient extent to enable the consistency with the normative definitions, and the comparability between Member States, of the good status boundary EQR values for those methods to be assessed during 2005. It was recognized however that this assumption might be problematic. An inventory on

[6] WFD Annex V, 1.4.1 (v)

[7] This is not the case for those quality elements that are significantly impacted by the hydromorphological alteration that has led to the water body to be designated as heavily modified.

[8] Intercalibration Guidance, section 3.5

[9] as described in the document' Overview of common Intercalibration types' (available at the intercalibration site submission web pages, http://wfd-reporting.jrc.cec.eu.int/Docs/typesmanual)

[10] If the results of the method are significantly affected by biogeographical or other ecological differences within the intercalibration type, different boundary EQR values may be appropriate for different parts of the type

the state-of-the-art in the developments of WFD compliant methods is carried out during the process of finalisation of the intercalibration network[11].

3. Geographical Intercalibration Groups (GIGs) and common Intercalibration types

3.1. Geographical Intercalibration Groups

For rivers, five Geographical Intercalibration Groups were established (Table 2.1). Each GIG was lead by a country, with the exception of the Eastern Continental GIG, tha was lead by an organisation, the International Commission for the Protection of the Danube (ICPDR). The very large Central-Baltic GIG established a steering group facilitate the work.

[11] The metadata questionnaire is available at the intercalibration site submission web pages, http://wfd-reporting.jrc.cec.eu.int/Docs/ metadata

Table 3.1 Overview of river Geographical Intercalibration Groups (GIGs) with participating countries. Lead countries are indicated in red.

Name of the GIG	**Member States comprising rivers GIGs**
Northern	Finland Ireland Norway Sweden United Kingdom
Central/Baltic	Austria Belgium Czech Republic Denmark Estonia France Germany Ireland Italy Latvia Lithuania Netherlands Poland Slovenia Slovakia Spain Sweden Luxemburg United Kingdom
Alpine	Austria France Germany Italy Slovenia Spain
Eastern Continental **(ICPDR)**	Austria Bulgaria Czech Republic Greece Hungary Romania Slovakia Slovenia
Mediterranean	Cyprus France Greece Italy Malta **Portugal** Slovenia Spain

:

3.2. Common Intercalibration types

For the Northern GIG, the intercalibration exercise was carried out for five common types (Table 3.2), hared by five countries.

Table 3.2: common river types in the Northern GIG

Type	River characterisation	Catchment area (of stretch)	Altitude & geomorphology	Alkalinity (meq/l)	Organic material (mg Pt/l)
R-N1	Small lowland siliceous moderate alkalinity	10-100 km^2	< 200 m or below the highest coastline	0.2 - 1	< 30 (<150 in Ireland)
R-N3	Small/medium lowland organic	10-1000 km^2		< 0.2	> 30
R-N4	Medium lowland siliceous moderate alkalinity	100-1000 km^2		0.2 - 1	< 30
R-N5	Small mid-altitude siliceous	10-100 km^2	Between lowland and highland	< 0.2	< 30

Countries sharing the types that have been intercalibrated
Type R-N1: Finland, Ireland, Norway, Sweden, United Kingdom
Type R-N3: Finland, Ireland, Norway, Sweden, United Kingdom
Type R-N4: Finland, Norway, Sweden, United Kingdom
Type R-N5: Finland, Norway, Sweden, United Kingdom

For the Central-Baltic GIG, the intercalibration exercise was carried out for six common types (Table 2.3), shared by 18 countries.

Table 3.3: common river types in the Central-Baltic GIG

Type	River characterisation	Catchment (km^2)	Altitude & geomorphology	Alkalinity (meq/l)
R-C1	Small lowland siliceous sand	10-100	lowland, dominated by sandy substrate (small particle size), 3-8m width (bankfull size)	> 0,4
R-C2	Small lowland siliceous - rock	10-100	lowland, rock material 3-8m width (bankfull size)	< 0,4
R-C3	Small mid-altitude siliceous	10-100	mid-altitude, rock (granite) - gravel substrate, 2-10m width (bankfull size)	< 0,4
R-C4	Medium lowland mixed	100-1000	lowland, sandy to gravel substrate, 8-25m width (bankfull size)	> 0,4
R-C5	Large lowland mixed	1000-10000	lowland, barbel zone, variation in velocity, max. altitude in catchment: 800m, >25m width (bankfull size)	> 0,4
R-C6	Small, lowland, calcareous	10-300	lowland, gravel substrate (limestone), width 3-10m (bankfull size)	> 2

Countries sharing the types that have been intercalibrated
Type R-C1: Belgium (Flanders), Germany, Denmark, France, Italy, Lithuania, the Netherlands, Poland, Sweden, United Kingdom
Type R-C2: Spain, France, Ireland, Portugal, Sweden, United Kingdom
***Type R-C3*:** Austria, Belgium (Wallonia), Czech Republic, Germany, Poland, Portugal, Spain, Sweden, France, Latvia, Luxembourg, United Kingdom

***Type R-C4*:** Belgium (Flanders), Czech Republic, Germany, Denmark, Estonia, Spain, France, Ireland, Italy, Lithuania, Luxembourg, the Netherlands, Poland, Sweden, United Kingdom
***Type R-C5*:** Czech Republic, Estonia, France, Germany, Spain, Ireland. Italy, Latvia, Lithuania, Luxembourg, Netherlands, Poland, Sweden, United Kingdom
Type R-C6: Denmark, Estonia, Spain, France, Ireland, Italy, Poland, Lithuania, Luxembourg, Sweden, United Kingdom

For the Alpine GIG, the intercalibration exercise was carried out for two common types (Table 3.4), shared by six countries.

Table 3.4: common river types in the AlpineGIG

Type	River characterisation	Catchment (km^2)	Altitude and geomorphology	Alkalinity	Flow regime
R-A1	Small to medium, high altitude calcareous	10-1000	800-2500 m (catchment), boulders/cobble	high (but not extremely high) alkalinity	
R-A2	Small to medium, high altitude, siliceous	10-1000	500-1000m (max. altitude of catchment 3000m, mean 1500m), boulders	Non-calcareous (granite, metamorphic). medium to low alkalinity	nival-glacial flow regime

Countries sharing the types that have been intercalibrated
Type R-A1: Germany, Austria, France, Italy, Slovenia
***Type R-A2*:** Austria, France, Italy, Spain, Slovenia

For the Eastern Continetal GIG, the intercalibration exercise was carried out for three common types (Table 3.5), shared by six countries.

Table 3.5: common river types in the Eastern Continental GIG

Type	River characterisation	Ecoregion	Catchment (km^2)	Altitude (m)	Geology	Substrate
R-E1	Carpathians: small to medium, mid-altitude	10	10 – 1000	500 – 800	siliceous	gravel and boulder
R-E2	Plains: medium-sized, lowland	11 and 12	100 – 1000	< 200	mixed	sand and silt
R-E4	Plains: medium-sized, mid-altitude	11 and 12	100 – 1000	200 – 500	mixed	sand and gravel

Countries sharing the types that have been intercalibrated
Type R-E1: Czech Republic, Hungary, Romania, Slovakia
Type R-E2: Czech Republic, Hungary, Romania, Slovakia
Type R-E4: Austria, Czech Republic, Hungary, Slovakia, Slovenia

For the Mediterranean GIG, the intercalibration exercise was carried out for four common types (Table 3.6), shared by eight countries.

Table 3.6: common river types in the Mediterranean GIG

Type	River characterisation	Catchment (km2)	Altitude (m)	Geology	Flow regime
R-M1	Small mid-altitude mediterranean streams	10-100	200-800	Mixed	Highly seasonal
R-M2	Small/Medium lowland mediterranean streams	10-1000	<400	Mixed	Highly seasonal
R-M4	Small/Medium mediterranean mountain streams	10-1000	400-1500	Non-silicious	Highly seasonal
R-M5	Small, lowland, temporary	10-100	<300	Mixed	Temporary

Countries sharing the types that have been intercalibrated
Type R-M1: France, Greece, Italy, Portugal, Slovenia, Spain
Type R-M2: France, Greece, Italy, Portugal, Spain
Type R-M4: Cyprus, France, Greece, Italy, Spain
Type R-M5: Cyprus, Italy, Portugal, Slovenia, Spain

4. References

EC (2005). Common implementation strategy for the water framework directive (2000/60/ec). Guidance on the Intercalibration process 2004-2006. Luxembourg, Office for Official publications of the European Communities. http://circa.europa.eu/Public/irc/env/wfd/library

Section 2 – Benthic Macroinvertebrates

1 Introduction

For the quality element Benthic Macroinvertebrates the intercalibration exercise has been completed for all five geographical intercalibration groups, covering all EU Member States (plus Norway).

2 Methodology and results

2.1 Central-Baltic Geographical Intercalibration Group

2.1.1 Intercalibration approach

Common intercalibration types and countries sharing the types

Within the Central-Baltic GIG GIG six common intercalibration types were defined (Table 2.1.1), that are shared by 18 countries (Table 2.1.2).

Table 2.1.1 Central-Baltic rivers common intercalibration types

Type	**River characterisation**	**Catchment area (km^2)**	**Altitude & Geomorphology**	**Alkalinity (meq/l)**
R-C1	Small lowland siliceous - sand	10-100	lowland, dominated by sandy substrate (small particle size), 3-8m width (bankfull size)	< 0,4
R-C2	Small lowland siliceous - rock	10-100	lowland, rock material 3-8m width (bankfull size)	< 0,4
R-C3	Small mid-altitude siliceous	10-100	mid-altitude, rock (granite) - gravel substrate, 2-10m width (bankfull size)	< 0,4
R-C4	Medium lowland mixed	100-1000	lowland, sandy to gravel substrate, 8-25m width (bankfull size)	> 0,4
R-C5	Large lowland mixed	1000-10000	lowland, barbel zone, variation in velocity, max. altitude in catchment: 800m, >25m width (bankfull size)	> 0,4
R-C6	Small, lowland, calcareous	10-300	lowland, gravel substrate (limestone), width 3-10m (bankfull size)	> 2

Table 2.1.2 Countries sharing the Central-Baltic common intercalibration types

	R-C1	**R-C2**	**R-C3**	**R-C4**	**R-C5**	**R-C6**
Austria			X			
Belgium (Flanders)	X			X		
Belgium (Wallonia)			X			
Czech Republic			X	X	X	
Denmark	X			X		X
Estonia				X	X	X
France	X	X	X	X	X	X
Germany	X		X	X	X	
Ireland		X		X	X	X
Italy	X			X	X	X
Latvia			X		X	
Lithuania	X			X	X	X
Luxemburg			X	X	X	X
The Netherlands	X			X	X	
Poland	X		X	X	X	X
Portugal		X	X			
Spain		X	X	X	X	
Sweden	X	X	X	X	X	X
United Kingdom	X	X	X	X	X	X

Intercalibration approach - General overview

The intercalibration approach followed in the Central-Baltic Rivers GIG was based on a hybrid of Options 2 and 3 outlined in Annex III of the Intercalibration Process Guidance (EC, 2005). In this approach boundaries are initially set separately by each Member State (as in Option 3), then compared to a common metric (as in Option 2), and harmonised where necessary. Common metrics enable a GIG-wide comparison of class boundaries. For this approach to be successful it is essential that there is agreement within the GIG on criteria to derive reference conditions; to ensure this, the procedure and criteria applied by each country for selecting reference sites were carefully evaluated as a part of the intercalibration process.
In this intercalibration approach it is not necessary to compile a single data set at the GIG level, avoiding the problem of collating data from different countries applying different methods. Instead, Member States use their own data to calculate a common metric, and compare this to their national assessment results. It was possible to follow this approach because most Member States had relatively well-developed river macroinvertebrate assessment methods in place at the start of the intercalibration exercise, and because a robust common metric was available (the ICMi was developed for this purpose within the STAR research project; see Buffagni et al., 2005).
Because initially the class boundaries are set by Member States using their own data and methods, it is necessary to compare and harmonise the different steps of the class boundary setting procedure within the GIG to ensure that the boundaries meet the requirements of the WFD.

The intercalibration approach comprises the following basic steps:

- Evaluation of national methods, reference conditions, and boundary setting: each Member State provided information on their national assessment method, including a explanation of how the high-good and good-moderate class boundaries were set. Methods and boundary setting procedure were evaluated in the GIG for compliance with the requirements of the Water Framework Directive The GIG agreed on common criteria to identify reference sites. Each Member State collated data according to the CB GIG common intercalibration types and identified the reference sites in the dataset applying the common criteria. The correct application of those criteria was evaluated in the GIG.
-
- Comparison of the boundaries on a common scale: The GIG agreed on a common metric (the 'Intercalibration Common Metric Index' – ICMi). Using data submitted by each fo the Member States, a linear regression between the ecological quality ratios (EQRs) of the common metric and each of the national assessment methods. The national high-good and good-moderate boundary values were transformed into ICMi values using the regression formula. This allowed MS boundaries to be compared with the boundaries of other MS on a common (ICMi EQR) scale.
- Harmonisation: GIG average high-good and good-moderate boundary ICMi values were calculated, including only those Member States whose methods and boundary setting procedures were accepted by the GIG in the first (evaluation) step. A range around this boundary value was then defined (the 'harmonisation band'). Member States whose ICMi boundaries fell below this band were required to adjust, unless they were able to provide a convincing scientific explanation why their boundaries should be different.

The steps involved in the evaluation, comparison and harmonisation stages of the process are summarised in Figure 2.1.1, and explained in further detail below.

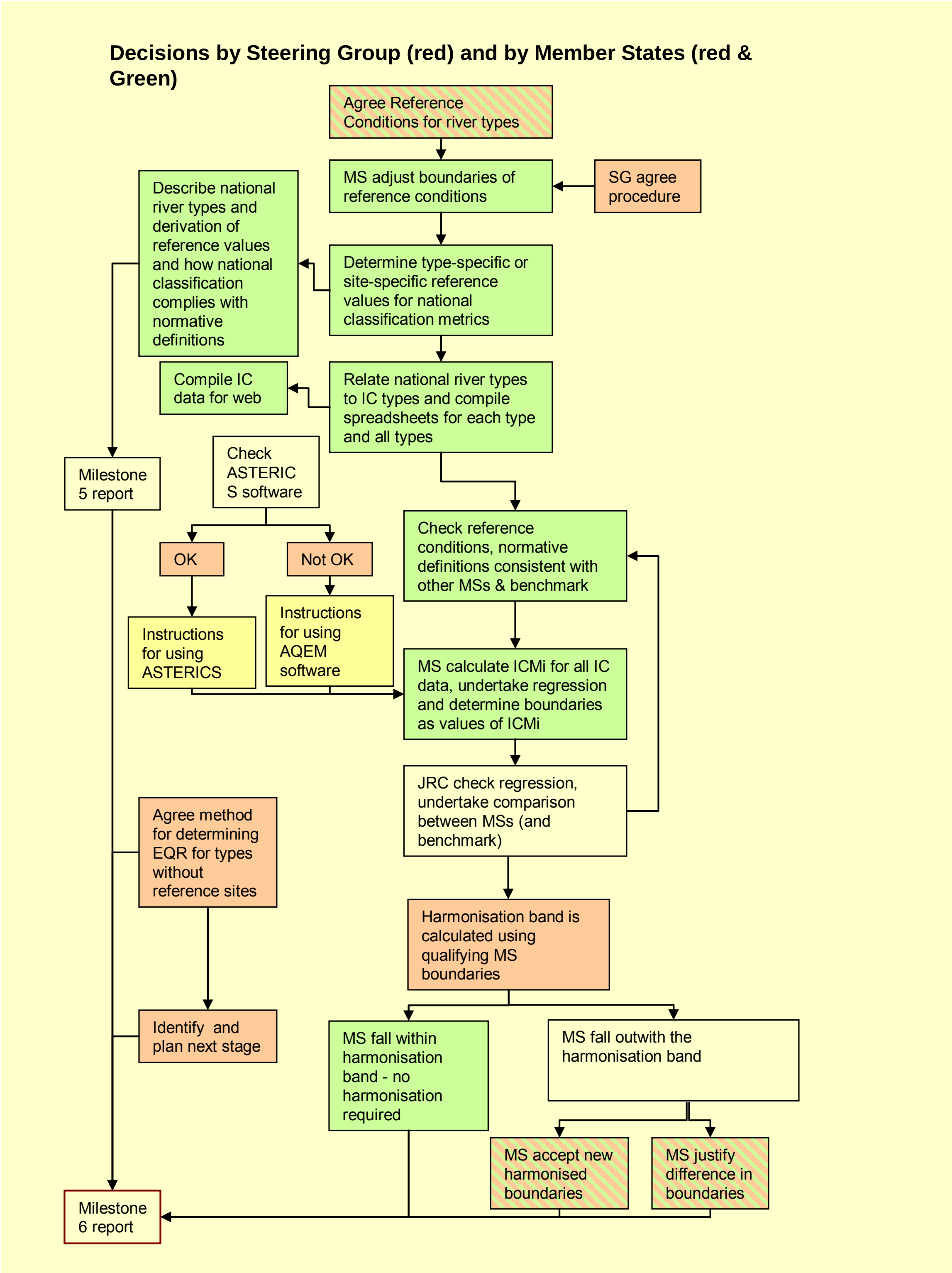


Figure 2.1.1: Flow diagram to demonstrate the CB GIG rivers comparison and harmonisation procedure, and the roles of the Steering Group and the Member States in the process (further explanations see text).

<u>Evaluation of national methods, reference conditions, and boundary setting</u>

Each Member State identified and described their national river macroinvertebrate classification method and explained how reference conditions and class boundaries were set, using common templates.
Reference sites were chosen by Member States following the principles outlined in the REFCOND guidance. The GIG agreed on more specific criteria for reference sites, based on catchment land use and type-specific concentrations of key chemical parameters. Two sets of thresholds were established – reference thresholds and rejection thresholds; Figure 2.1.2 shows how the criteria were applied. Member States were asked to complete a checklist indicating which of the GIG defined reference criteria were used for the screening exercise and to specify the sources of information that were used by the Member State for this process. The Steering Group of the GIG verified this information and ensured that Member States adhered to the correct screening procedure using the information provided in the check list.

The national methods were initially evaluated by members of the Central-Baltic river GIG Steering Group, whose conclusions were endorsed by the GIG as a whole, taking into acount the following aspects:

- Review of the compliance of national assessment and classification methods with WFD requirements
- Completion of the boundary setting template
- Completion of the reference conditions template
- Evaluation if the reference condition criteria were correctly applied
- Evaluation if the Member State assessment method and boundary setting procedure were in agreement with the requirements of the WFD

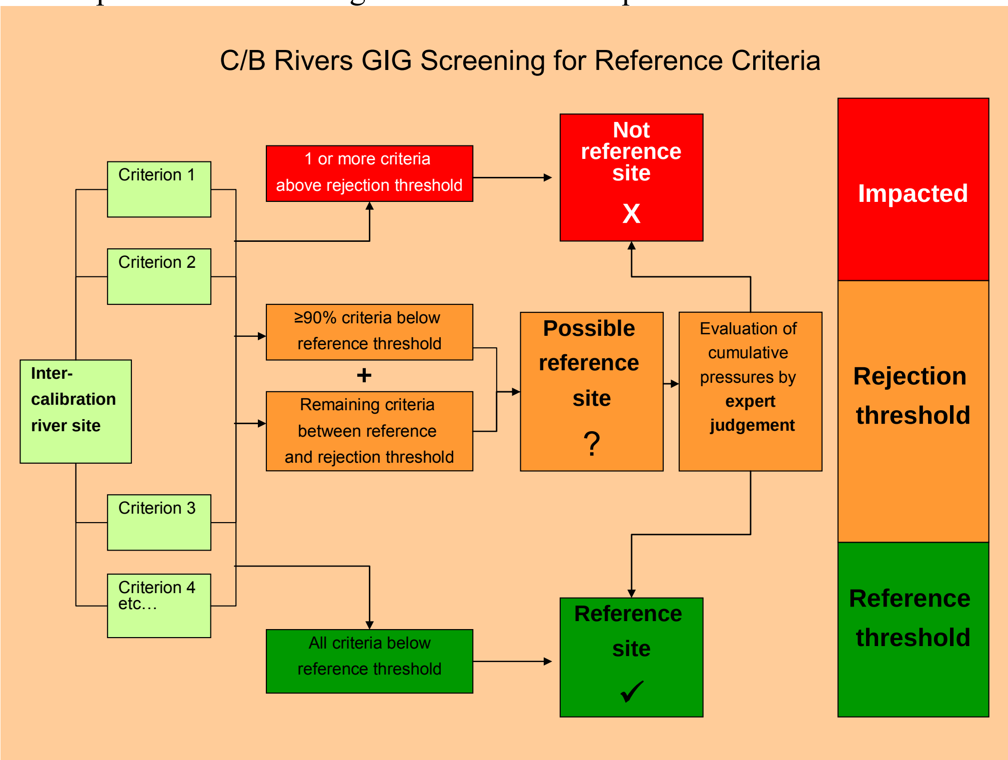


Figure 2.1.2: Flow diagram of the procedure for validating reference sites. "Reference thresholds" and "rejection thresholds" were agreed within the GIG; if one or more of the criteria are above the rejection threshold a site should rejected as a reference site, if up to 10% of the criteria are between the reference and rejection threshold the refernce site should be validated using expert judgment.

The Intercalibration Common Metric Index (ICMi)

The ICMi is a multimetric index, covering the four main aspects of the definitions for high, good and moderate ecological status for river benthic invertebrates (WFD Annex V, 1.2.1). The following six metrics were used (see Table 2.1.3 for more details):

- Average Score Per Taxon (ASPT)
- Log_{10}(sel_EPTD+1)
- 1-GOLD
- total number of taxa (families)
- number of EPT (Ephemeroptera, Plecoptera and Trichoptera) taxa (families)
- Shannon-Wiener diversity index

The ICMi was calculated as a weighted average of all the metrics, taking into account the conceptual group to which each metric belongs (Table 2.1.3). This gives an equal weighting to each of the three groups.

The ICMi fulfils the requirements of the WFD normative definitions because each criterion is addressed by 2 or 3 of the metrics combined in the ICMi (Figure 2.1.3).

- The change in taxonomic composition and abundance is mainly evaluated through Number of taxa, EPT taxa, and diversity (Shannon) index.
- Diversity is evaluated through Number of taxa and Shannon index.
- Sensitive taxa are mainly evaluated with ASPT (for organic + nutrient), abundance of selected EPT (mainly accounting for hydro-morphological degradation).
- The balance of important functional groups is evaluated with the 1-GOLD metric.

Table 2.1.3: The Intercalibration Metrics (ICMs) used in the Intercalibration Common Metric index (ICMi) (Buffagni et al., 2005).

Intercalibration Common Metrics (ICMs) studied					
Information type	Metric type	Metric name	Taxa considered in the metric	Literature reference	weight
Tolerance	Index	ASPT	Whole community (Family level)	e.g. Armitage *et al.*, 1983	0.333
Abundance/ Habitat	Abundance	Log_{10} (Sel_EPTD +1)	Log(sum of Heptageniidae, Ephemeridae, Leptophlebiidae, Brachycentridae, Goeridae, Polycentropodidae, Limnephilidae, Odontoceridae, Dolichopodidae, Stratyomidae, Dixidae, Empididae, Athericidae & Nemouridae)	Buffagni *et al.*, 2004; Buffagni & Erba, 2004	0.266
	Abundance	1-GOLD	1 - (relative abundance of Gastropoda, Oligochaeta and Diptera)	Pinto *et al.*, 2004	0.067
Richness and Diversity	Taxa number	Total number of Families	Sum of all Families present at the site	e.g. Ofenböch *et al.*, 2004	0.167
	Taxa number	number of EPT Families	Sum of Ephemeroptera, Plecoptera and Trichoptera taxa	e.g. Ofenboch et al., 2004; Böhmer et al., 2004.	0.083
	Diversity index	Shannon-Wiener diversity index	$D_{S-W} = -\sum_{i=1}^{s}\left(\frac{n_i}{A}\right)\cdot\ln\left(\frac{n_i}{A}\right)$	e.g. Hering *et al.*, 2004; Böhmer *et al.*, 2004.	0.083

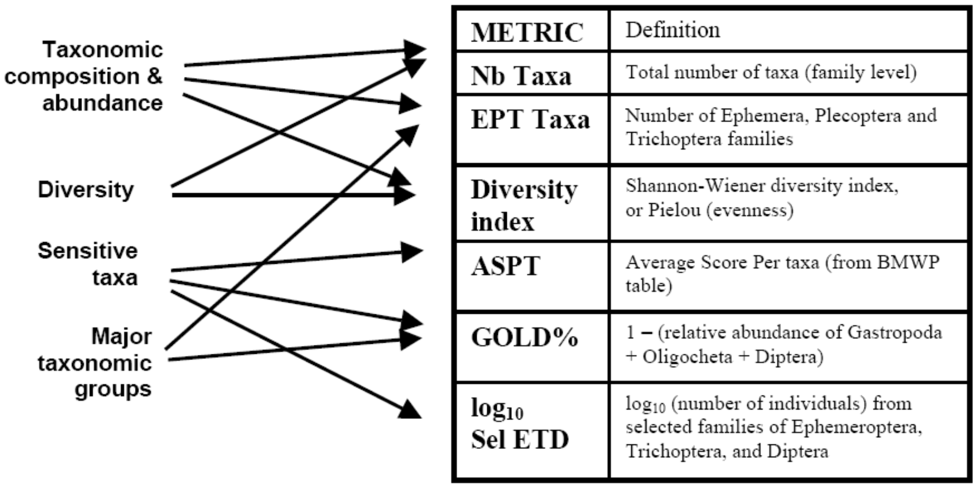


Figure 2.1.3: Coverage of different aspects included in the WFD normative definitions of high, good and moderate status for river benthic invertebrates by the metrics included in the Intercalibration Common Metric index (ICMi)

An overall introductory overview of the ICMi and the included metrics is reported among the results of the research project STAR (Buffagni et al., 2005). Examples of the response of the ICMi to general and/or specific pressure indicators (from Member States' data and the research project REBECCA dataset) are provided in Annex 2.1.5. It was concluded that most of the metrics included in the ICMi respond both to single stressors and to general degradation (see Table 2.1.4) and that the ICMi takes into account all important stressors occurring in European rivers (Buffagni et al., 2005).

Table 2.1.4: Estimated response of the metrics included in the Intercalibration Common Metric Index (ICMi) to the most important stressors (modified from Buffagni et al., 2005)

Metrics	Organic + Nutrients	Hydro-morphology	Toxics	General
Total NB taxa	X	X	X	XX
EPT taxa	XX	(X)	(X)	XX
Diversity index	X		X	X
ASPT'	XXX		(X)	
1 – GOLD	X			
Log Sel. ETD	X	XX		XX

Comparison of class boundaries

For the comparison, each Member State provided a dataset using standardised Excel spreadsheets. The basis information for each sample were macroinvertebrate family-level abundances, allowing the calculation of the common metric ICMi. Additionally, the EQRs (Ecological Quality Ratios) of national assessment method and MS class boundaries were provided.
As a minimum, each dataset included 6 samples from reference sites (identified according to the common criteria), and 4 samples of high, good, and moderate class according to the national classification. Each country compiled a separate data set for each of the common intercalibration types they shared. Detailed instructions were issued explaining the procedure for converting national class boundaries to ICMi EQR values (Murray-Bligh et al., 2006). The EQRs from the national assessment method were correlated with the corresponding EQRs from the ICMi, using the median value of the samples from reference sites in the dataset as reference value. A linear regression was performed and the r^2 value was calculated. National boundary values were transformed into ICMi EQR values using the regression formula.

All calculations were initially carried out by the Member State experts in the GIG. For each country the data set and the calculations were screened by members of the Central-Baltic river GIG Steering Group within the GIG, whose conclusions were endorsed by the GIG as a whole. The evaluation criteria are summarised in Table 2.1.5. As part of the evaluation of the datasets, a report detailing the full list of acceptance criteria required for the comparison was compiled for each Member State along with the evaluations of whether the Member State met each of the required criteria (see Annex 2.1.3.3)

Table 2.1.5: Acceptance criteria for inclusion of national datasets in the calculation of a GIG boundary for the CB GIG macro-invertebrate intercalibration exercise.

Acceptance Criteria
Provision of raw family lists in the national dataset
Provision of physio-geographical parameter values (catchment size, altitude, geology, substrate, additional parameters) for checking type allocations
Reference sites and samples available (checked by the GIG criteria) - minimum number of sites: 2 - minimum number of samples: 6
Number of test sites/samples per quality class according to national classification - high: 4 samples (incl. reference samples) - good: 4 samples (incl. reference samples) - moderate: 4 samples - poor: if not provided → still acceptable - bad: if not provided → still acceptable
Exploration of relationship between national method and ICMi R square is checked, low values (R^2<0.5) are flagged and excluded from confidence interval averaging in boundary comparison and harmonisation
Discontinuous national indices: in class boundary translation via regression, use only values that occur in national method (no artificially derived mean values); in each case boundary values generally belong to next higher class
Review of intercalibration typology data
Median of MS EQR derived from reference samples according to GIG criteria should be around 1; if not, countries must provide a satisfactory explanantion
Method and boundary values are finalised and officially endorsed by the Member State

GIG average values were calculated for the high-good and the good-moderate ICMi EQR boundaries. Only the boundaries from those Member States meeting all the acceptance criteria were included, in order to ensure compliance with the requirements of the Water Framework Directive.

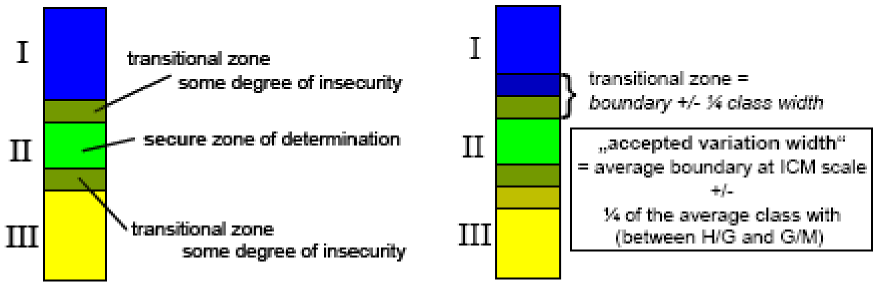


Figure 2.1.4 Uncertainty in classification and harmonisation bands (see Annex 2.3.1 for further explanation). **Left:** Every status class can be segmented in an zone where the status evaluation is secure (high confidence) and in transitional zones to the neighbouring status classes, where the status evaluation is to some extent uncertain. **Right**: The transitional zone between two status classes is supposed to the "accepted width of variation", estimated at ¼ of the average class width, defining the harmonisation band

Ranges around the GIG average high-good and good-moderate boundary values were defined and termed the harmonisation band with a width of 0.1 on the ICMi EQR scale (GIG average +/- 0.05 – corresponding to ¼ of the average class width). The width of the harmonisation band is based on a basic assessment of uncertainty in the boundary values expressed on the ICMi scale that was carried out within the Alpine river GIG (see Annex 2.3.1 and Figure 2.1.4). Uncertainty is caused by data limitations, natural variability and the simplification principles of the ICMi. Due to these and other sources of variation, classification uncertainty is high in a transitional between neighbouring status classes. This "insecure" zone of assessment is estimated to be ¼ of the status class width (equivalent to 0.05 of the ICMi_EQR scale). More detailed and quantitative estimates of accuracy and precision are lacking in most countries at the moment.
The Central-Baltic GIG also considered and tested an alternative method, calculating harmonisation bands as the average value of the 95 percentile confidence intervals of the ICMi boundaries; this option resulted in narrower bands (typically ca. 0.03 instead of 0.05), but it was concluded that this approach was problematic from the statistical point of view, and that the uncertainties introduced in the different steps of the intercalibration process are too large to justify such a narrow band.
The high-good and good-moderate boundaries of all Member States were compared to the GIG average boundary value (and associated harmonisation band) with appropriate warning "flags" for those boundaries derived from non-compliant datasets.

The whole procedure was initially carried out in two different ways:

- separately for each of the common intercalibration types. Type-specific regressions were used to calculate type-specific harmonisation bands.
- Combining the normalised (EQR) data of all common intercalibration types. A single regression was used per Member State, combining all

common types, still using type-specific references to calculate the EQRs to take account of typological differences.

The outcomes of both methods were compared and reviewed, and it was decided by the GIG to use the latter (all types combined) for the harmonisation step (see chapter 2.1.4 for further details).
A substantial majority decision was reached by the GIG on the recommended use of the 'all types combined' option based on a single regression for each country. The Central-Baltic GIG Steering Group considered comments from participating countries received prior to their meeting in September 2006 in reaching this recommendation.
A number of issues underlined the recommended use of the 'all types combined' option based on a single regression for each country; these included:

- There is uncertainty in the exercise caused by the distribution of data and statistics.
- The variation in MS IC river type boundary values attributable to differences in IC typology cannot be quantified.
- Type-specific variation within each IC river type cannot be ruled out.
- The range of variability between MS boundary values is greater than the variability between river types.

Bearing in mind these uncertainties, it was considered appropriate to determine GIG boundary values for H/G and G/M that were not river type specific. This was not a unanimous view, however, and one MS was of the strong opinion that type-specific boundaries would be more appropriate.
Calculations were carried out by Wouter van de Bund (JRC) and Nicolas Mengin (CEMAGREF, France) using the data provided by each MS.

<u>Harmonisation of MS Boundaries</u>
The high-good and good-moderate boundaries of each Member State were checked against the harmonisation band. For each boundary the following outcomes were possible:

1. The Member State's ICMi_EQR boundary lies within the harmonisation band.
   - No action required.
2. The Member State's ICMi_EQR boundary lies above the harmonisation band.
   - No action required.
3. The Member State's ICMi_EQR boundary lies below the harmonisation band. The Member State has two options:
   - **Harmonise the boundaries**: Member States whose ICMi_EQR boundary values that fall below the harmonisation band should adjust their national class boundaries so that the equivalent ICMi_EQR boundary falls within the GIG harmonisation band.
   - **Justify the differences**: Member States should justify why they do not accept the GIG mean boundary. In this case, a scientific or technical note explaining why the MS boundary differs from the GIG boundary and harmonisation band (e.g. due to typological differences) should be provided, and accepted by the GIG.

2.1.2 National methods that were intercalibrated

Table 2.1.6 below indicates the name of the national classification method for macroinvertebrates used by each country in Central-Baltic GIG, with information on the status of development of the method and a reference. Each country compiled a fact sheet describing their national methodology and the criteria used for boundary setting at a national level in detail (Annex 2.1.2). Many of the methods were still in development and/or were changed during the period when the intercalibration exercise was carried out. The status of the methods in the table shows the situation at the time that this part of the report was edited (December 2007).

Table 2.1.6: Member State national classification methods compared in the Central-Baltic GIG macro-invertebrate intercalibration exercise (status of the method reflects the situation in December 2007)

	Method	**Status**	**Reference**
Austria	Austrian System for Ecological River Status Assessment (Worst case between Multimetric Indices for General Degradation and Saprobic Index	Agreed national method	Ofenböck et al., 2007
Belgium (Flanders)	Multimetric Macroinvertebrate Index Flanders (MMIF)	Agreed national method	Gabriels, 2007
Belgium (Wallonia)	Indice Biologique Global Normalisé (IBGN)	Agreed national method	Norme AFNOR NF T 90 350, 1992
Czech Republic	Multimetric index	Under development	-
Denmark	Danish Stream Fauna Index (DSFI)	Agreed national method	-
Estonia	British Average Score Per Taxon (ASPT)	National method in development	-
France	French WFD classification Indice Biologique Global Normalisé (IBGN	Agreed national method	Norme AFNOR NF T 90 350 (1992) and circular MEDD/DE 05 n°14 (July 05
Germany	PERLODES –Bewertungsverfahren von Fließgewässern auf Basis des Makrozoobenthos	Agreed national method	LAWA-AO, 2006
Ireland	Quality Rating System (Q-value)	Agreed national method	-
Italy	STAR Intercalibration Common Metric index	Agreed national method	ÌRSA-CNR, Notiziario dei Metodi Analitici, Marzo 2007; Buffagni et al., 2005
Latvia	Saprobic Index	Under development	-
Lithuania	Biotic index (BI), Danish stream fauna index (DSFI). Also possible calculation of ASPT, BMWP	Still in development	-
Luxembourg	Indice Biologique Global Normalisé (IBGN)	Agreed national method	Norme AFNOR NF T 90 350, 1992
Netherlands	KRW-maatlat	Agreed national method	Van der Molen & Pot, 2007
Poland	BMWP (BMWP-PL) verified by modified Margalef diversity index	Development of new method compliant with WFD under development	
Spain	North Spain Multimetric Indices	Agreed national method for Type B rivers in North Spain	Pardo, Álvarez & Roselló, 2007
Sweden	Multimetric index; DJ-index)	Agreed national method (NFS 2008:1)	Dahl & Johnson 2004
United Kingdom	RIVPACS	Agreed national method	-

2.1.3 Reference conditions and class boundary setting

Reference conditions
The reference data selection and validation process is documented in detail in Annex 2.1.2, containing the following documents:

- Annex 2.1.2.1 Rationale for Reference Thresholds of selected chemical parameters for Central-Baltic GIG Intercalibration.
- Annex 2.1.2.2: Chemical thresholds values.
- Annex 2.1.2.3: Reference criteria checklist completed by MS.

A total number of 888 reference sites were identified within the GIG (see Table 2.1.7). The number of reference sites varies considerably, both between countries (from 0 in Belgium/Flanders and the Netherlands to 243 in France) and between types (from 24 for RC5 to 253 for RC3).
No reference sites could be identified in Belgium (Flanders) and the Netherlands according to the agreeed criteria, and therefore alternative approaches for quantifying reference conditions were needed for those countries. Both countries provided reports describing how they derived their reference conditions, and demonstrating a sufficiet level of comparability with the other Member States in the GIG (see Annex 2.1.4).

Table 2.1.7: Number of samples from reference sites selected by Member States according to the CB GIG defined criteria for each common intercalibration river type.

	RC1	RC2	RC3	RC4	RC5	RC6	Grand Total
AT			25				25
BE-F	0			0			0
BE-W			20				20
CZ			7				7
DE	6		20	6			32
DK	5			9		7	21
EE				6	5	5	16
ES		16	35	10	10	6	77
FR	23	50	107	21		42	243
IE		116		13	9	66	204
IT	32						32
LT				6		10	16
LU			39	18		26	83
NL	0			0			0
PL	8						8
SE		14					14
UK	25	16		30		19	90
Grand Total	99	212	253	119	24	181	888

Class boundary setting

Each country compiled a fact sheet describing their national methodology and the criteria used for high-good and good-moderate boundary setting at a national level in accordance with the normative definition outlined in Annex V of the WFD. This information is provided in Annex 2.1.2.

On a general level, the compliance of the intercalibration outcome with the WFD normative definitions is guaranteed because the ICMi takes into account all relevant aspects.

The GIG agreed that the main issues for the interpretation of the normative definitions are the following:

- how to quantify taxonomic composition, abundance, disturbance sensitive taxa, diversity and major taxonomic groups. This has been done by defining the ICMi as described in section 2.1.1 (see also Buffagni *et al.*, 2006).
- what constitutes a slight and a moderate deviation from reference conditions. The normative definitions themselves do not give any clarification of the meaning of 'slight' and 'moderate'. In most cases the Member States indicate that there is a lack of obvious break points or thresholds in the relationship between their classification metrics and pressures, and as a consequence the interpretation of 'slight' and 'moderate' is rather arbitrary. Even so, some Member States have been able to justify the boundary setting protocol template in an objective way. The approach followed in the intercalibration process for macro-invertebrates has been to compare the results of each Member State's method to a common set of WFD compliant metrics, combined in an ICMi. The data screening procedure and acceptance criteria (described in Section 2.1.1) aimed to ensure that MS class boundaries would be comparable on the ICMi-EQR scale. Only Member States that fulfilled all the agreed CB GIG criteria were included in the calculation of the harmonisation band.

In additiona, an independent "benchmark classification" was made available from the STAR project. Here the class boundaries were set independently by the scientists involved in the project, according to the methodology outlined in Buffagni *et al.* (2005). The underlying dataset did not cover all CB GIG countries and the derivation of the classification used by the project partners was not completely transparent and consistent. The CB GIG decided therefore not to use the benchmark boundaries as a basis for harmonisation within the GIG, but to include it in the comparison exercise to check if the Member State's boundaries are in line with the benchmark classification.

2.1.4 Results of the comparison

Data from a total of almost 15,000 samples was brought together. Table 2.1.8 shows the numbers broken down by common intercalibration type and by country.
Annex 2.1.3 contains the full results of the comparison of MS (and benchmark) high-good (H/G) and good-moderate (G/M) boundary values on the common ICMi_EQR

scale. MS boundary values for the 'all types combined' regression for H/G and G/M boundaries are summarised in Table 2.1.9.

Table 2.1.8: Number of macro-invertebrate samples per Member State for each common intercalibration river type.

	Bench-mark	RC1	RC2	RC3	RC4	RC5	RC6	Grand Total
AT				67				67
BE-F		193			185			378
Benchmark	401							401
BE-W				50				50
CZ				101				101
DE		68		170	88			326
DK		49			36		45	130
EE					22	16	27	65
ES			97	158	220	44	26	545
FR		127	378	462	185		424	1,576
IE			2,319		1,071	221	2,815	6,426
IT		365						365
LT					72		73	145
LU				98	58		140	296
NL		374			508			882
PL		59						59
SE		71						71
UK		502	188		924		1,338	2,952
Grand Total	401	1,737	3,053	1,106	3,369	281	4,888	14,835

After reviewing the information provided by the Member States desribing their methods, the reference conditions setting, and the class boundary setting procedure (see Annex 2.1.1, 2.1.2 and 2.1.3), the GIG decided to include the methods from nine countries in the calculation of the harmonisation bands: Austria, Belgium (Wallonia), Germany, Spain, United Kingdom, Italy, Luxemburg, and Ireland. Data from eight countries (LT, NL, PL, BE-F, CZ, DK, EE, SE) were not included in the calculation of the GIG boundaries for reasons including:

- National boundaries not agreed yet.
- National assessment method not fully developed.
- Reference values were chosen using an approach that differs to that outlined by the CB GIG (described in Section 2.1.3).
- Data quality issues (insufficient number of samples or reference sites; poor regression between the national system and the ICMi).

Of the remaining countries, the Netherlands, Belgium (Flanders), and Denmark provided explanations at a later stage (see Annexes 2.1.4.2, 3 and 5). It was agreed within the GIG that those explanations were sufficient to demonstrate consistency with the WFD normative definitions, and that the comparison exercise gave valid results for those MS. Data from nine countries were included in the calculation of the

H/G and G/M boundaries (AT, BE-W, DE, ES, FR, UK, IT, LU, IE); these MSs occur to the left of the red line in Figures 2.1.5 and 2.1.6.
Please refer to Annex 2.1.3.3 for a detailed summary of the Type Coordinators recommendations of MS datasets that should be included in the calculation of the GIG boundary and the reasons for excluding other MS datasets from this calculation.

Table 2.1.9 H/G and G/M boundary values for national methods (MS H/G and MS G/M) and boundary EQR values (EQR H/G and EQR G/M) based on the all types combined comparison. Equivalent EQR_ICMi values are also shown for each MS. Countries indicated in green were included in the calculation of the GIG boundary and harmonisation band, countries indicated in yellow and red were not because they failed one of the criteria. Countries indicated in yellow have provided additional information at a later stage demonstrating consistency with the WFD normative definitions.

MS	MS H/G	MS G/M	EQR_MS H/G	EQR_MS G/M	EQR_ICMi H/G	EQR_ICMi G/M
band					0.93	0.76
benchmark					0.95	0.79
AT	0.80	0.60	0.80	0.60	0.93	0.72
BE-W	17.00	13.00	0.97	0.74	0.95	0.73
DE	0.80	0.60	0.80	0.60	0.93	0.82
ES	0.93	0.73	0.93	0.73	0.97	0.82
FR	0.92	0.80	0.94	0.80	0.88	0.78
UK	0.97	0.86	0.97	0.86	0.92	0.74
IT	0.96	0.72	0.96	0.72	0.96	0.72
LU	14.70	11.00	0.96	0.72	0.95	0.74
IE	4.50	4.00	0.85	0.75	0.93	0.82
NL	0.80	0.60	0.80	0.60	0.93	0.77
BE-F	0.80	0.60	0.80	0.60	0.77	0.58
DK	7.00	5.00	1.00	0.71	1.00	0.76
LT	9.00	7.00	0.95	0.74	0.93	0.82
PL	0.90	0.68	0.89	0.68	0.88	0.71
SE	0.80	0.60	0.80	0.60	1.03	0.92
CZ	0.80	0.60	0.80	0.60	0.93	0.82
EE	6.00	5.00	0.92	0.77	0.91	0.70

2.1.5 Harmonisation

The GIG boundary values for H/G and G/M based on the average boundary values of all accepted MS datasets are shown in Table 2.1.10 along with the harmonisation band (or acceptable range). The harmonisation band represents the GIG boundary values for H/G and G/M +/- 0.05 of the ICMi_EQR scale.
The position of each MS ICMi_EQR boundary in relation to the GIG boundary and harmonisation band for the H/G and G/M boundaries is illustrated in Figures 2.1.5 and 2.1.6, respectively. Details on the calculations are presented in Annex 2.1.3.2.

Table 2.1.10 EQR_ICMi boundary values for H/G and G/M including upper and lower limits of the harmonisation bands.

EQR_ICMi H/G	EQR_ICMi G/M		H/G	G/M
0.94	0.76	Upper band	0.99	0.81
		Lower band	0.89	0.71

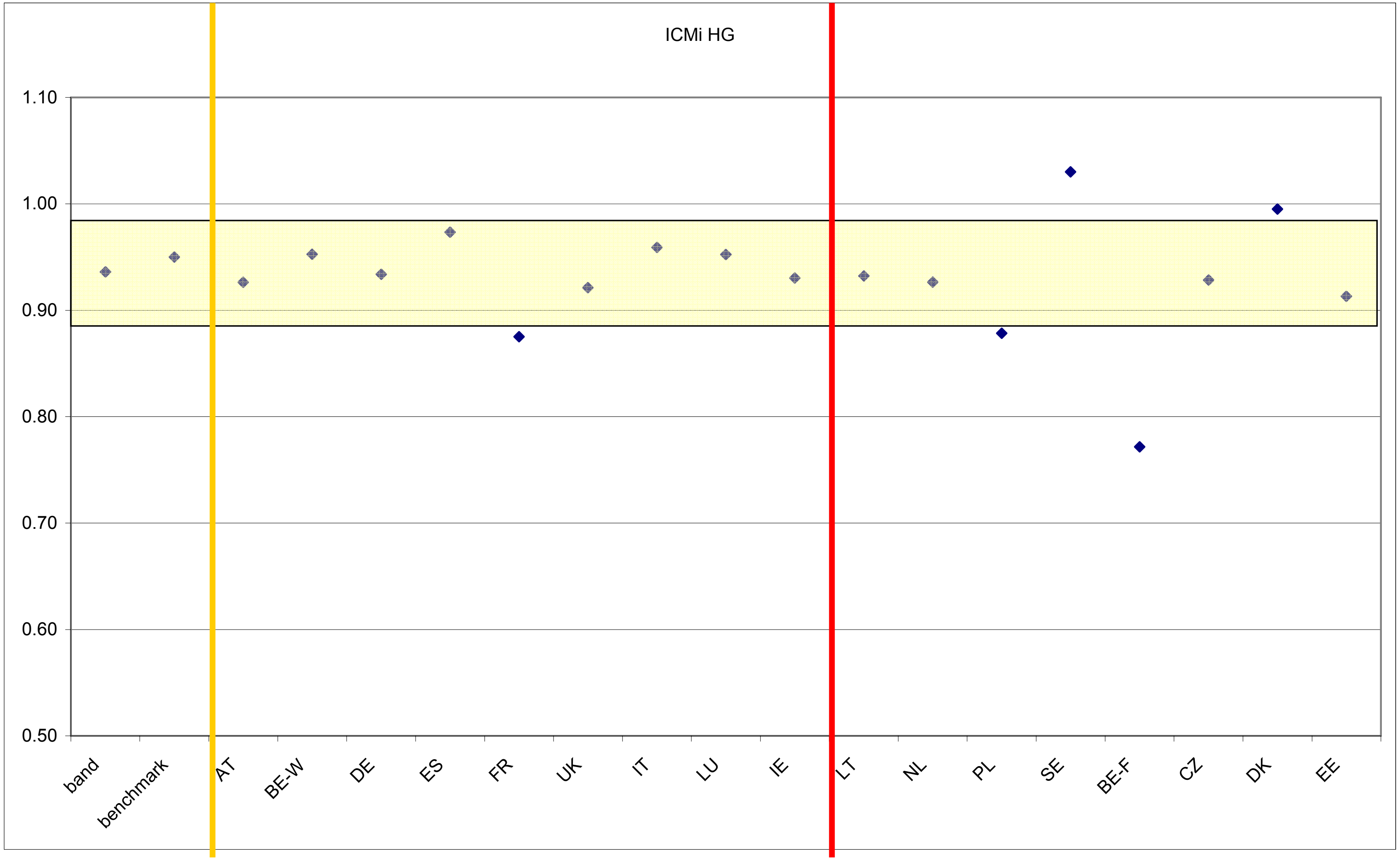


Figure 2.1.5: Results of the 'all types combined' comparison showing MS EQR_ICMi values for the H/G boundary. 'Band' represents the GIG H/G boundary value. MS to the left of the red line contributed to the calculation of the GIG boundary. MS to the right of the red line did not contribute to the GIG boundary. The yellow 'harmonisation band' represents +/- 0.05 of the ICMi_EQR scale around the GIG boundary value.

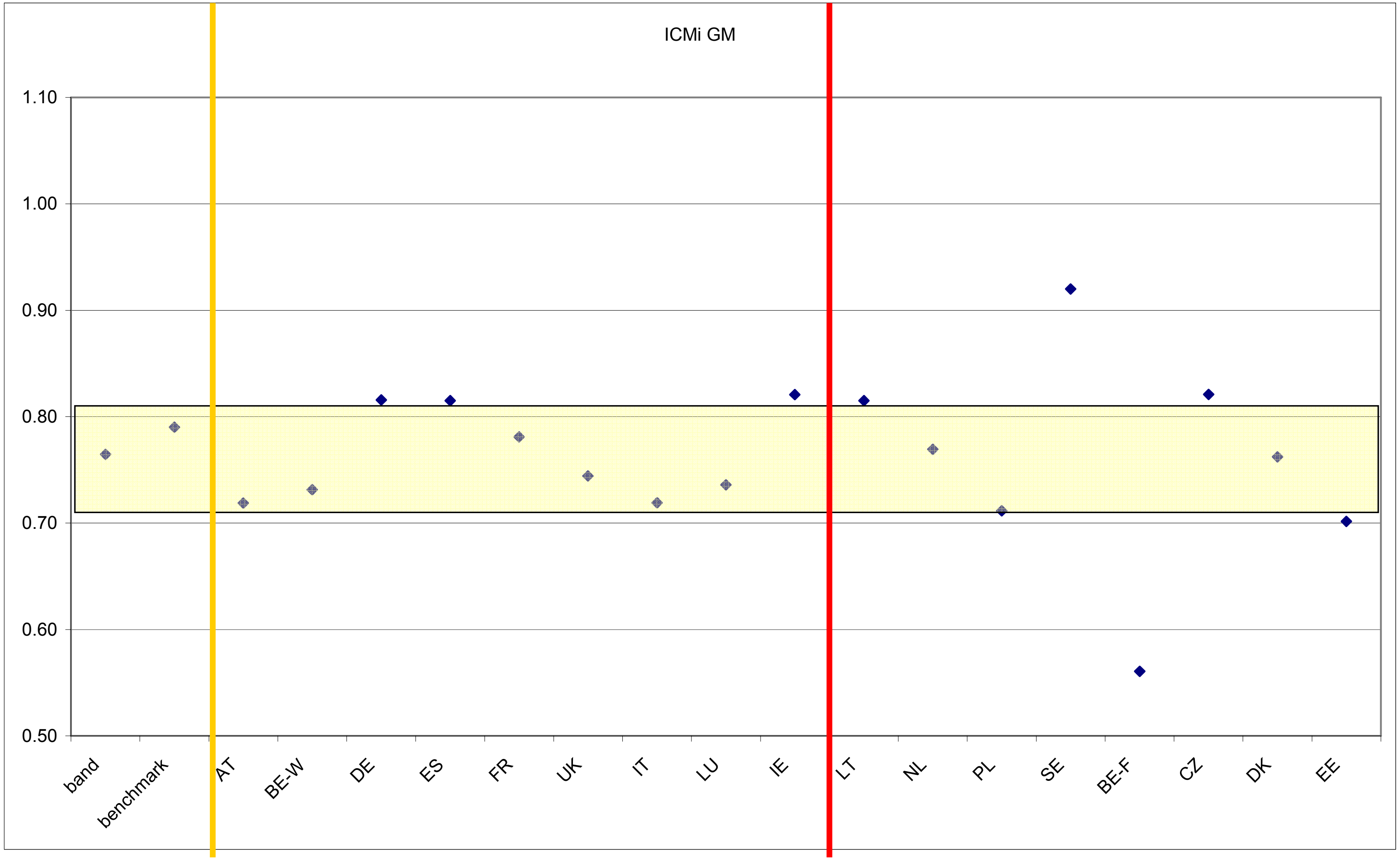


Figure 2.1.6: Results of the 'all types combined' comparison showing MS EQR_ICMi vlues for the G/M boundary. 'Band' represents the GIG G/M boundary value. MS to the left of the red line contributed to the calculation of the GIG boundary. MS to the right of the red line did not contribute to the GIG boundary. The yellow 'harmonisation band' represents +/- 0.05 of the ICMi_EQR scale around the GIG boundary value**.**

High/Good Boundary

The results of the comparison for the H/G boundary (Table 2.1.10) indicate that the following countries fall within the harmonisation band:

- AT, BE-W, DE, ES, UK, IT, LU, IE, NL, [LT, CZ, EE].

The following countries lie above or below the H/G harmonisation band:

- FR (below), PL (below), , BE-F (below), DK (above), SE (above).

Good/Moderate Boundary

The results of the 'all types combined' comparison for the G/M boundary (Table 2.1.11) indicate that the following countries fall within the harmonisation band:

- AT, BE-W, FR, UK, IT, LU, NL, DK.
- PL lies on the lower limit of the harmonisation band.

The following countries lie above or below the G/M harmonisation band:

- DE (above), ES (above), IE (above) , BE-F (below), [LT (above), CZ (above), EE (below), SE (above).

Table 2.1.11: EQR_ICMi values for the <u>High-Good</u> boundary in relation to the harmonisation band for each Member State.

MS	EQR_ICMi for H/G boundary	Harmonisation Band EQR_ICMi	Harmonisation Required	Required Adjustment	MS Comment (cf. Annex 2.1.4)
band	0.94				
benchmark	0.95				
AT	0.93		No		
BE-W	0.95		No		Comment
DE	0.93		No		
ES	0.97		No		
FR	0.88	0.89- 0.99	Yes	0.01	Adjust
UK	0.92		No		
IT	0.96		No		
LU	0.95		No		
IE	0.93		No		
NL	0.93		No		
BE-F	0.77		Yes	0.12	Adjust
DK	1.00		No		
LT	0.93		No		
PL	0.88		Yes	0.01	Comment
SE	1.03		No		
CZ	0.93		No		
EE	0.91		No		

Notes:

'Required Adjustment' indicates the difference on the ICMi scale between the MS boundary and the harmonisation band.

MS boundaries shown in red were not required to indicate their preference for harmonisation until the national boundary values/assessment systems were fully developed and endorsed for use by the MS.

Tables 2.1.11 and 2.1.12 show MS EQR_ICMi boundary values in relation to the harmonisation band and identifies those MSs that are required to harmonise. Responses from MSs are summarised under 'MS Comment' in Tables 2.1.11 and 2.1.12; detailed MS responses are provided in Annex 2.1.4.

Table 2.1.12: EQR_ICMi values for the Good-Moderate boundary in relation to the harmonisation band for each Member State.

MS	EQR_ICMi for G/M boundary	Harmonisation Band EQR_ICMi	Harmonisation Required	Required Adjustment	MS Comment (cf. Annex 2.1.4)
band	0.76				
benchmark	0.79				
AT	0.72	0.71- 0.81	No		
BE-W	0.73		No		Comment
DE	0.82		No		
ES	0.82		No		Adjust
FR	0.78		No		
UK	0.74		No		
IT	0.72		No		
LU	0.74		No		
IE	0.82		No		
NL	0.77		No		
BE-F	0.58		Yes	0.13	Adjust
DK	0.76		No		Comment
LT	0.82		No		
PL	0.71		No		
SE	0.92		No		
CZ	0.82		No		
EE	0.70		Yes	0.01	Adjust

Notes:

'Required Adjustment' indicates the difference on the ICMi scale between the MS boundary and the harmonisation band. [12]

MS boundaries shown in red were not required to indicate their preference for harmonisation until the national boundary values/assessment systems were fully developed and endorsed for use by the MS.

Outcome of Harmonisation

Figures 2.1.7 and 2.1.8 highlight boundary values that were adjusted by MSs following the outcome of the Intercalibration exercise; please refer to Annex 2.1.4.1 for full details of these changes as well as comments from MSs who have justified the position of their boundaries. Some MSs chose to adjust their boundaries or justify the position of their boundaries regardless of the requirement to do so by the GIG. Please also refer to Annex 2.1.4 for a specific comment from Latvia to explain why the data submitted to the macro-invertebrate Intercalibration exercise was not suitable for use. Table 2.1.13 shows MS EQR and MS EQR ICMi boundary values for H/G and G/M boundaries following the incorporation of changes that were submitted during the harmonisation process.

[12] Results for the Swedish method have been added after the work for the rest of the methods was completed

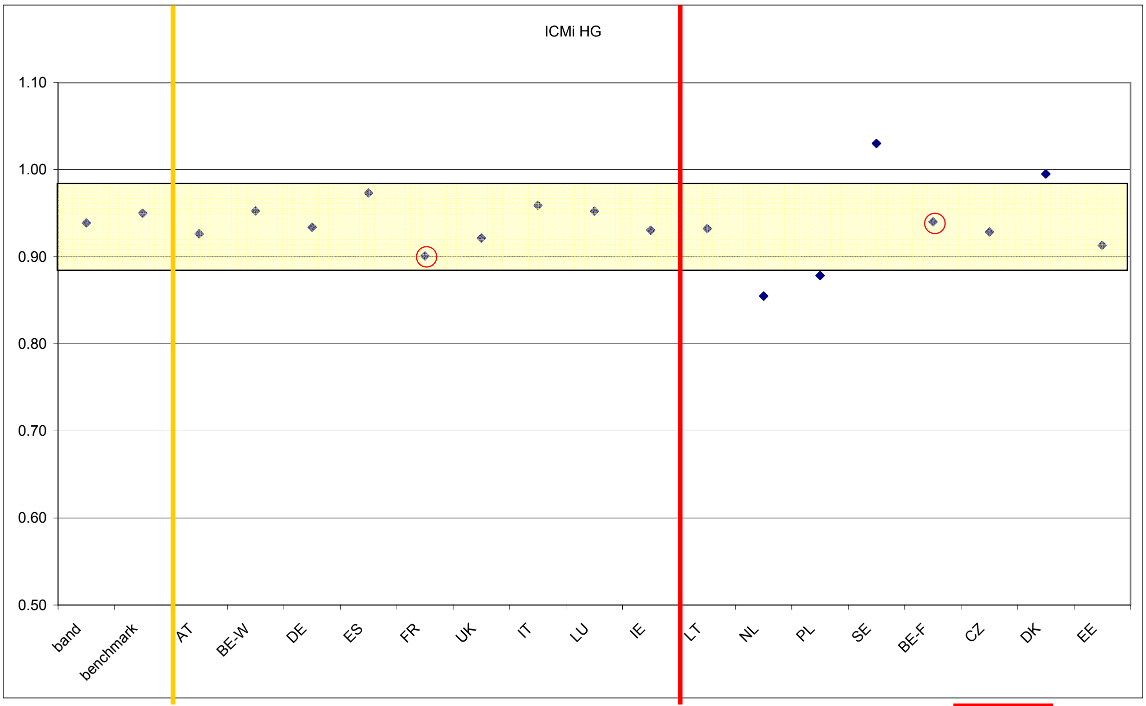


Figure 2.1.8: MS EQR_ICMi values for the H/G boundary following the incorporation of changes (red circles) made during the harmonisation stage of the Intercalibration exercise (*cf.* Annex 2.1.4 for full details of boundary changes).

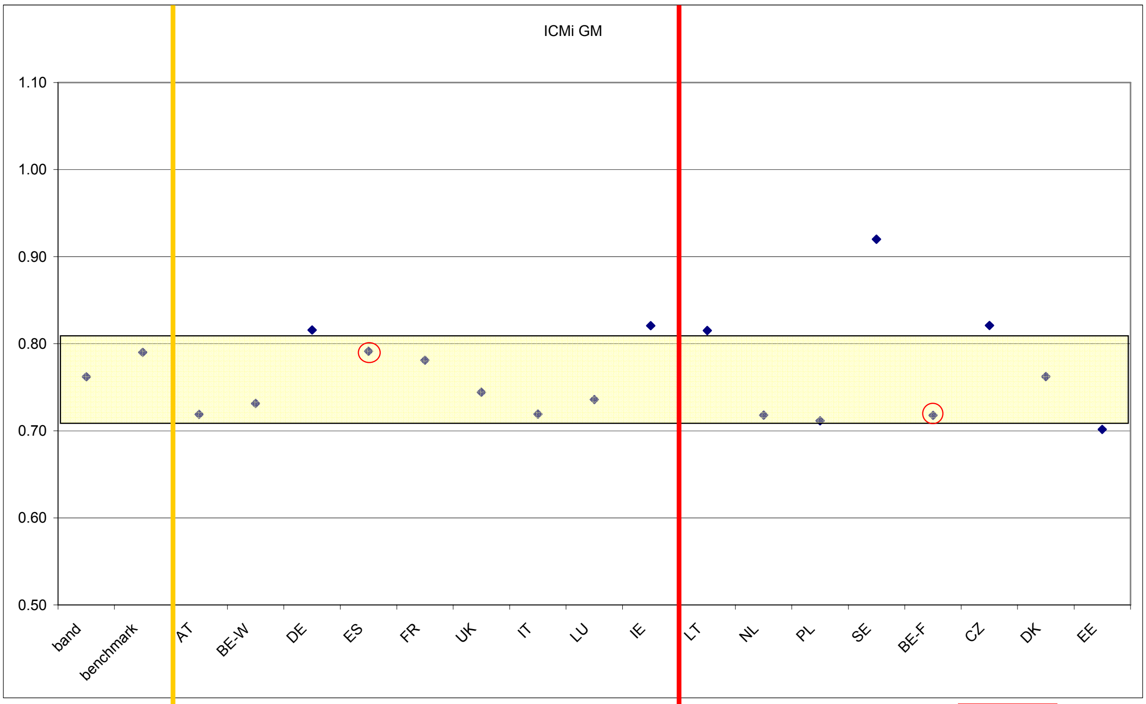


Figure 2.1.9: MS EQR_ICMi values for the G/M boundary following the incorporation of changes (red circles) made during the harmonisation stage of the Intercalibration exercise (*cf.* Annex 2.1.4 for full details of boundary changes).

Table 2.1.12: MS EQR and MS EQR ICMi boundary values for High/Good and Good/Moderate boundaries following the incorporation of changes made during harmonisation; 'grey' cells highlight boundary values that were adjusted following the outcome of the first Intercalibration exercise. NL, BE-F and DK are included in the decision. (Table 2.8).

MS	EQR_MS H/G	EQR_MS G/M	ICMi H/G	ICMi G/M
AT	0.80	0.60	0.93	0.72
BE-W	0.97	0.74	0.95	0.73
DE	0.80	0.60	0.93	0.82
ES	0.93	0.70	0.97	0.79
FR	0.94	0.80	0.90	0.78
UK	0.97	0.86	0.92	0.74
IT	0.96	0.72	0.96	0.72
LU	0.96	0.72	0.95	0.74
IE	0.85	0.75	0.93	0.82
NL	0.80	0.60	0.93	0.77
BE-F	0.90	0.70	0.94	0.72
DK	1.00	0.71	1.00	0.76
LT	0.95	0.74	0.93	0.82
PL	0.89	0.68	0.88	0.71
SE	0.80	0.60	1.03	0.92
CZ	0.80	0.60	0.93	0.82
EE	0.92	0.77	0.91	0.70

List of CB GIG Annexes

Annex 2.1.1 – Reference criteria and reference sites

- Annex 2.1.1.1: Rationale for Reference Thresholds of selected chemical parameters for the Central-Baltic GIG Intercalibration (Jean-Gabriel Wasson).
- Annex 2.1.1.2: Chemical thresholds values.
- Annex 2.1.1.3: Reference screening questionnaire completed by MS.

Annex 2.1.2 – Class boundary setting procedure for national methods

Annex 2.1.3 – Results of the comparison

- Annex 2.1.3.1: Results of type-specific comparison.
- Annex 2.1.3.2: Results of combined regression comparison (comparison and harmonisation).
- Annex 2.1.3.3: Acceptance/inclusion criteria/tables for macro-invertebrate datasets.

Annex 2.1.4 – Comments from Member States on Results of the Comparison

- Annex 2.1.4.1: General comments from MS.
- Annex 2.1.4.2: Proposal for adjusting the Flemish class boundaries according to the Intercalibration exercise for river macroinvertebrates.
- Annex 2.1.4.3: 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.
- Annex 2.1.4.4: Short comments on the current IC approach and proposal for further refinements derived from the Spanish IC exercise.
- Annex 2.1.4.5: Danish comments on the high/good boundary, reference selection and the national method (DFSI).
- Annex 2.1.4.6: Swedish justification of comparability of their updated boundary values with the CB GIG outcome

Annex 2.1.5 – Response of the ICMi vs. general and/or specific pressure indicators

- Annex 2.1.5.1: Relationship between pressure data and ICMi in Northern Spain CB GIG rivers.
- Annex 2.1.5.2: Part I: REBECCA short contribution to the IC exercise for Rivers. Part II: Relationships between the Intercalibration Common Metric index (ICMi) and a Land Cover Pressure Index for the French invertebrates IC datasets.

2.2 Northern Geographical Intercalibration Group

2.2.1 Intercalibration approach

River types and countries participating - Within the Northern GIG five common intercalibration types were initially defined (Table 2.2.1), shared by five countries (Table 2.2.2).

The intercalibration for the benthic macroinvertebrates has been completed for all Northern common intercalibration types, except R-N5 (small mid-altitude siliceous rivers). Although this type is shared by four of the five countries in the GIG, only the UK was able to provide sufficient data, making it impossible to complete the intercalbration for that type at this stage.

Table 2.2.1 Northern rivers common intercalibration types

Type	River characterisation	Catchment area (of stretch)	Altitude & geomorphology	Alkalinity (meq/l)	Organic material (mg Pt/l)
R-N1	Small lowland siliceous moderate alkalinity	10-100 km^2	< 200 m or below the highest coastline	0.2 - 1	< 30 (<150 in Ireland)
R-N3	Small/medium lowland organic	10-1000 km^2		< 0.2	> 30
R-N4	Medium lowland siliceous moderate alkalinity	100-1000 km^2		0.2 - 1	< 30
R-N5	Small mid-altitude siliceous	10-100 km^2	Between lowland and highland	< 0.2	< 30

Table 2.2.2 Countries sharing the Northern common intercalibration types

	R-N1	R-N3	R-N4	R-N5
Finland	X	X	X	X
Ireland	X	X	-	-
Norway	X	X	X	X
Sweden	X	X	X	X
United Kingdom	X	X	X	X

Pressures - The methods that were intercalibrated are aimed to detect the effects of general organic pollution and nutrient pressure. Acidification is being dealt with in a separate working group in the Northern GIG; this work has not been completed and is therefore not included in this report.

Interalibration methodology - The same general approach using the common metric ICMi was followed as was used by the Central-Baltic GIG for macroinvertebrates (described in detail in Chapter 2.1.1). The ICMi approach was developed and tested using a pan-European dataset (Buffagni et al., 2005). Sites proposed for reference conditions were screened using agreed reference criteria (Annex 2.2.3). A minimum of 4 samples were required for each river type in each presumed class – high, good, moderate, for a Member State to be included in the analysis for that NGIG river type. Sites of poor and bad status were included in the exercise, where available. For each country and for each site across the range of status classes, MS_EQR values were calculated by dividing the Member State metric by the median value of the same metric calculated for sites in reference condition. Member State status classes were assigned to all sites based on the value of the national metric at this site. The agreed common metric, ICMi, was also calculated for all sites in each Member State following the standard procedure issued by CBGIG. National class boundary values, expressed as ICMi values, were then compared in Annex 2.2.1.

Harmonisation - If a boundary would lie outside an acceptable ± 5 % band, it would be necessary for a Member State either to adjust their boundary in order to fall within the tolerance limits, or to provide a scientific explanation why the boundary is different to the mean GIG boundary – (e.g. due to ecoregional differences).

Data - A common data set was established consisting of data from all MS was used, with a total of 4502 samples; the large majority of the data was contributed by Ireland (2939 samples) and United Kingdom (1382 samples).

Table 2.2.3 Number of samples submitted for intercalibration in each river type

	R-N1	**R-N3**	**R-N4**	**All types**
Finland	-	33	-	33
Ireland	620	2319	-	2939
Norway	11	-	41	52
Sweden	-	96	-	96
United Kingdom	907	140	335	1382
Total	1538	2588	376	4502

2.2.2 National methods that were intercalibrated

The methods that were indicated are identified in Table 2.2.4; detailed descriptions of all methods can be found in Annex 2.2.2.

Table 2.2.4. National methods for river macroinvertebrates

QE: Benthic macroinvertebrates	Method	Status
Finland	Multimetric system, first version established (Hämäläinen, H. et al. 2006). Metrics in the system include ASPT, number of type-specific taxa (Hämäläinen et al. 2002), number of EPT families and PMA-index (Percent Model Affinity, Novak & Bode 1992).	Under development. National methods for classification are planned to be completed this year for most of the national types
Ireland	Quality Rating System (Q-value)	Agreed national method
Norway	Classification system under development; ASPT was used in the intercalibration exercise	Being developed to meet WFD requirements.
Sweden	Multimetric index; DJ-index (Dahl & Johnson 2004)	Agreed national method (NFS 2008:1)
United Kingdom	ASPT component of General Quality Assessment Classification (RIVPACS)	Being revised to meet WFD requirements

2.2.3 Reference conditions and class boundary setting

Reference conditions

Reference sites were chosen by MS using REFCOND guidance. A list of more detailed criteria and type-specific concentrations of key chemical parameters were agreed by the GIG. MS were asked to screen selected reference sites against agreed catchment landuse limits, and when proposed reference sites were over agreed limits, a validation with physico-chemical parameters threshold at the site scale was necessary or strongly recommended.

Reference sites are in general close to pristine with upstream catchments having minimal intensive agriculture, low population density and low levels of other pressures. Nutrients and indicators of organic pollution are also low at reference sites.

The procedure for setting reference conditions is detailed in Annex 2.2.3.

Boundary setting

All countries are currently using or are developing classification systems which are WFD compliant. Annex 2.2.2 described how boundaries were set in each of the Member State's methods.
The ICMi method (Annex 2.2.4.) is specifically designed to match the normative definitions in Annex V of the WFD. By comparing the MS status boundaries using the ICMi approach, the intercalibration process is effectively WFD compliant and takes account of all the normative definitions. The procedure is described by Buffagni et al. (2005, 2006).

2.2.4 Results of the comparison

Official NGIG Boundary Calculation Method
The High/Good (HG) Boundary for each Member State (MS) within each NGIG river type is calculated as the half-way point between the average ICMi value for the adjacent status classes of High and Good.
The Good/Moderate (GM) Boundary for each Member State within each NGIG river is calculated as the half-way point between the average ICMi value for the adjacent status classes of Good and Moderate.
The graphs below show the mean ICMi values for HG and GM calculated by simple averaging of the ICMi values for those MS with data for the individual river types.
The ±5% tolerance bands are also shown. The individual MS values are shown as points. No further harmonisation is deemed to be required if these data points fall within the ±5% tolerance bands.
More detailed results are given in the Annex 2.2.1.

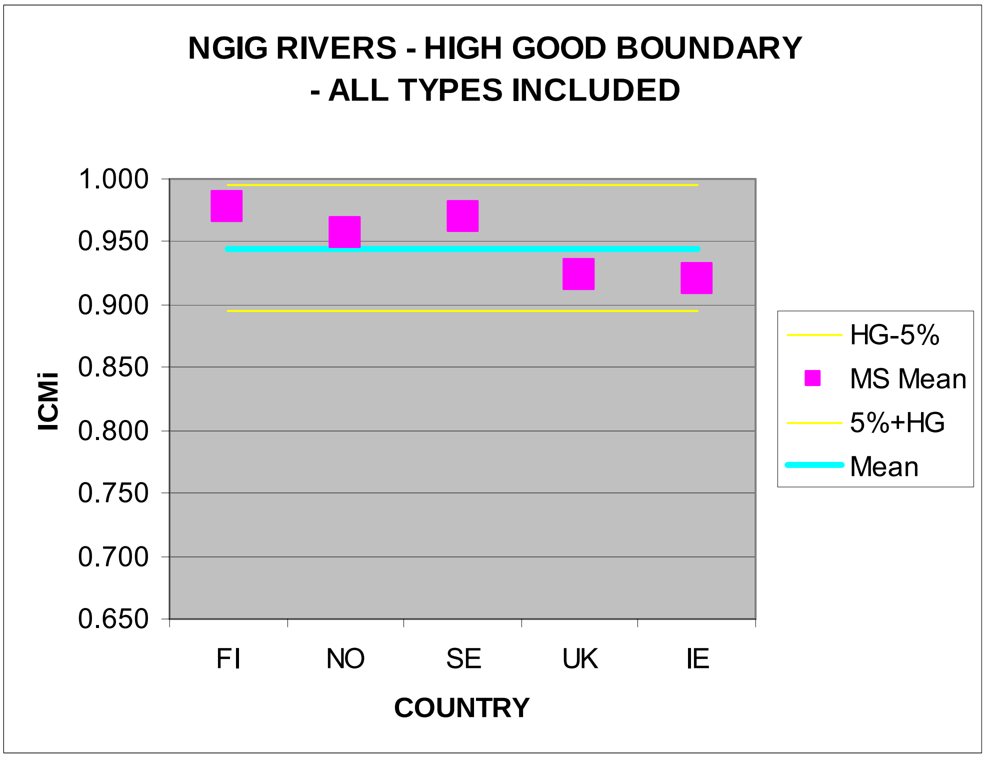


Figure 2.2.1: Results of the comparison showing MS EQR_ICMi values for the H/G boundary. 'Band' represents the GIG H/G boundary value. The 'harmonisation band' represents +/- 0.05 of the ICMi_EQR scale around the GIG boundary value.

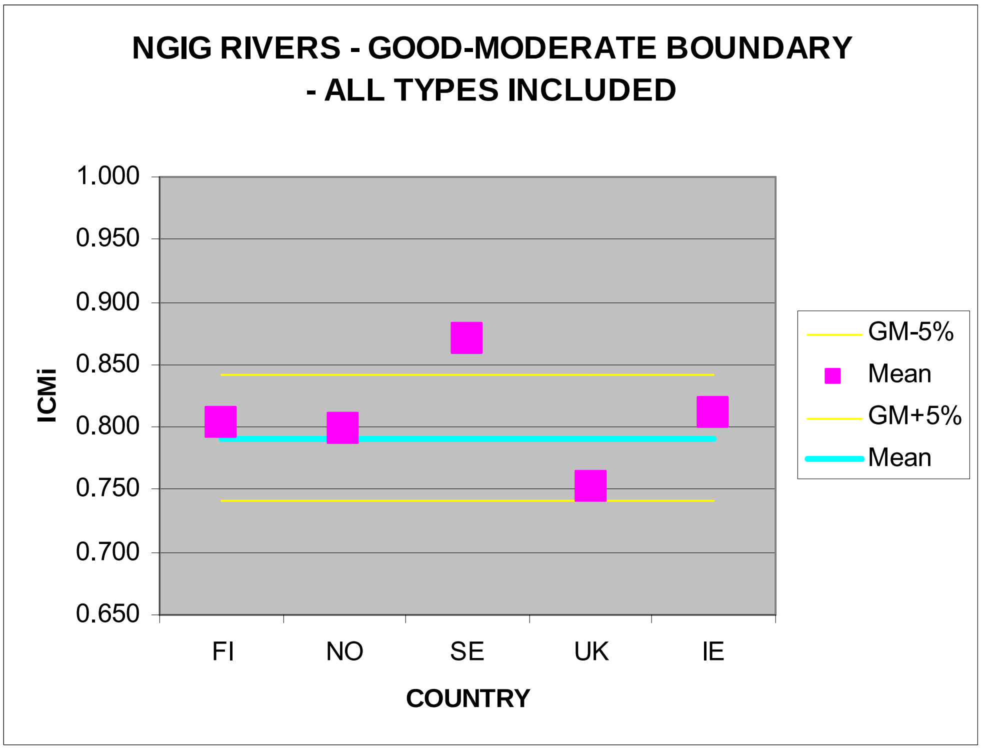


Figure 2.2.2: Results of the comparison showing MS EQR_ICMi values for the G/M boundary. 'Band' represents the GIG H/G boundary value. The 'harmonisation band' represents +/- 0.05 of the ICMi_EQR scale around the GIG boundary value.

Calculation of the Intercalibration Common Metric ICMi

The methodology for calculating ICMi is that described by Murray Bligh *et al.* (2006) (see also Annex 2.2.4) and is the same as for the Central Baltic GIG documentation.

Boundary Calculation for the ICMi

In order to agree on a common definition of the H/G and G/M boundaries, in particular the mean of all MS boundary values, calculations were carried out in the following way. The H, G and M sites were grouped and the mean ICMi scores calculated for each status class. The boundary was then calculated for each MS dataset (on a type by type and also on a combined all-type basis) by taking the half-way point between the mean of H and G status sites for the High/Good boundary and the half-way point between the mean of the G and M sites for the Good/Moderate boundary. Thus, boundaries were calculated as the half-way point between the mean of the High and Good status classes or as the half-way point between the mean of the Good and Moderate Status classes.

This method allows for accurate calculation of errors in the vicinity of the boundaries - i.e. it should not be affected by low R-squared values or non-linearity in the data.

The status class for an individual macroinvertebrate site or sample is calculated from the regression equations supplied relating ICMi_EQR to MS_EQR.

The average H/G and G/M boundaries for a type is calculated as the average of the individual country boundaries - it is not calculated based on all the samples supplied as this would weight it towards those countries supplying most data.

The calculations for the individual boundaries are shown for R-N1, R-N3 and R-N4 and for the combined all types (in Annex 2.2.1.1).

Additional test with direct regression method

Northern GIG also tested the approach used by CB GIG in comparison of boundaries. This approach compares MS boundaries as ICMi EQRs derived directly from regressions of ICMi EQR and MS EQR values. This comparison also showed comparability within a band of ± 5% of the average all MS boundary values for types. This testing provided additional assurance that N GIG boundaries have achieved comparability of ecological status, whichever method is used.

2.2.5 Results of the harmonisation – Boundary EQR values

Harmonisation of boundaries

Harmonisation Needs

The results indicate that harmonisation is not required at the ± 5 % tolerance level indicated above. All MS are within the ± 5 % band for individual NGIG river types and for all types combined using the method described for comparison above (in point *Boundary Calculation for ICMi)*

Thus it was observed that the results of the Northern GIG were harmonised.

Consideration of the results

Method of comparison showed the N GIG results to give a harmonised view for the comparisons within individual types and for all river types combined. All MS fell within the ± 5 % tolerance bands above and below the H/G and G/M boundaries. It is noted that in comparing any future additional data from, e.g. new river types or rivers from other MS not included in this initial intercalibration, the results should be compared using exactly the same method (described above in point *Boundary Calculation for ICMi).*

The ICMi metric is deliberately calculated to include all the important aspects of the normative definitions of status as defined in Annex V of the Directive. Similarly, the national metrics used are believed to adequately represent ecological status for macroinvertebrates as required by the Directive and to be WFD compliant. While some of the metrics are quite new others have been in use for a number of years. The obvious spread about the central regression lines comparing MS metrics with the ICMi values indicates that as with all biological data relatively wide uncertainty bands are to be expected especially where the R-Squared values are lower. Thus, the ± 5 % tolerance boundary is felt to be a realistic goal for the intercalibration process. The ICMi is aimed primarily at eutrophication and organic pollution and thus, where other pressures are concerned, other metrics and/or other biological quality elements may be more suitable for intercalibration of status affected by such pressures.

Relevance of the results to all types

The boundaries are expected to be appropriate to be used for all types intercalibrated and for those types not intercalibrated but falling within the extremes of the type factor values defined for NGIG types. For types that are markedly different (e. g. large rivers) the boundary values may not apply and more data would need to be gathered for these.

The river types RN-3 (and RN-5) are present in Norway, but Norway has not had satisfactory data to participate in the intercalibration of these types. If future Norwegian monitoring data show that the national boundary values for RN-3 (and RN-5), set with the combined data approach, do not reflect the observed ecological conditions of the border values in the normative definitions, Norway will choose a type-specific approach to revise the national value.

In Finland the all-types-boundaries are expected to be relevant for several national types. However, in addition to the size of the catchment, in Finland at least two geological factors have to be taken into account. In parts of Finland clay soils can dominate in catchments of small or medium-sized rivers and these types of rivers might need differing boundaries. Rivers with high humic content and naturally acidic conditions most likely need to have more specific classification criteria. Also south – north climatic gradient might have some influence inside Scandinavian countries.

There seem to be limitations of the ICMi at least in the most humic river types in the N GIG (the Finnish sites). For example, it has been observed that some of the metrics (ASPT and GOLD) have poor ability to distinguish community differences between polluted and reference river sites.

Sweden has participated in the intercalibration process. The national metric, DJ-index, is WFD compliant and will be used for the national classification. The national classification boundaries have, where changed at a late stage due to technical problems with the calculations Therefore the Swedish boundaries were not included in the calculation of the acceptable band, but the results of the comparison show that the boundaries are within (for the HG boundary) or above (for the GM boudary) the acceptable band that was calculated using the data from the other countries in the GIG.

Boundaries as ICMi Values

The boundaries are presented in the following table for all types. Values for various types are presented in Annex 2.2.1.

Table 2.2.5 Boundaries as ICMi values for all types

MS	National Boundary ICMi	ICMi, mean of all MS (range ± 5 %)
H/G		
FI	0,98	0,95 (0,90 – 1,0)
UK	0,92	
IE	0,92	
SE	0,97	
NO	0,96	
G/M		
FI	0,80	0,79 (0,74 – 0,84)
UK	0,75	
IE	0,81	
SE	0,87	
NO	0,80	

It is important to note the N GIG MS boundaries all fall within the ± 5 % tolerance band about the average boundary value for H/G and for G/M and this provides strong reassurance that NGIG original MS classification systems boundary values are comparable and do not require adjustment. This was also tested by the use of a direct regression comparison, which further assured the N GIG results.

It is noted that in comparing any future additional data from, e.g. new river types or rivers from other MS not included in this initial intercalibration, the results should be compared using exactly the same method (described above in point *Boundary Calculation for ICMi)*. Furthermore, it is noted that the boundaries according to the method of comparison (in point *Boundary Calculation for ICMi)* are the ones for the comparison and not as such compliant with the actual boundaries used in the different MS.

List of N GIG Annexes:

Annex 2.2.1 – Comparison of boundaries.
Annex 2.2.2 – Description of national classification methods.
Annex 2.2.3 – Reference criteria and reference sites.
Annex 2.2.4 − Setting of boundaries

2.3 Alpine GIG

2.3.1 Intercalibration approach

Common intercalibration types and countries sharing the types

Within the Alpine GIG GIG two common intercalibration types were defined (Table 2.3.1), that are shared by 18 countries (Table 2.3.2)..

Table 2.3.1 Alpine rivers common intercalibration types

Type	River characterisation	Catchment (km^2)	Altitude and geomorphology	Alkalinity	Flow regime
R-A1	Small to medium, high altitude calcareous	10-1000	800-2500 m (catchment), boulders/cobble	high (but not extremely high) alkalinity	
R-A2	Small to medium, high altitude, siliceous	10-1000	500-1000m (max. altitude of catchment 3000m, mean 1500m), boulders	Non-calcareous (granite, metamorphic). medium to low alkalinity	nival-glacial flow regime

Table 2.3.2 Countries sharing the Alpine common intercalibration types

	R-A1	R-CA2
Germany	X	
Austria	X	X
France	X	X
Spain		X
Slovenia	X	X
Italy	X	X

Details of the intercalibration approach followed in the Alpine GIG can be found in Annex 2.3.3.
Intercalibration was carried out for the two common intercalibration types described in
A qualitative $ICMi_{alpine}$ was used for the comparison of national boundaries, consisting from the following metrics:

- Total # taxa
- # EPT taxa
- # selected (sensitive) taxa
- ASPTiberian-2

Quantitative data were not available from all countries participating, but from Austria, France, Germany and Slovenia only. For these countries a comparison between qualitative and quantitative ICM_{alpine} was carried out. The quantitative ICM_{alpine} consists of $ASPT_{Iberian} - 2$, Log $_{10}$ (sel_sens_taxa), RETI, Total number of taxa, Number of EPT-taxa, Shannon-Wiener Diversity Index.

For setting the reference value common reference criteria were used (see Annex 2.3.2) As there is no common benchmark available for the alpine river types, the median+/- 0,05 was used to define the acceptable range of boundary values.

For the comparison national biomonitoring data covering – if possible - the entire quality gradient (acc. to national index) for all common stream types were collected. For macroinvertebrates calculations on GIG level were carried out by: Franz Wagner (Federal Agency for Water Management/Austria)
All MS data were used to set the boundary (acceptable range).

2.3.2 National methods that were intercalibrated

Table 2.3.3 National methods that were intercalibrated

QE 1: Macroinvertebrates	***Assessment Method***
Austria	Multimetric Indices for General Degradation (Structural Diversity, nutrients,…), Saprobic Index
France	French WFD classification Indice Biologique Global Normalisé (IBGN) - norm AFNOR NF T 90 350 (1992) and circular MEDD/DE 05 n°14 (July 05)
Germany	Handbuch zur Untersuchung und Bewertung von Fließgewässern auf der Basis des Makrozoobenthos vor dem Hintergrund der EG-WRRL, April 2005, www.fliessgewaesserbewertung.de
Italy	STAR Intercalibration Common Metric Index (STAR_ICMi), type adapted
Slovenia	Multimetric index (Hydromorphology), Saprobic Index
Spain	IBMWP-Iberian BMWP, IASPT

A detailed description of these methods and their relation to normative definitions is given in Annex 2.3.1.

The methods of Austria, France, Germany and Spain are officially accepted WFD methods. The method of Slovenia is finalized and in verification.The method of Italy is still in development. and included for information only.

2.3.3 Reference conditions and class boundary setting

The selection of reference sites was based on common criteria (see Annex 2.3.2). The resulting numbers of reference sites for the two common intercalibration types are given in the table below:

Table 2.3.4: Numbers of reference sites for each common intercalibration type and each country

Country	Number of reference sites	
	R-A1	R-A2
Austria	7	7
France alpine	4	21
France pyrenean	-	16
Germany	2	-
Italy*	28**	80***
Slovenia	5	-
Spain	-	12****

* new Italian dataset, delivered spring 2007
** 4 reference sites, 28 samples
*** 6 reference sites, 80 samples
**** 3 reference sites, 12 samples

For each common intercalibration type, reference values were calculated by using the median of reference sites.

In the intercalibration approach followed by the Alpine GIG, class boundary setting was done separately by each Member State. Annex 2.3.1 includes a description how this was done for each of these methods.

2.3.4 Results of the comparison

The following figure shows the comparison of the boundary values for the two intercalibration types of the Alpine GIG accomplished with the ICMi with qualitative metrics. Please note that the figure already presents the new Italian boundary values based on STAR ICMi method and the new Spanish values, but these values were not included in the calculation of the band.

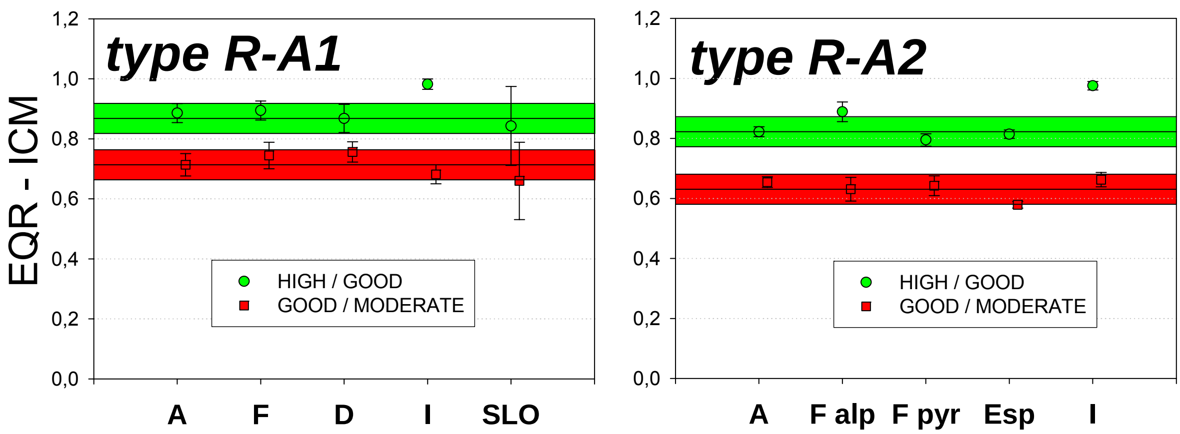


Figure 2.3.1: Boundary values high/good and good/moderate at the ICMi scale +/- 95% confidence limits. Values are taken from the regression between the EQR values of the national method and the EQR values from the ICMi method. The band indicates an „*acceptable range of variation*"and consist of the median of the boundary values of all member states +/- ¼ of the the median status class width of all member states

The ICMi with quantitative metrics is highly correlated with the ICMi with qualitative metrics – statistical values und the analysis of the ICMi*quantitative* (as seen above) is included in Annex 2.3.3.

2.3.5 Results of the harmonisation – Boundary EQR values

The following table includes boundary values for each national method, the corresponding boundary value for the ICMi and the agreed acceptable range.

Table 2.3.5 Boundary values for each national method

MS	National boundary		National boundary ICMi		Agreed range ICMi		o.k.
Type R-A1	**H/G**	**G/M**	**H/G**	**G/M**	**H/G**	**G/M**	
Austria	0,80	0,60	0,89	0,71			√
France	0,93	0,79	0,89	0,74			√
Germany	0,80	0,60	0,87	0,76	0,82 - 0,92	0,66 - 0,76	√
Italy	0,97	0,73	0,98	0,68			
Slovenia	0,8	0,6	0,84	0,66			√
Type R-A2							
Austria	0,80	0,60	0,82	0,65			√
France alp	0,93	0,71	0,89	0,63			√
France pyr	0,94	0,81	0,80	0,64	0,77 - 0,87	0,58 - 0,68	√
Spain	0,83	0,53	0,81	0,58			√
Italy	0,95	0,71	0,98	0,66			√

*: Spain agrees on changing its National G/M boundary to be inside the proposed range: ICMi: 0,58 ⇒ New National Boundary: 0,53

List of Alpine GIG Annexes:
Annex 2.3.1 – National methods and normative definitions;
Annex 2.3.2 – Criteria for reference conditions;
Annex 2.3.3 – Technical aspects of the comparison of the boundary values by using the ICMi – method and Final results.

2.4 Mediterranean GIG

2.4.1 Intercalibration approach

Common intercalibration types and countries sharing the types

Within the Mediterranean GIG GIG 4 common intercalibration types were defined (Table 2.4.1), that are shared by 7 countries (Table 2.4.2).

Table 2.4.1 Mediterranean rivers common intercalibration types

Type	River characterisation	Catchment (km2)	Altitude (m)	Geology	Flow regime
R-M1	Small mid-altitude mediterranean streams	10-100	200-800	Mixed	Highly seasonal
R-M2	Small/Medium lowland mediterranean streams	10-1000	<400	Mixed	Highly seasonal
R-M4	Small/Medium mediterranean mountain streams	10-1000	400-1500	Non-silicious	Highly seasonal
R-M5	Small, lowland, temporary	10-100	<300	Mixed	Temporary

Table 2.4.2 Countries sharing the Mediteranean common intercalibration types

	R-|M1	R-M2	R-M4	R-M5
France	X	X	X	
Greece	X	X	X	
Italy	X	X	X	X
Slovenia	X			X
Portugal	X	X		X
Spain	X	X	X	X
Cyprus			X	X

Only 6 Member States (MS) participated actively in the intercalibration providing data:
Cyprus, France, Greece, Italy, Portugal, and Spain.
This report includes comparison and harmonization of national boundaries from all these MS. Results were produced in two phases: April 2007 (boundaries for R-M1+M2+M4) and June 2007 (boundaries for R-M5 and for all types).

Malta participated in some meetings and wishes to be directly involved in future stages of the IC, if possible, but not with benthic invertebrates. According to the current knowledge on the streams of Malta, they are very peculiar, with an ephemeral character and unstructured biota. Apparently invertebrate assemblages are very poor and quite different from the Mediterranean rivers included in the IC. For this reason, invertebrates are not considered as a suitable biological quality element for Maltese streams.

The Med GIG has followed for Rivers the hybrid Option 2 described in the ECOSTAT Boundary Setting Protocol (IC process guidance, Annex III). Nonetheless, within the MedGIG, three countries used the full Option 1 (i.e. countries adopted as 'official National method' the same method used as the Common Index (i.e. ICMi) that was selected for the GIG intercalibration).

The GIG understands that MS methods differ in compliance and state of development in relation to WFD normative definitions. The MedGIG therefore agreed on the construction of a common metric (Intercalibration Common Metric index (ICMi)) which is intrinsically compliant with the normative definitions so that MS data can be converted to ICMi.

For national methods, the interpretation of the WFD normative definitions concerning good and moderate status within the GIG therefore relies on:

- agreeing how to quantify taxonomic composition, abundance, disturbance sensitive taxa, diversity and major taxonomic groups. For macro-invertebrates, this has been done by defining an "Intercalibration Common Metric index" (ICMi) as described in Buffagni *et al.* (2006).
- agreeing on what constitutes a slight and a moderate deviation from reference conditions. Because the normative definitions do not give any clarification of the meaning of 'slight' and 'moderate', and the lack of obvious break points or thresholds, interpretation of 'slight' and 'moderate' is rather arbitrary. Even so, member states have been able to justify the boundary setting protocol template in an objective way. The approach followed in the intercalibration process for macro-invertebrates has been to compare the results of each Member State's method translated into an Intercalibration Common Metric index (ICMi) combining a set of Water Framework Directive compliant metrics. The data screening procedure and acceptance criteria aimed to ensure that MS class boundaries would be comparable on the ICMi scale. Only MS that fulfilled all the agreed criteria were included in the calculation of the harmonization band.

Member States (MS) collated their data according to Common Intercalibration river types, identified and screened reference sites against pressure criteria agreed by the MedGIG and converted their national classification boundaries to values of the Intercalibration Common Metric index (ICMi) by regression.

Two particular ICMs were initially developed and tested by the MedGIG: a qualitative ICM (MedQual_ICM) and a quantitative one (MedQuant_ICM). However the index selected for the intercalibration of MedGIG rivers was the STAR_ICMi, for two main reasons: a) it provides a direct trans-GIG comparability i.e. with Central and Nordic GIGs; b) its performance against pressures is not very different from the Med ICMs.

The following six metrics are used in the STAR_ICMi: (Table 2.4.3):

- o ASPT - 2
- o Log10(sel_EPTD+1)
- o 1-GOLD
- o N-taxa

- o EPT taxa
- o Shannon-Wiener diversity index

The STAR_ICMi value is calculated by the weighted sum of all the metrics, according to the conceptual group to which they belong (Table 2.4.3), giving the same weight to each of the three groups. In the calculation, two normalization steps are performed, to re-scale single metrics before combining them and to re-adjust the ICMi values around 1 for Reference site samples (see Buffagni et al., 2005, 2006 for details). Both normalisations were performed by dividing the value observed for a sample by the Median value calculated for Reference sites. The two steps are essential to make ICMi values comparable across river types, MS and GIGs.

The STAR_ICMi fulfils the requirements of the WFD normative definitions because each criterion is addressed by 2 or 3 of the metrics combined in the ICMi (Table 2.4.3).

- The change in taxonomic composition and abundance is mainly evaluated through: Number of taxa, EPT taxa, and diversity (Shannon) index.
- The diversity is evaluated through Number of taxa and Shannon index.
- Sensitive taxa are mainly evaluated with ASPT (for organic + nutrient), abundance of selected EPTD (mainly accounting for hydro-morphological degradation).
- The 1-GOLD metric refers to quantitative changes in the balance of important functional groups.

Table 2.4.3. The Intercalibration Common Metrics (ICMs) used for the MedGIG Intercalibration (Buffagni et al., 2005; 2006).

Intercalibration Common Metrics (ICMs) studied					
Information type	Metric type	Metric name	Taxa considered in the metric	Literature reference	weight
Tolerance	Index	ASPT	Whole community (Family level)	e.g. Armitage *et al.*, 1983	0.333
Abundance/ Habitat	Abundance	Log_{10} (Sel_EPTD +1)	Log(sum of Heptageniidae, Ephemeridae, Leptophlebiidae, Brachycentridae, Goeridae, Polycentropodidae, Limnephilidae, Odontoceridae, Dolichopodidae, Stratyomidae, Dixidae, Empididae, Athericidae & Nemouridae)	Buffagni *et al.*, 2004; Buffagni & Erba, 2004	0.266
	Abundance	1-GOLD	1 - (relative abundance of Gastropoda, Oligochaeta and Diptera)	Pinto *et al.*, 2004	0.067
Richness and Diversity	Taxa number	Total number of Families	Sum of all Families present at the site	e.g. Ofenböch *et al.*, 2004	0.167
	Taxa number	number of EPT Families	Sum of Ephemeroptera, Plecoptera and Trichoptera taxa	e.g. Ofenboch et al., 2004; Böhmer et al., 2004.	0.083
	Diversity index	Shannon-Wiener diversity index	$D_{S-W} = -\sum_{i=1}^{s}\left(\frac{n_i}{A}\right)\cdot\ln\left(\frac{n_i}{A}\right)$	e.g. Hering *et al.*, 2004; Böhmer *et al.*, 2004.	0.083

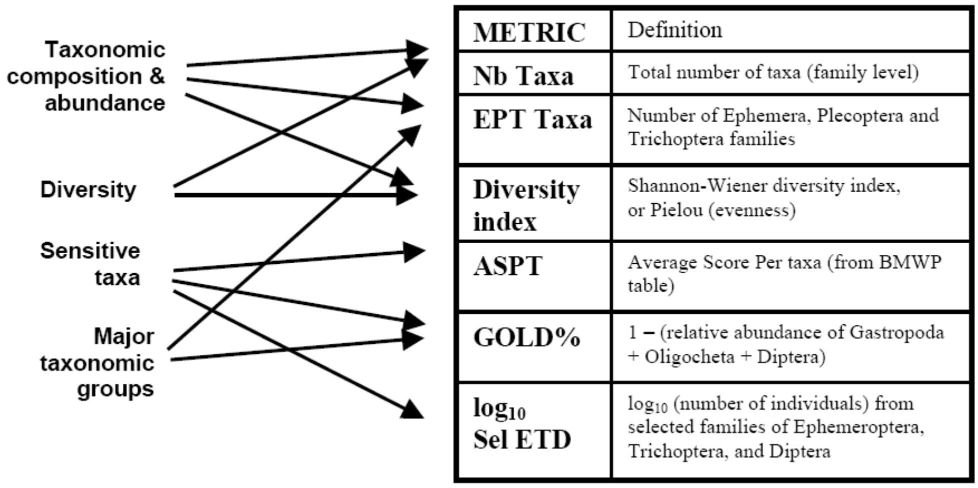


Figure 2.4.1: Each WFD criterion for the normative definition of 'good status' is addressed by more than one qualitative or quantitative metric in the ICMi.

Most metrics respond to general degradation (or combined stressors) (see Table 2.4.4).

Table 2.4.4: Each metric in the ICMi respond to a given stressor or to general degradation; the whole range of possible stressors is taken into account.

Metrics	Organic+Nutrients	Hydro-Morphology	Toxics	General
Total Nb taxa	X	X	X	XX
EPT taxa	XX	(x)	(x)	XX
Diversity index	X		X	X
ASPT	XXX		(x)	
1-GOLD	X			
Log Sel.ETD	X	XX		XX

An overall introductory analysis of the ICM and ICMi is reported among the results of the STAR project (Buffagni et al., 2005).

The response of the STAR_ICMi vs. general and/or specific pressure indicators can be found in the Annexes provided by MS to the CBGIG and in specific scientific literature.

<u>The Comparison Approach</u>
For the comparison, each MS provided a dataset with:

- raw macro-invertebrate data,
- MS EQRs (Ecological Quality Ratio of national assessment method and metrics)
- MS class boundaries
- Formula to derive the STAR_ICMi value from the National method boundary (when the official method was different from the ICMi).

Instructions were issued to the GIG to explain the procedure for converting national class boundaries to ICMi values.
For each member state, the EQRs from the national assessment method were correlated with the corresponding EQRs from the ICMi. A regression was performed and the regression formula and r^2 value calculated. National boundary values were transformed into ICMi – EQR values using the regression formula. The transformation of national boundaries (MS EQR values) to ICMi-EQR values allows MS boundaries to be compared with the boundaries of other MS on a common scale.
Following the screening of MS datasets (explained below), all calculations were re-run and the comparison was carried out centrally both by CNR and CEMAGREF.

Main checking steps:

- Official methods classifications were checked for WFD compliancy (→ based on general description of the classification system provided by MS).
- Checking of the criteria used to accept Reference sites.
- Checking calculations of the ICMi values and normalization options used.
- Checking regression between EQR MS_value and EQR ICMi (including R^2, regression equation).
- Screening for sources of natural variability (e.g. when boundaries are too low).

<u>The Harmonization Approach</u>

A first comparison was done for each IC type. Then, as the results for the different types were very similar, a GIG mean boundary value on the ICMi scale was calculated by averaging the different MS boundary values from all Intercalibration river types combined. Due to the low number of MS involved per type and the similar boundaries among types, the combination of boundary values from all types was preferred as a more robust definition of the harmonized MedGIG boundary.

In this first phase, only the R-M5 type was excluded from the averaging, due to its particular character (i.e. temporary rivers with higher natural variability).
The option considered to derive the harmonization band (i.e. the range of acceptability for MS boundaries to lie within) was a 5% band (one quarter of class width). This option is identical to the one used by the CBGIG and Alpine GIG.

Only data from MS that provided the required information to allow the GIG checking (see steps reported above) were considered for the calculation of the MedGIG mean boundary values (for High/Good and Good/Moderate boundaries). The most important requirement was that Reference criteria used to accept a Reference site at the MS level are fully in agreement with GIG requirements.

The following outcomes of the comparison are possible:
4. MS boundary lies within the harmonization band or is higher than its upper limit.
   - No action required.
5. MS boundary lies outwith the harmonization band.
   - MS accepts the GIG mean boundary and adjusts MS boundary to fall within the GIG harmonization band.

- MS justifies why it does not accept the GIG mean boundary. In this case, MS should provide scientific or technical reasons explaining why their boundaries differ from the GIG boundary (e.g. due to typological differences).

2.4.2 National methods that were intercalibrated

Table 2.4.5. National classification methods included in MedGIG Invertebrate Intercalibration.

MS	Method
Cyprus	STAR Intercalibration Common Metric Index (STAR_ICMi) (Buffagni et al., 2005 and Water Development Department, 2008)
France	French WFD classification Indice Biologique Global Normalisé (IBGN) - norm AFNOR NF T 90 350 (1992) and circular MEDD/DE 05 n°14 (July 05)
Greece	STAR Intercalibration Common Metric Index (STAR_ICMi) (Buffagni et al., 2007)
Italy	STAR Intercalibration Common Metric Index (STAR_ICMi), type specific (Buffagni et al., 2007 and IRSA-CNR, Notoziario dei Metodi Analitici, Marzo 2007)
Portugal	IPtI Invertebrate Portuguese Index: $IPtI_N$ and $IPtI_S$
Spain	IBMWP (Alba-Tercedor & Sánchez-Ortega, 1988, Alba-Tercedor et al., 2004)

National methods from MS are indicated in Table 2.4.5. All method involved in the IC were presented by the MS as official. "Official" means those methods will be used in the monitoring programs to assess the ecological status.
For benthic invertebrate fauna, each country in the MedGIG compiled fact sheets describing their national methodology and the criteria used for boundary setting at a national level (Please refer to Annex 2.4.1). All MS provided this information.

2.4.3 Reference conditions and class boundary setting

Reference conditions were derived from data observed in reference sites; these sites were chosen by MS on the basis of the procedure and criteria agreed in the REFCOND guidance. A detailed list of criteria was agreed by the MedGIG; this list is similar to that used by the Central Baltic and Alpine GIGs, with minor adaptations to the Mediterranean context. MS were asked to screen selected reference sites against agreed catchment land cover limits and other pressure criteria, and when proposed reference sites exceeded the agreed reference limits for urban land cover, a validation with physico-chemical parameters thresholds at the site scale was necessary. MS were also asked to complete a check list to indicate which of the GIG defined reference criteria were used for the screening exercise and what sources of information were available to the MS for this process (see annex 2.4.2). Table 2.4.6 indicates the numbers of reference sites and samples for each type and MS.
The GIG checking process ensured that MS adhered to the correct screening procedure using the information provided in the check list. Reference sites have been identified for each MS and for each river type. The table below shows the number of reference sites/samples identified for the different MS and common IC types.

Table 2.4.6. Number of reference sites and samples selected by MS according to the GIG defined criteria for the different IC types in which they participate (- : MS not participating to the IC type).

MS	RM1		RM2		RM4		RM5	
	Sites	Samples	Sites	Samples	Sites	Samples	Sites	Samples
CY	-	-	-	-	8	16	-	-
FR	10	16	-	-	-	-	-	-
GR	4	4	3	3	3	3	-	-
IT	3	8	2	7	3	6	3	5
PT	33	33	31	31	-	-	18	18
SP	30	42	-	-	34	51	25	42
TOTAL	80	103	36	41	48	76	46	65

Due to the lack of data for newly selected reference sites, the occasional use of expert judgment was an accepted means of validating reference sites especially in interpreting the use of different forms of chemical determinants throughout the GIG.
A central pressures database was not available to verify reference sites. Access to pressure data from reference sites may be requested and provided by individual MS.

Table 2.4.7 presents the number of sites and samples for each IC type which were used in the intercalibration.

Table 2.4.7. Number of sites and samples per Member State used in the analysis for the invertebrate common intercalibration types.

MS	R-M1		R-M2		R-M4		R-M5	
	Sites	Samples	Sites	Samples	Sites	Samples	Sites	Samples
CY	-	-	-	-	29	60	-	-
FR	36	72	-	-	-	-	-	-
GR	17	17	15	15	12	12	-	-
IT	11	33	34	249	11	33	13	25
PT	62	62	68	68	-	-	29	29
SP	64	177	-	-	48	87	59	236
TOTAL	190	361	117	332	100	192	101	290

The Reference Value used to derive EQR is set at the median of the values observed in reference samples.

In general terms, the Mediterranean MS used the so-called REFCOND approach to derive class boundaries. According to this approach, the High/Good boundary is equated to the 25%ile value of Reference site samples. The Good/Moderate boundary is then derived by multiplying the HG boundary by 0.75.
Cyprus, Greece and Italy applied such an approach directly to the ICMi selected for the MedGIG comparison (i.e. the same as the National method).

France, Spain and Portugal applied similar approaches to the national method - direct REFCOND or with some adaptations due to the non-linear behaviour of the index (Spain) - and secondarily derived the ICMi value by regression.

The relationship between the IBMWP Spanish national method and the ICMi is not linear, and values must be transformed using non-linear regression. That is why EQR values for Spanish boundaries are lower than ICMi.

The relationship between the IBMWP Spanish national method and the ICMi is not linear. For this reason, the national EQR was first transformed into an inverse variable to linearize the relationship, and then a linear regression was derived between the transformed variable and the ICM. Although different of the direct regression used by the others MS, this procedure ensures that the relationship between the ICMi and the national EQR value is derived from a linear regression, and that only one ICM value can correspond to each national EQR value.

The details of the procedures and the minor adjustments adopted in few circumstances are described below and in the Annexes provided by MS.
The GIG required each country to compile a fact sheet describing their national methodology and the criteria used for H/G and G/M boundary setting at a national level in accordance with the WFD normative definitions (see Annex 2.4.1).

Table 2.4.8 presents the values of the High-Good and Good-Moderate boundaries for each country and river type as EQR values of the national classification systems.

As cited above, the MS participating to the WFD IC exercise have used the so-called REFCOND approach to derive class boundaries for their assessment methods. France has defined, for all its river types at the national level, G/M boundaries slightly higher than simply applying the REFCOND approach; these G/M boundaries values were validated in consideration of the results of the Alpine and Central GIGs. Cyprus and Greece have applied directly the REFCOND approach for all types, without any corrections. Italy has applied directly the REFCOND approach for all types but the R-M2, where the boundaries were equated to those obtained for the R-M4 to increase homogeneity among similar types. Portugal applied the REFCOND approach and adopted slightly higher boundaries for some types. Spain has used the National class boundaries tested for a long time and used in the Mediterranean Spanish rivers.

G/M boundary was checked for presence or absence of major taxonomic groups. Further details on the class boundaries setting can be found in the Annexes provided by MS.
The data underlying the setting of national boundaries for the H/G and G/M accordance to WFD Annex V normative definitions is held at the MS level. The MedGIG does not have a central database.

Table 2.4.8. EQR National values for High-Good and Good-Moderate boundaries presented by each MS for the IC types.

Type and country	**Ecological Quality Ratios for the national classification systems**	
	High-Good boundary	**Good-Moderate boundary**
R-M1		

France	0.94	0.81
Greece	0.95	0.71
Italy	0.97	0.72
Portugal	0.92	0.69
Spain	0.78	0.48
R-M2		
Greece	0.94	0.71
Italy	0.94	0.70
Portugal	0.87	0.66
R-M4		
Cyprus	0.97	0.73
Greece	0.96	0.72
Italy	0.94	0.70
Spain	0.83	0.51
R-M5		
Italy	0.97	0.73
Portugal	0.98	0.72
Spain	0.91	0.55

2.4.4 Results of the comparison

The results of the comparison are presented separately for each river type. High-Good and Good-Moderate boundary values are here presented on the ICMi scale. For France Portugal and Spain, these values were derived through regression from the National EQR values. Table 2.4.9 presents the regression equations which translate the national methods into ICMi for each MS. The HG and GM boundaries translated into ICMi are presented in Table 2.4.10. The only non-linear regression was used for the index of Spain. The inverse regression is considered to be the best description for the IBMWP-ICMi relationship and poses no problems of translation from one index to the other.

Table 2.4.9: Regression equations used to translate the national methods into ICMi.

Member State	Translation from National Method into ICMi	
Cyprus	ICMi = STAR_ICMi Index	R^2= 1.00
France - R-M1	ICMi = 0.9322 x IBGN - 0.0004	R^2= 0.89
Greece	ICMi = STAR_ICMi Index	R^2= 1.00
Italy	ICMi = STAR_ICMi Index	R^2= 1.00
Portugal - R-M1	ICMi = 0.9555 x $IPtI_N$ + 0.0108	R^2= 0.95
Portugal - R-M2	ICMi = 1.014 x $IPtI_N$ + 0.0049	R^2= 0.96
Portugal - R-M5	ICMi = 0.8488 x $IPtI_S$ + 0.0475	R^2= 0.87
Spain – R-M1	ICMi = 2 - (1.94 / ("IBMWP EQR"+1))	R^2 = 0.88
Spain – R-M4	ICMi = 1.76 - (1.58 / ("IBMWP EQR"+1))	R^2 = 0.63
Spain – R-M5	ICMi = 1.93 - (1.87 / ("IBMWP EQR"+1))	R^2 = 0.83

Table 2.4.10: HG and GM boundary values for national methods converted by regression into ICMi values.

	Cyprus	France	Greece	Italy	Portugal	Spain
R-M1						
HG		0.88	0.95	0.97	0.89	0.91
GM		0.76	0.71	0.72	0.67	0.69
R-M2						
HG			0.94	0.94	0.89	
GM			0.71	0.70	0.67	
R-M4						
HG	0.97		0.96	0.94		0.90
GM	0.73		0.72	0.70		0.71
R-M5						
HG				0.97	0.88	0.95
GM				0.73	0.66	0.73

2.4.5 Results of the harmonization

Because the number of participating MS is relatively small and boundaries from the different IC types are similar, it was considered more robust to calculate a common average of the boundaries provided by each MS for each type. The approach of having a common boundary was also followed by the CBGIG.

The HG and GM boundary values for each type and MS fulfilling all MedGIG requirements were used to define mean boundary values and an acceptability band. This band is calculated by removing/adding 0.05 to the mean boundary value. The critical value is obviously the lower limit of this band.

Table 2.4.11: Mean values (ICM) for High-Good and Good-Moderate boundaries and the minimum acceptable values (lower limits of the acceptability bands) for R-M1+M2+M4, R-M5, and all types.

	Mean	Minimum acceptable values
R-M1+M2+M4		
H/G	**0.94**	**0.89**
G/M	**0.72**	**0.67**
R-M5		
H/G	**0.93**	**0.88**
G/M	**0.71**	**0.66**
All IC types		
H/G	**0.93**	**0.88**
G/M	**0.71**	**0.66**

The boundaries for M1+M2+M4 (Table 2.4.11) were derived during the previously mentioned phase 1; results were presented to and approved by ECOSTAT in the April 2007 Meeting and included in the Decision Annex. Countries whose data were used to derive the averaged values were Cyprus, France, Greece, and Italy. In that Meeting it was decided to keep M5 out of this group because of its particular character, i.e. temporary streams with higher natural variability which could increase the range of reference values and possibly decrease all boundaries. For this reason, in phase 2 (June 2007) mean values and bands for M5 boundaries were derived. For R-M5 mean boundary values and bands only Italy, Portugal and Spain provided data. These boundaries for M5 are actually very similar to the M1+M2+M4 ones (Table 2.4.11). This fact questions the separation of this river type from the other IC types. For this reason, the GIG decided to derive and present boundaries also for a single group with all types (Table 2.4.11). In practical terms this approach means that the minimum acceptable boundaries are set for all river types (even for those not included in the Mediterranean Rivers IC) as the same values, with the possible exception of large rivers.

It must be emphasised that a possible way to deal with the natural temporal variability of M5 (or others if it is the case) is to derive different reference conditions for different sets of hydrological years (e.g. dry years) and use them according to the characteristics of the data subsets. With this approach, EQRs are calculated with the appropriate reference values (e.g., *dry* reference value for *dry* year samples) and, because of the standardisation properties of EQR, the problems related to inter-annual variability are solved or at least strongly reduced.

Boundary ICM values for all MS and types are presented in Table 2.4.12. For France, HG boundary is marginally lower than the acceptability band for M1+M2+M4 but GM boundary is the highest among all MS and types. No other MS presents boundaries lower than the minimum acceptable values, for all the combining options (M1+M2+M3 and M5, or all types grouped), no harmonization being required.

Table 2.4.12. ICMi values for the High-Good and Good-Moderate boundaries, lower limits of the acceptability band for each IC type, and the need for harmonization. M1+M2+M4 boundaries were approved in phase 1 (april2007).

MS	IC type	HG boundary	Harmonization required	GM boundary	Harmonization required
CY	RM4	0.97	No	0.73	No
FR	RM1	*0.88*	slightly lower	0.76	No
GR	RM1	0.95	No	0.71	No
	RM2	0.94	No	0.71	No
	RM4	0.96	No	0.72	No
IT	RM1	0.97	No	0.72	No
	RM2	0.94	No	0.70	No
	RM4	0.94	No	0.70	No
	RM5	0.97	No	0.73	No
PT	RM1	0.89	No	0.67	No
	RM2	0.89	No	0.67	No
	RM5	0.88	No	0.66	No
SP	RM1	0.91	No	0.69	No
	RM4	0.90	No	0.71	No
	RM5	0.95	No	0.73	No

lower band limits				
M1+M2+M4	**0.89**		**0.67**	
M5	**0.88**		**0.66**	
All IC types	**0.88**		**0.66**	

List of MedGIG Annexes

Annex 2.4.1 – National methods: Class boundary setting procedure
Annex 2.4.2 – Reference criteria

2.5 Eastern Continental GIG

2.5.1 Intercalibration approach

<u>Common intercalibration types and countries sharing the types</u>

Within the Eastern Continental GIG GIG three common intercalibration types were defined (Table 2.5.1), that are shared by 18 countries (Table 2.5.2).

Table 2.5.1 Central-Baltic rivers common intercalibration types

Type	River characterisation	Ecoregion	Catchment (km^2)	Altitude (m)	Geology	Substrate
R-E1	Carpathians: small to medium, mid-altitude	10	10 – 1000	500 – 800	siliceous	gravel and boulder
R-E2	Plains: medium-sized, lowland	11 and 12	100 – 1000	< 200	mixed	sand and silt
R-E4	Plains: medium-sized, mid-altitude	11 and 12	100 – 1000	200 – 500	mixed	sand and gravel

Table 2.5.2 Countries sharing the Eastern Continental common intercalibration types

	R-E1	R-E2	R-E4
Czech Republic	X	X	X
Romania	X	X	
Hungary	X	X	X
Slovakia	X	X	X
Slovenia			X
Austria			X

Within the intercalibration exercise the definition of reference conditions is of major importance for the comparison of national quality assessment methods. In this regard, two problems became obvious in the EC GIG:

- Either existing reference site are not available (esp. for lowland river types) or
- reference criteria to screen for existing reference sites differ among countries.

The EC GIG agreed to follow an *alternative approach* to resolve these issues by defining IC type specific, harmonised quality criteria. In general, the GIG set common high-good (R-E1) respectively good-moderate (R-E2-4 quality class boundaries for the national biological assessment methods using existing data assembled within the EC GIG intercalibration exercise. The main idea of using this approach is to overcome the difficulties of lacking (near-natural) references by defining *alternative references*. . The EC GIG countries commonly agreed on a specific level of impairment, which is acceptable for alternative references. The available data sets have been screened by defined threshold values of selected biotic and abiotic criteria.
This practical approach comprises two steps, that are explained in further detail below:

-A Harmonised definition of quality criteria/thresholds for the high and good ecological status

-B Class boundary setting based on 25th percentile value of common metrics using all sampling sites meeting the criteria defined in section A

A. Harmonised definition of quality criteria/thresholds for the high and good ecological status

Based on criteria for saprobiological quality - commonly agreed for monitoring purposes in the Danube River Basin - biological threshold values are derived using the common metric ASPT (Average Score Per Taxon). Sites with samples showing ASPT values above these thresholds (=better values) are screened by additional chemical, morphological and land use parameters. The set of sites complying with all criteria/thresholds are considered too be in a commonly agreed, ecologically high (R-1) respectively high and good status (R-E 2, 4).

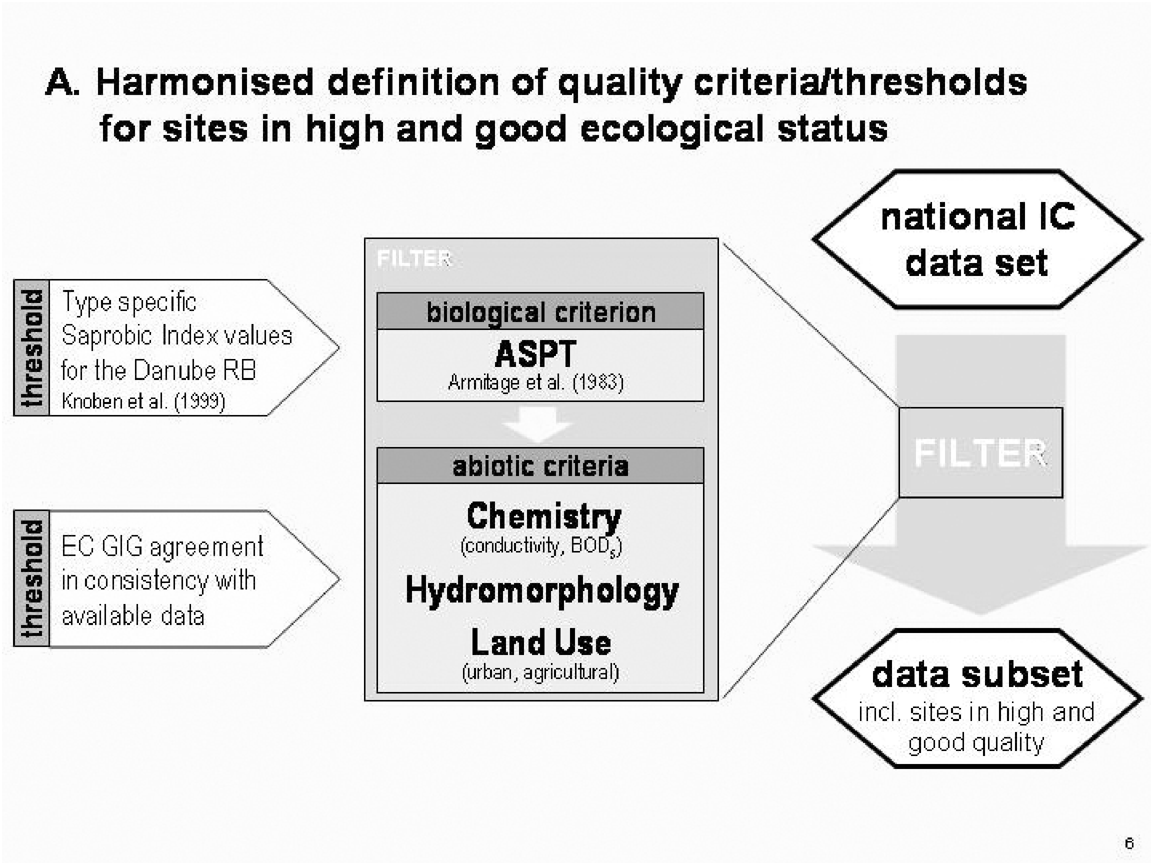


Figure 2.5.1: Harmonised definitios of quality criteria/thresholds for sites in high and good ecological status

B. Class boundary setting based on 25th percentile value of common metrics using all sampling sites meeting the criteria defined in section A

The ecological quality class boundaries are expressed on an ICMi-EC scale – Intercalibration Common Metric Index for the Easter Continental Region to comply with the normative definitions of the WFD. These boundaries are derived by selecting the 25th percentile values of each common metric from the set of sites in high respectively high and good status. By means of regression analysis the boundary values are translated into values of the national assessment method (= final result). annex 2.5.3 provides a more detailed description of this approach.

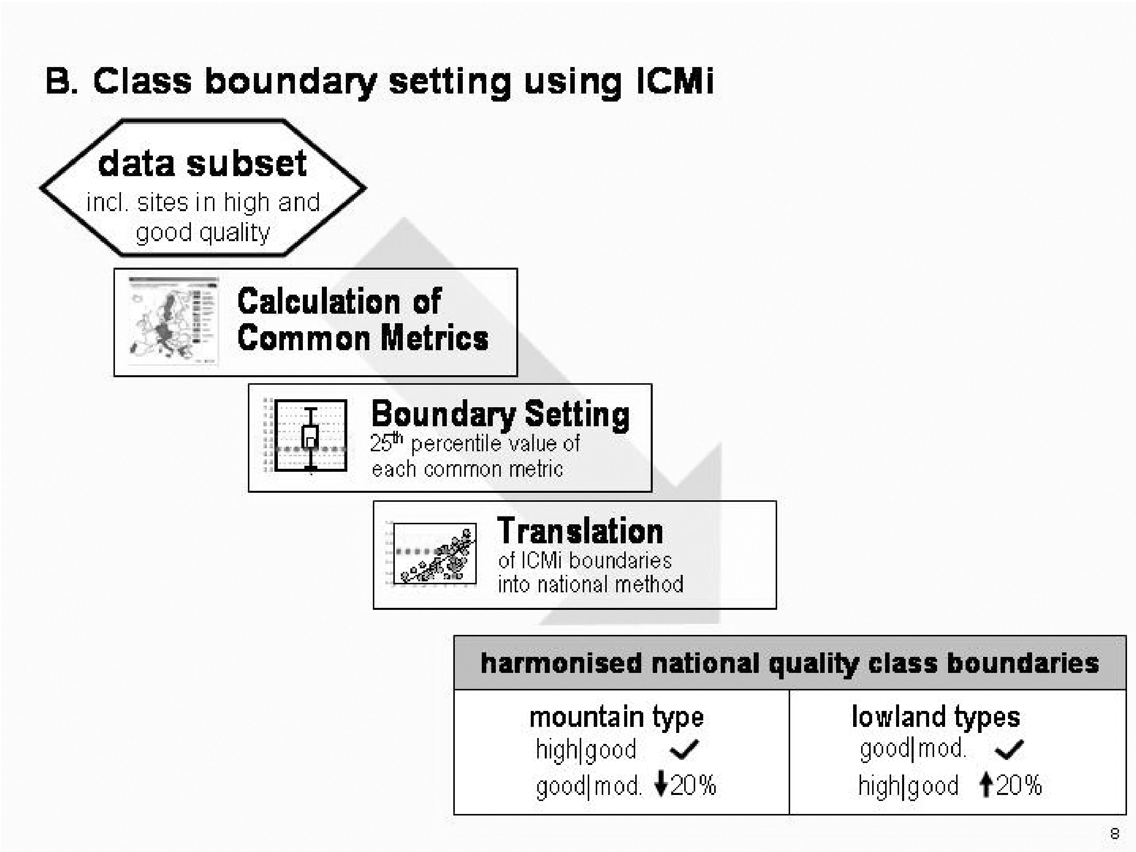


Figure 2.5.2: Class boundary setting using ICMi

R-E6 (Danube River):

Biological assessment of the Danube River on the basis of the benthic macroinvertebrate community is limited to the application of Saprobic Systems or Biotic Indices to evaluate the degree of organic water pollution. So far, the ecological quality/status of the Danube River using benthic macroinvertebrates is assessed by classification method, which are not WFD compliant . The development of WFD compliant methods for large rivers are an European wide challenge and are underway . Therefore, the intercalibration exercise performed for the Danube River (R-E6) focused on the comparison of national methods which have been used in regular water quality monitoring of the Danube River. Preliminary results of this intercalibration exercise are presented in Annex 2.5.4.

2.5.2 National methods that were intercalibrated

The natioanbl methods that were used in the Eastern Continental intercalibration exercise are summarised in Table 2.5.3.

Except for Austria and the Slovak Republic none of the other countries in the EC GIG hold biological assessment methods that are fully compliant with the requirements of the EU-WFD. WFD compliant methods are currently being developed in those countries. Therefore, intercalibration of EQR class boundary values was only fully completed for the methods of Austria and Slovak Republic. The results of the intercalibration exercise for these two assessment methods are have been successfully completed. However, the IC exercise was also performed for the non-WFD compliant methods. These preliminary results including intercalibrated boundary values are included in Annex C 2.5.3

o

Table 2.5.3 Respective national assessment methods used in the intercalibration exercise.

Country	Name	Category	WFD compliant
Austria[1]	Slovak System for Ecological River Status Assessment	Multimetric Index	yes
Slovak Republic[1]	Austrian System for Ecological River Status Assessment	Multimetric Index	yes
Czech Republic	Czech Saprobic Index following Zelinka & Marvan (1961)	Saprobic Index	no
Hungary	Hungarian Average Score Per Taxon	Biotic Index	no
Romania	Romanian Saprobic Index following Pantle & Buck (1955)	Saprobic Index	no
Bulgaria	Bulgarian Biotic Index for River Quality Assessment (Q-Scheme)	Biotic Index	no

[1] For the intercalibration of R-E6 the national Saprobic Indices were used instead of the methods listed in this table.

2.5.3 Reference conditions and class boundary setting

Reference sites were chosen by the GIG countries using the REFCOND guidance following the procedures and criteria agreed in the Central-Baltic GIG (see Chapter 2.1.3). A list of more detailed criteria and type-specific concentrations of key chemical parameters were agreed by the EC GIG. Countries were asked to screen selected reference sites against agreed chemical, hydromorphological and catchment land-use threshold limits. Countries were also asked to complete a check list to indicate which reference criteria - defined in the GIG - were used for the screening exercise.

This process showed that (near) natural reference sites could only be described for the common stream type R-E1 (Carpathian rivers).

Table 2.5.4: Number of R-E1 reference sites identified for the different countries:

Country	Number of reference sites	Number of samples at reference sites
CZ	3	10
HU	16	41
RO	20	42
SK	22	48

For all other Eastern Continental IC types reference sites are currently not available, and the alternative approach based on harmonised definition of quality criteria and thresholds for the high and good status as described in Chapter 2.5.1 was used for all common types (including R-E1). Comparison of median values for common metrics of “true” R-E1 reference sites with 25th percentile values from sites of the R-E1 data subsets revealed that “true” references and sites in high ecological status (defined by the common criteria) cover similar biological quality status.

The EC GIG realized that methods used by the GIG countries differ in compliance and state of development in relation to WFD normative definitions. The GIG therefore agreed on the design of a common metric (Intercalibration Common Metric index (ICMi)) which is intrinsically compliant with the normative definitions so that the countries' diverse data can be converted to common scale (=ICMi scale). The ICMi-EC developed for the Eastern Continental GIG consists of four common metrics combined to a common multimetric index by using the average of normalised metric values.

Table 2.5.5: Common metrics, WFD indicative parameters addressed and pressures indicated (based on pressure analysis of EC GIG dataset):

Common Metric	WFD indicative parameter	Indicated Pressure
Average Score Per Taxon (ASPT)	Sensitive Taxa	Organic Pollution, General Degradation
Austrian Structure Index (family level)	Sensitive Taxa	Structural and General Degradation
Total Number of Families	Taxonomic composition, diversity	General Degradation
[%] EPT Abundance	Taxonomic composition, abundance, major taxonomic groups	Organic Pollution, Structural and General Degradation

The boundary setting process consisted of three steps (see also Figure 2.5.2), that are explained in further detail below: (a) Setting of class boundaries on the common metric scale, (b) Translation of boundary values into national classification schemes and (c) Definition of national class boundaries not specified by data subsets.

(a) Setting of class boundaries on the common metric scale

For each national data subset comprising sampling sites in selected quality status (see section 2.5.1 and Annex 2.5.3) the single metrics of the ICMi were calculated. This procedure resulted in a range of common metric values per country and intercalibration stream type, that were indicative of the macrozoobenthic fauna at sites in at least high (Carpathian rivers: R-E1) or good (rivers of the Plains: R-E2, R-E3, R-E4) quality status.

The lower boundaries of the quality status specified by the data subset were set by the 25^{th} percentile value of each single metric. Setting of validated boundaries per common type was only applied if national data subsets contained at least 8 samples taken at minimum 2 sites.

(b) Translation of boundary values into national classification schemes

Single common metrics were normalised by the boundary values defined above, so that the combined ICMi value of "1" represented the high-good boundary for the Carpathian rivers (R-E1) and the good-moderate boundary for the rivers in the Plains (R-E2, R-E3, R-E4), respectively. The translation method followed the approach described by Birk & Hering (2006): Based on the complete national datasets regression analysis was performed per country and common intercalibration type using ICMi as predictor variable and the national indices as target variables. By means of the regression equation the national index value corresponding to an ICMi value of "1" was modelled, resulting in harmonised boundaries on the national index scale.

(c) Definition of national class boundaries not specified by data subsets

The above described procedure resulted in harmonisation of either the upper (R-E1) or lower (R-E2, R-E3, R-E4) biological quality limit of the good status. Respective class boundaries not specified by the data subsets were defined in the last step of the setting process. Basis for this setting was the premise that a 20 percent deviation in biological quality generally represents a shift in status class. Thus, the good-moderate boundaries of the national classification schemes intercalibrated for the Carpathian rivers (R-E1) were defined by decreasing the harmonised values by 0.2 EQR units (WFD compliant methods) or 20 percent (absolute values of non-WFD compliant methods). In reverse, the high-good boundaries of the national schemes intercalibrated for the rivers in the Plains (R-E2, R-E3, R-E4) were set by increasing the harmonised values by 0.2 EQR units (WFD compliant methods) or 20 percent (absolute values of non-WFD compliant methods).

Data will be made available on DANUBIS, the database of the International Commission for the Protection of the Danube River (ICPDR), which is coordinating the work of the EC GIG. The database is accessible via the Internet.

2.5.4 Results of the comparison

In the Eastern Continenal GIG harmonised class boundaries were defined within a GIG-wide agreed framework. The GIG decided that national class boundaries will be adjusted according to the results of the intercalibration analysis. Therefore, national class boundaries were not compared between countries but against the boundary values obtained in the intercalibration analysis. These boundaries are presented in chapter 2.5.5.

Data will be made available on DANUBIS, the database of the ICPDR accessible via the Internet.

2.5.5 Results of the harmonisation – Boundary EQR values

Table 2.5.6 below presents the results of the EC GIG intercalibration exercise for the WFD compliant national assessment methods of Austria and Slovak Republic regarding the common intercalibration types R-E1, R-E2 and R-E4. Results of further country/type combinations (based on non-WFD compliant methods) are described in Annex 2.5.3.

Table 2.5.6 Boundary values and confidence limits for RE1 – R-E4 river types

common stream type	country	boundary type	boundary value	confidence limit[13]	
				lower	upper
R-E1	Slovak Republic	high-good	0,74	0,69	0,79
		good-moderate	0,54	0,49	0,59
R-E2	Slovak Republic	high-good	0,74	0,69	0,79
		good-moderate	0,54	0,49	0,59
R-E4	Slovak Republic	high-good	0,72	0,67	0,77

[13] Confidence intervals are specified as the 5 percent deviation from the respective boundary value.

		good-moderate	0,52	0,47	0,57
R-E4	Austria	high-good	0,79	0,74	0,84
		good-moderate	0,59	0,54	0,64

List of EC GIG Annexes

Annex 2.5.1 – National methods
Annex 2.5.2 – Reference criteria
Annex 2.5.3 – Class boundary setting
Annex 2.5.4 – Intercalibration Danube River

3 Discussion

3.1 Comparability between GIGs

For the river macroinvertebrate intercalibration there is a high degree of comparability between the GIGs. The intercalibration procedure is in all cases based on a intercalibration common metric index (ICMi). National classification scales are compared by translating them into a common ICMi scale by regression, and ensuring that all boundaries fall within agreed harmonisation band. The advantage of this methodology is that each country uses its own dataset to calculate the ICMi boundary values, and the tricky problem of merging raw data from different sources is avoided, allowing to compare classification boundaries of countries employing different sampling and sorting methods.
This intercalibration methodology relies on the identification of reference sites within each of the national data sets. The values of the classification methods for the reference sites are used to normalise each data set, making it possible to directly compare the resulting EQR values on the ICMi scale (see Buffagni et al., 2005).
As a consequence, the comparison between class boundaries is only valid if it is possible to find sufficiently comparable reference sites. This is very difficult to accomplish for certain river types, e.g. very large rivers. For this rivers another approach is required, e.g. by basing the comparison not on undisturbed reference sites, but on sites with an equal level of anthropogenic disturbance. This approach was already appleid in the Eastern Continental GIG, and may be the right way forward in continued intercalibration activities aimed at large rivers.

3.2 Open issues and need for further work for river macroinvertebrates

3.2.1 Central-Baltic GIG

Open issues
The following open issues have been be identified:

- Not all countries have been able to provide reference sites according to the agreed criteria, limiting the validity of the comparison produced in this intercalibration exercise for those countries; other approaches (e.g. bilateral comparisons with neighbouring countries that were involved in the comparison) could allow them to adjust their boundaries (or modify their present intercalibrated boundaries, if a country is still modifying their national system) according to the results of the intercalibration exercise.
- Not all countries have fully WFD compatible assessment methods in place at this time, and will need to set their boundaries taking into account the results of the intercalibration exercise as soon as they are ready.
- River types for which the CB GIG believes there are no reference sites available in Europe (e.g. very large rivers) have been excluded thus far; another intercalibration approach is required for those types.
- The CB GIG Steering Group is of the opinion that further coordinated work is required to establish the extent and nature of reference conditions in the EU across all river types and all geographic regions.

Rivers Types that were not intercalibrated
Some river types that fall into the CB GIG list of common river types were not included in the first intercalibration exercise for various reasons such as lack of available data across MSs. In this case, the SG recommended that the boundaries defined under the combined regression option should apply until further data becomes available and a review of intercalibration can be carried out. MSs must provide reference data for these other river types. This includes all river types falling within the extremes of the type factor values (e.g. medium or large lowland rivers with alkalinity < 0,4 meq/l), however boundary values may not be applicable in all situations e.g. very large rivers.

Data Availability
All data underlying the comparison of class boundaries is stored centrally in a restricted access folder on the EEWAI CIRCA website:

- Separate Excel files are compiled by each MS for each of the river types applicable to them, containing for each site raw macro-invertebrate data (family level), the national assessment metric, the classification according to the national metric, and whether or not it is a reference site. The Excel files also includes the results of the calculations – a regression of the MS assessment vs. the ICMi, and the translation of national boundary values to ICMi values.
- A single Excel file contains all CB GIG results for each of the >14,000 data points:
  - Site ID
  - Common intercalibration river type
  - MS boundary value
  - ICMi boundary value
  - MS status class (H,G,M,P,B)
  - Whether or not a site is a reference site according to the GIG criteria.

This data will be adequate to verify the comparison procedure but not to verify reference status. Most MSs have expressed their willingness to make the IC data publicly available for independent validation. Accompanying qualifications from MSs regarding the public availability of IC data vary from requiring acknowledgment of the source of the information, to a request that the data is only made available subject to consent from the relevant MS.

There has been unprecedented cooperation between MSs in the GIG to provide practical and convenient solutions to intercalibration. To illustrate this point, the combined number of samples included in the analysis of macro-invertebrate intercalibration totals 14,835 (Table 2.5) across participating MSs!

3.2.2 Northern GIG

The approach followed in this phase of the intercalibration process can be extrapolated to new river types not considered to date and to other countries. for the applicability of the current results to the river type R-N5 needs to be verified. Some MS are still in the process of refining their methodology for macroinvertebrate assessment but the intercalibration process used here can be used to ensure that ongoing compatibility between status classifications in different river types and in different MS can be maintained.

The use of the ICMi metrics has proved successful in comparing widely differing ecoregions in the five countries involved due to the incorporation of a normalisation step into the methodology – i.e. dividing by the appropriate sub-metrics for the reference conditions for the river type in question. While some differences appear in the three river types compared in all cases the H/G and G/M boundaries lie within the ±5 tolerance limits. The differences in actual boundary values are felt to be largely data-dependent rather than being due to real inherent differences. The lack of large numbers of moderate, poor and bad status sites has undoubtedly biased the results towards the good and high end of the scale. This was, however, to be expected in view of the generally high quality of waters in Scandinavian countries in particular. For this reason an all-types approach to the intercalibration is probably well justified in the NGIG case. Attempts to harmonise more tightly based on, for example, 95% confidence limits would be likely to lead to spurious results and a false sense of accuracy. Larger numbers of samples across the full range of quality are required before attempting a more detailed comparison. In the light of the huge natural variation encountered within the NGIG rivers, however, the intercalibration exercise must be seen as highly successful. The overall agreement on what constitute reference conditions together with discussions concerning the 'real-world' meaning of the five WFD status categories (e.g. as included in earlier milestone reports), plus some informal intercalibration between the UK system and the Irish system within Ecoregion 17 leads to the conclusion that the MS assessments of where the H/G and G/M boundaries lie is very similar.

The wide variation in ecoregional type within the NGIG is indicated by the following:

1) Varying geographical conditions inside the N GIG area.
a) In Finland and in eastern parts of Middle and Northern Sweden the bedrock is very old, whereas in Norway the bedrock is mostly younger. This has implications for the water quality.
b) High relief in the western part, low in the eastern part. This difference influences significantly the conditions in surface waters.

c) Overall retention of water in river basins is longer in the eastern than in the western parts of the NGIG area.
d) Coverage of mires is significant in the eastern part, especially to the east and north of the Gulf of Bothnia in Finland and in parts of Northern Sweden. Also occurrence of clay soils is relevant in some parts of Finland for natural conditions of rivers.
e) Ireland in Ecoregion 17 has a limited fauna and flora due to island biogeographical reasons and the relatively recent glaciation (12,000 years). This has resulted in macroinvertebrate communities that have a naturally low number of species. This also applies to Ireland's fish communities. Also in Norway, especially in the Western region, the fauna and flora has a naturally low number of taxa due to recent glaciations (8000 years) and immigration barriers.While the ICMi metric does use reference conditions it is believed that the low number of taxa in Irish and (Western) Norwegian ecosystems, especially in more acidic areas, may result in non-linear response at high and good status.

2) Climate
Among other issues the duration of winter varies inside the N GIG. In all the Scandinavian countries the north-south climate gradient is rather strict, which might have influence e.g. on the use of the all-types approach for specific national types.

3) Monitoring practice
Although biological elements have been used in all N GIG countries for a long time, the use of these elements and especially specific status assessment methods has not been so widespread in other N GIG countries as the UK and Ireland. This has had influence on the amount of data available for the intercalibration.

3.2.3 Alpine GIG

The Alpine GIG has not identified any other open issues and considers the invertebrate intercalibration completed.

3.2.4 Mediterranean GIUG

For the river types that the GIG has not intercalibrated, but which fall into MedGIG, it is recommended that the boundaries defined under the combined mean value should apply until further data becomes available and a review of intercalibration can be carried out. MS must provide reference data for these other types. Boundary values may not be applicable in all situations, e.g. large rivers.
Large rivers are considered to be a priority for future work. Most countries lack data on this type of rivers and no present reference sites are available.

3.2.5 Eastern Continental GIG

The intercalibration exercise performed within the Eastern Continental GIG and co-ordinated by the ICPDR PS addresses exclusively the biological quality element (BQE) macroinvertebrates. This results from the fact that data availability for the other BQEs within the Danube River Basin is currently scarce. However, as Austria and Slovakia are already using WFD compliant methods and do have data on macrophytes as well as on phytobenthos available, the intercalibration of these two parameters will be performed and additionally reported by June 2007.

The analysis of the EC GIG are primarily based on data which have not been assessed with WFD compliant methods - only AT and SK are currently using WFD compliant methods whereas the other countries are developing their methods. Due to this fact most of the analysis' results are part of this report's Annex 2.5.3. As soon as data – based on WFD compliant methods - will be available the analysis will be improved. This improvement will very likely be performed by the end of 2008 and can further be included in the updated version of the Technical IC Report (JRC) by 2011 (see Draft Mandate of Working Group A/ECOSTAT).

Regarding the continuation of the intercalibration exercise within the EC GIG the following issues will be addressed:

- Filling of existing data gaps (see Annex 2.4.3) by June 2007.
- Intercalibration using the other BQEs: Improvement related to information on other BQEs is expected during the upcoming years. Increasing data sets will be available from assessments of the WFD compliant monitoring networks (by mid 2008) and should be used for the improvement of the intercalibration exercise results.
  - The intecalibration between AT and SK regarding the BQEs macrophytes and phytobenthos will be performed by and reported by February respectively May 2007.
- Improvement of the intercalibration analysis for the types RE-2, 3 and 4: Currently an adapted approach had to be chosen due to the lack of reference sites. Further, not all countries are using WFD compliant sampling/assessment methods. Expected results from the WFD compliant monitoring networks will be integrated.
- Intercalibration of type RE-6 (Danube River): The results regarding the intercalibration of the Danube River (Type RE-6) have to be considered preliminary and will have to be revised. The ICPDR is organising Joint Danube Survey2 (JDS2), which will be performed during summer 2007. All BQEs will be addressed, sampled and assessed using WFD compliant methods for the entire Danube River and the main tributaries. The results of this homogenous data set will be used to supplement the current intercalibration of the Danube River. The improvement of the current IC results should be improved by mid/end 2008.

The ICPDR and therefore the countries of the Eastern Continental GIG will continue the intercalibration exercise after 2006.

The above-mentioned issues should be the objectives of this continued intercalibration exercise. The inclusion of additional countries of the Eastern Continental Region (currently only the EU MS (AT, HU, SK, CZ) and EU Accession Countries (BG, RO) are participating) is intended.

4 References

Alba-Tercedor & Sánchez-Ortega, 1988 [MED]

Alba-Tercedor et al., 2004 [MED]

Armitage, P. D., D. Moss, J. F. Wright & M. T. Furse, 1983. The performance of a new biological water quality scores system based on macroinvertebrates over a wide range of unpolluted running-water sites. Wat. Res. 17: 333-347.

Birk, S. & D. Hering, 2006. Direct comparison of assessment methods using benthic macroinvertebrates: a contribution to the intercalibration exercise. Hydrobiologia 566: 401-415

Böhmer, J., C. Rawer-Jost & A. Zenker, 2004. Multimetric assessment of data provided by water managers from Germany: assessment of several different types of stressors with macrozoobenthos communities. In: *D. Hering, P.F.M. Verdonschot, O. Moog & L. Sandin (eds), Integrated Assessment of Running Waters in Europe. Kluwer Academic Publishers. Printed in the Netherlands.* Hydrobiologia 516: 215-228.

Buffagni, A., S. Erba, M. Cazzola & J. L. Kemp, 2004. The AQEM multimetric system for the southern Italian Apennines: assessing the impact of water quality and habitat degradation on pool macroinvertebrates in Mediterranean rivers. In: *D. Hering, P.F.M. Verdonschot, O. Moog & L. Sandin (eds), Integrated Assessment of Running Waters in Europe. Kluwer Academic Publishers. Printed in the Netherlands.* Hydrobiologia 516: 313-329.

Buffagni, A. & S. Erba, 2004. A simple procedure to harmonize class boundaries of European assessment systems. Discussion paper for the Intercalibration process – WFD CIS WG 2.A ECOSTAT, 6 February 2004, 21pp.

Buffagni A., S. Erba, S. Birk, M. Cazzola, C. Feld, T. Ofenböck, J. Murray-Bligh, M.T. Furse, R. Clarke, D. Hering, H. Soszka & W. van de Bund (2005). Towards European Inter-calibration for the Water Framework Directive: Procedures and examples for different river types from the E.C. project STAR. 11th STAR deliverable. STAR Contract No: EVK1-CT 2001-00089. *Quad. Ist. Ric. Acque* 123, Rome (Italy), IRSA, 467 pp.

Buffagni A., Erba S., Cazzola M., Murray-Bligh J., Soszka H. & Genoni P. 2006. 'The STAR Common Metrics approach to the WFD Intercalibration Process: full application across Europe for small, lowland rivers'. Hydrobiologia 566:379-399.

EC (2003a). Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Guidance on typology, reference conditions, and classification systems for transitional and coastal waters. Luxembourg, Office for Official publications of the European Communities. http://circa.europa.eu/Public/irc/env/wfd/library

EC (2003b). Common implementation strategy for the water framework directive (2000/60/EC). Towards a guidance on establishment of the intercalibration network and on the process of the intercalibration exercise. Luxembourg, Office for Official publications of the European Communities. http://circa.europa.eu/Public/irc/env/wfd/library

EC (2003c). Common implementation strategy for the water framework directive (2000/60/ec). Overall approach to the classification of ecological status and ecological potential. Luxembourg, Office for Official publications of the European Communities. http://circa.europa.eu/Public/irc/env/wfd/library

EC (2005). Common implementation strategy for the water framework directive (2000/60/ec). Guidance on the Intercalibration process 2004-2006. Luxembourg, Office for Official publications of the European Communities. http://circa.europa.eu/Public/irc/env/wfd/library

Dahl & Johnson (2004). A multimetric macroinvertebrate index for detecting organic pollution of streams in Southern Sweden. Archiv fur Hydrobiologie, 160: 487-513.

Gabriels, W. (2007). Proposal for adjusting the Flemish class boundaries according to the intercalibration exercise for river macroinvertebrates. Technical note for the Steering Group of the Central/Baltic Rivers Geographical Intercalibration Group. Flemish Environment Agency, Aalst, Belgium. 20 p. + appendices

Hämäläinen et al. 2002 [NO]

Hämäläinen et al. 2006 [NO]

Hering, D., O. Moog, L. Sandin & P. F. M. Verdonschot, 2004. Overview and application of the AQEM assessment system. In: *D. Hering, P.F.M. Verdonschot, O. Moog & L. Sandin (eds), Integrated Assessment of Running Waters in Europe.Kluwer Academic Publishers. Printed in the Netherlands.* Hydrobiologia 516: 1–20.

LAWA-AO (2006). RaKon Monitoring Teil B. Arbeitspapier III: Untersuchungsverfahren für biologische Qualitätskomponenten. Entwurf 2.1. Stand 22.11.2006. Ständiger Ausschuss "Oberflächengewässer und Küstengewässer" der Bund/Länder-Arbeitsgemeinschaft Wasser (LAWA-AO), Mainz.

Murray-Bligh J, Buffagni A, Cazzola M, Birk S and Vlek H (2006). Procedure for Calculating the ICMi Index: Step-By-Step Instructions For Central-Baltic river GIG. Version 2.6, 19 March 2006

Novak & Bode, 1992 [NO]

Ofenböck, T., O. Moog, J. Gerritsen & M. Barbour, 2004. A stressor specific multimetric approach for monitoring running waters in Austria using benthic macro-invertebrates. In: *D. Hering, P.F.M. Verdonschot, O. Moog & L. Sandin (eds), Integrated Assessment of Running Waters in*

Europe.Kluwer Academic Publishers. Printed in the Netherlands.
Hydrobiologia 516: 251-268.

Ofenböck, T., O. Moog, A. Hartmann & I. Stubauer (2007): Leitfaden für die Erhebung der biologischen Qualitätselemente, Teil A2 – Fließgewässer /Makrozoobenthos, BMLFUW, Wien.)

Pardo, I., Álvarez, M. & E., Roselló. (2007). Asistencia científica y técnica para la aplicación de los anejos II y V de la Directiva Marco del Agua en la Demarcación Hidrográfica del Norte. Technical report.

Pinto, P., J. Rosado, M. Morais & I. Antunes, 2004. Assessment methodology for southern siliceous basins in Portugal. In: *D. Hering, P.F.M. Verdonschot, O. Moog & L. Sandin (eds), Integrated Assessment of Running Waters in Europe.Kluwer Academic Publishers. Printed in the Netherlands.*
Hydrobiologia 516: 191-214.

Van der Molen, D., & R. Pot. (2007) Referenties en concept-maatlatten voor rivieren voor de Kaderrichtlijn Water. Technical report.

Section 3 - Phytobenthos

1 Introduction

For the quality element Phytobenthos the intercalibration exercise has been completed for all of the five geographical intercalibration groups – the Central/Baltic, Alpine, Mediterranean and Northern GIG. The Eastern Continental GIG will complete and report the work for this quality element at a later stage.

2 Methodology and results

2.1 Northern GIG

2.1.1 Introduction

The Northern GIG (N GIG) includes (parts of) Finland, Ireland, Norway, Sweden, and UK. Four of these MS are taking part in the phytobenthos IC exercise: Finland (FI), Ireland (IE), Sweden (SE) and the United Kingdom (UK).

Seven[14] common IC river types were identified for N GIG (Table 2.1.1) and are characterised by the following descriptors:

- catchment area, following System A typology.
- Altitude and geomorphology - three classes: lowland (altitude <200m or below highest coastline), mid-altitude (between lowland and highland), and high (above treeline).
- Alkalinity was used as a proxy for siliceous/calcareous geology, with two classes: low alkalinity (< 0.2 meq/l) and medium alkalinity (0.2-1 meq/l).
- Organic/peat content – two water colour classes: low level (< 30 mg Pt/l) and high level (> 30 mg Pt/l).

However, this river typology was derived primarily for the macro-invertebrate intercalibration. The CB GIG phytobenthos group carried out an evaluation of the CB GIG common IC typology using reference data from eleven participating countries. Their results suggested that the common IC river types for CB GIG did not distinguish between diatom assemblages and consequently the CB-GIG exercise did not use common IC river types. Due to time constraints and the experience of the CB GIG process, the N GIG working group agreed that the "no types approach" was fit for purpose, providing the data submitted to the exercise fitted one of the N GIG common IC river types.

[14] Nine common river types were initially identified in the N GIG but two types (R-N6 and R-N8 were subsequently deleted because only Norway could assign sites to those river types.

Table 2.1.1: Northern GIG common intercalibration river types.

Type	River characterisation	Catchment area (of stretch)	Altitude & geomorphology	Alkalinity (meq/l)	Organic material (mg Pt/l)
R-N1	*Small lowland siliceous moderate alkalinity*	*10-100 km2*	*< 200 m and HC**	*0.2 - 1*	*< 30***
R-N2	*Small-medium lowland siliceous low alkalinity, clear*	*10-1000 km2*	*< 200 m and HC**	*< 0.2*	*< 30*
R-N3	*Small lowland organic*	*10-100 km2*	*< 200 m and HC**	*< 0.2*	*> 30*
R-N4	*Medium/large lowland siliceous moderate alkalinity*	*100-10000 km2*	*< 200 m and HC**	*0.2 - 1*	*< 30*
R-N5	*Small mid-altitude siliceous*	*10-100 km2*	*Between lowland and highland*	*< 0.2*	*< 30*
R-N7	*Small highland siliceous low alkalinity, clear*	*10-100 km2*	*Above treeline*	*< 0.2*	*< 30*
R-N9	*Small – medium mid-altitude siliceous low alkalinity organic (humic)*	*10-1000 km2*	*Between lowland and highland*	*< 0.2*	*> 30*

* - highest coastline
** - Ireland has indicated that they need a higher threshold of 150 mg Pt/l

2.1.2 National approaches to assessing ecological status using phytobenthos

Compliance with normative definitions

Annex V of the WFD treats 'macrophytes and phytobenthos' as a single biological element for the purpose of ecological status assessment and identifies four characteristics of this biological element (taxonomic composition, abundance, likelihood of undesirable disturbances and presence of bacterial tufts) that need to be considered when setting status class boundaries. All MS taking part in the N GIG intercalibration exercise have chosen to develop separate methods for macrophytes and phytobenthos and, in addition, to use diatoms as proxies for phytobenthos. There are, however, differences in national concepts of 'macrophytes' with some MS including larger algae such as *Cladophora* in macrophyte methods whilst others treat these as part of the phytobenthos.

All MS participating in phytobenthos IC were asked to justify their methods in terms of the normative definitions (NDs) and their responses will be considered below. It should be borne in mind that a phytobenthos assessment method does not necessarily need to consider all properties defined in the NDs either because these are considered in a macrophyte method that will be used in parallel with the phytobenthos method or because the MS can demonstrate a relationship between properties defined in the NDs which means that measurement of one property provides an indication of the state of another. In such cases, MS can use a cost-effective method for routine estimation of ecological status whilst, at the same time, demonstrating *de facto* compliance with the NDs.

Table 2.1.2 shows the extent to which the four properties listed in the NDs are incorporated into the national assessment methods. All methods assess taxonomic composition of diatoms alone, however, Ireland and UK have also evaluated the potential for using non-diatoms (Kelly *et al*., 2006a; Kelly, 2006).

Abundance is problematic. Finland and Sweden report that abundance is assessed, but both measure relative, rather than absolute abundance of diatom taxa. Relative abundance is assessed by Ireland and the UK but neither regard this as an assessment of abundance within the meaning of the NDs. The relationship between taxonomic composition, abundance and ecological status was assessed by Ireland and the UK as part of a joint project. The results of this project revealed a relationship between EQR and the upper 90th percentile of biomass measurements, suggesting that the trophic gradient determined the upper limit of biomass at a site but that other factors acted locally to reduce this (Kelly *et al*., 2006a). These findings are broadly in line with those found in other studies (Bernhardt and Likens 2004, Pan *et al.,* 1999, Biggs & Close, 1989; Biggs, 1996) and suggest that routine evaluation of absolute abundance may not yield significant extra information about ecological status.

This suggests that the requirement for assessment of abundance as outlined in the NDs might be better served by macrophyte survey methods, particularly where these include macroalgae. Phytobenthos biomass is very spatially and temporally heterogeneous and therefore quantitative assessment is unlikely to yield detailed insights about ecological status at low or moderate pressure levels. However, at higher pressure levels, visually-obvious growths of macroalgae such as *Cladophora* are likely to be conspicuous, often at the expense of macrophyte diversity more generally, and routine assessment of such growths using straightforward survey techniques may well yield more useful information than quantitative assessment of phytobenthos abundance.

'Undesirable disturbances' are not defined any further in the WFD itself, but ECOSTAT (2005) defines an undesirable disturbance as: 'a direct or indirect anthropogenic impact on an aquatic ecosystem that appreciably degrades the health or threatens the sustainable human use of that ecosystem.' None of the participants in N GIG phytobenthos consider this to be assessed as part of their national methods. Several of the examples of 'undesirable disturbances' listed in ECOSTAT (2005) relate to the effects of macrophytes and phytobenthos on other biological elements, however, it is difficult to differentiate between direct effects of the pressure gradient on these biological elements and interactions with other biological elements.

Similarly, assessment of 'bacterial tufts' are not included directly in any of the assessment systems evaluated here although Sweden includes these growths in other parts of their overall assessment method. Again, a precautionary approach to boundary setting should ensure that the probability of such growths should be minimal when ecological status is good or better.

The view of the phytobenthos expert groups both in N GIG (like Central Baltic GIG) is that if a precautionary approach to boundary setting is taken using other properties (e.g. taxonomic composition), then the probability of undesirable disturbances and bacterial tufts should be minimal when ecological status is good or better.

Table 2.1.2: Northern GIG phytobenthos methods: compliance with WFD normative definitions. ✓ = assessed as part of national metric; X = not included in national metric; 0 = assessed but not included in national metric.

MS	Taxonomic composition	Abundance	Undesirable disturbances	Bacterial tufts
FI	✓	✓	**X**	**X**
Comment	Diatoms only.	Relative abundance of diatom taxa.		
IE	✓	**X**	**X**	**X**
Comment	See comments for UK.	See comments for UK.	See comments for UK.	See comments for UK.
SE	✓	✓	**X**	**0**
Comment	Diatoms only.	Relative abundance of diatom taxa. Percent cover of all benthic algae noted on field protocol, and used in expert assessment of status class.		Noted in field protocol, used in expert assessment of status class.
UK	✓	**X**	**X**	**X**
Comment	Diatoms only. The relationship between diatoms and other algae has been tested (Kelly et al., 2006b; Kelly, 2006). Macroalgae are included in the UK macrophyte method.	There is a negative relationship between EQR and abundance (as chlorophyll a concentration) but abundance is not measured routinely and was not used to set status class boundaries – see Kelly et al. (2006b).	Undesirable disturbances have not been considered.	Bacterial tufts have not been considered.

Evaluation of taxonomic composition

Only two national metrics are currently being used in N GIG by the four participating MS (Table 2.1.3), both of which use existing metrics based on weighted averaging to relate taxonomic composition to ecological status (Table 2.1.2).

Table 2.1.3: National metric/assessment methods for Northern GIG phytobenthos intercalibration.

MS	National metric
FI/SE	Swedish assessment method, Swedish EPA regulations (NFS 2008:1), based on Indice de Polluosensibilité (IPS) (Coste, in CEMAGREF, 1982).
IE/UK	Revised form of Trophic Diatom Index (TDI) (Kelly et al., 2006b)

Placement of status class boundaries

The metrics used by MS convert the response to a pressure gradient into a continuous variable which then has to be converted into an EQR, computed from Observed (O) and Expected (E) values. MS adopted a variety of approaches to split this EQR scale into separate status classes. Table 2.1.4 summarises these approaches.

The NDs define high, good and moderate status in terms of their deviation from the biota expected at the reference state and, therefore, a national method, if it is to be compliant with the NDs, has to be able to express each status class in terms of change from the reference state.

Table 2.1.4: Rationales for defining phytobenthos high/good and good/moderate class boundaries in Northern GIG.

	High / Good Boundary	Good / Moderate Boundary
FI	High/good boundary: IPS=17	Good/moderate boundary: IPS=15
	Preliminary national boundaries for IPS are based on the study of Eloranta & Soininen (2002). The study was based on data of 56 streams with varying degree of alteration of water chemistry. Streams were first classified into five classes according to land use and alteration of water chemistry: 1) near pristine streams with only minor degree of human activities in drainage area, 2) good quality streams with some forestry activities and low degree of agriculture, but with low load of nutrients or suspended materials, 3) moderate quality streams with moderate degree of agriculture and forestry or/and more dense populated areas, 4) poor quality streams with more intense agriculture and forestry, fish farming or small waste water plants, 5) bad quality streams loaded with effluents from different sources. However, none of the studied sites were heavily polluted. Boundaries for ecological quality classes for IPS were then derived from this classification.	
IE	The high/good boundary is set at the 75th percentile of EQR values for reference sites within a particular type. See comments for UK	'Crossover' between nutrient-sensitive and nutrient-tolerant species (Pollard and van de Bund, 2005).
SE	High/good boundary: IPS=17,5 High status: River/stream fulfils the national reference criteria, e.g. Tot-P < 10 µg/l or no eutrophication (area-specific loss of Tot-P = class 1); no acidification, pH > 6	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 ca. 30 % (and the amount of sensitive species falls below ca. 30 %).
UK	The high/good boundary is set at the 75th percentile of EQR values for reference sites within a particular type.	'Crossover' between nutrient-sensitive and nutrient-tolerant species (Pollard and van de Bund, 2005).
	Biological metrics tend to show gradual change as the level of nutrient/organic pressure increases, with no distinct discontinuities that could act as criteria for setting class boundaries. An alternative approach – based on the proportions of nutrient-tolerant, nutrient-sensitive and indifferent taxa within samples – was used to define	

	High / Good Boundary	Good / Moderate Boundary
	status class boundaries in the UK, with the good/moderate boundary set at the point where the proportion of sensitive taxa falls below that of tolerant taxa. In ecological terms, the diatom flora at high and good status is characterised by a number of taxa, often with relatively broad niches (e.g. *Achnanthidium minutissimum, Fragilaria capucina*) which occur at different phases of a microsucession from colonisation of bare rock up to a mature biofilm (see Biggs *et al*., 1989). At high status, these are accompanied by other nutrient-sensitive taxa but as nutrient concentrations increase, the most sensitive of these taxa disappear whilst the numbers of nutrient tolerant taxa increases. The 'crossover' is, therefore, the point at which the taxa which form the 'association' characteristic of a site in the absence of pressure become subordinate to taxa which are favoured by a pressure (nutrients, in this case). The EQR gradient below the good/moderate boundary is then divided into three equally-spaced portions from which the moderate/poor and poor/bad boundaries are derived.	

2.1.3 Test datasets

A summary of the number of sites available in each quality class (including reference sites) from each MS is presented in Table 2.1.5. In the N GIG, seven common IC river types were defined (Table 2.1.1). The data submitted for the IC exercise was required to fit into one of these seven IC common river types defined by N GIG even though the expert group also agreed to consider intercalibrating using a common river types approach. Those parts of UK which met criteria for N GIG tended to occur in regions well away from large towns and, consequently, the datasets had relatively few sites with status classes that were moderate or lower. The UK dataset used for intercalibration is, therefore, composed of sites that fulfil criteria for either N GIG or CB GIG in order to cover the entire status gradient. The national assessment systems use a site-specific prediction of expected values which compensates for any typological differences between N GIG and CB GIG sites.

Also the SE dataset is composed like the one from UK. The national approach for SE includes only one type, as there were no significant differences between reference values.

Table 2.1.5: Number of reference sites and phytobenthos samples available in each quality class from each Member State in the Northern GIG.

	Reference	H	G	M	P	B	Total
FI	66	79	23	10	4		116
IE	36	139	33	18	6	1	197
SE	61	82	16	24	4	1	127
UK	69	454	394	438	124	6	1,416
Total	232	718	466	490	138	8	1,856

2.1.4 Standardisation of reference conditions

Introduction to Reference Conditions

The concept of 'type-specific reference conditions' is central to the WFD as ecological status is defined in terms of deviation from the biota expected under such conditions. Different interpretations of 'reference conditions' may lead to different values being used as the denominator in EQR calculations leading, in turn, to the same 'observed' biota having different ecological status assessments. On the other hand, the WFD also recognises that the 'expected' biota will vary from place to place depending on local factors such as climate, underlying geology and stream order and this too will have an effect on ecological status class boundaries. The challenge facing the IC exercise is to differentiate between those differences in national reference states that reflect genuine biogeographical variability across the GIG and those that reflect differences in approach by those responsible for implementation.

Evaluation of reference conditions and principles of setting classification boundaries within the GIGs assumes a cascade of effects, with alterations to catchments (removal of natural vegetation, replacement by agriculture or urban development) leading to increases in pressure variables in surface water which, in turn, affect the biota. Ideally, evaluation of reference conditions focuses on changes to the catchment, and incorporates data on land use and supports this with data on pressure variables (nutrients, BOD etc). The final approach – use of the biota to define reference conditions – is not encouraged as the NDs define ecological status in relation to the biota expected under undisturbed conditions (Annex V, article 1.2) and the use of land-use and pressure data to define 'undisturbed conditions' ensures rigour and objectivity in the definition of the 'expected' value.

In common with most members of CB-GIG, N-GIG participants used the median metric values of reference samples as the 'expected' value. This is a more stable property than alternatives (e.g. use of 95^{th} percentile values), especially when the population of reference sites is small; however, one consequence is that a number of high status sites will have EQR >1. In such cases, EQR values >1 can be automatically set to 1 for reporting.

Reference screening procedures

The phytobenthos expert group adopted an approach that is consistent with other Intercalibration working groups (Central Baltic (CB) GIG phytobenthos and N GIG/CB GIG macro-invertebrate groups) to define what is meant by reference conditions. Member States followed REFCOND guidance (Working Group 2.3 - REFCOND Guidance Document No 10.) when initially choosing reference sites. A list of the more detailed criteria and type-specific concentrations ("reference thresholds") of key chemical parameters were developed by the N GIG macro-invertebrate working group for rivers. The thresholds aim to interpret the WFD requirement of "very minor anthropogenic impact".

Representatives from each MS were asked to screen reference sites, chosen using REFCOND guidance, against agreed catchment land use and chemical reference thresholds. The thresholds (Table 2.1.6) were principally derived from datasets linking invertebrates to general chemical elements, but other values taken from national water quality classifications, diatoms datasets (in the case of nutrients), specific studies and expert opinions were also considered. The proposed reference thresholds allow the same criteria to be applied to the selection of all reference samples used in the IC exercise in N GIG rivers and were intended for use in conjunction with other general pressure criteria. Both mean values and 90- or 95-percentile values were proposed for some parameters. The mean is the most robust statistic when few data are available, as is frequently the case for new reference sites. The 90th or 95th percentile should be used only when sufficient data are available (at least 12 monthly chemical samples).

Table 2.1.6: Northern GIG guidelines for physico-chemical characteristics and general characteristics of reference river sites. Physico-chemical values to be regarded as maximum threshold values for screening reference sites. Values may vary according to national typologies. *Cf.* Appendix Table A1 for guidance from REFCOND and N GIG on reference sites.

Quality Element of Characteristic	**Concentration or Descriptor at Reference Condition**	**Countries Using this Criterion**
Pollution Status	Pristine, Unpolluted	ALL
Organic Waste Load	No Observed Effect	ALL
Nutrient Loads	Background	ALL
90%ile B.O.D.	< 2.7 mg/l	IE
Mean BOD	<1.6	IE
Dissolved Oxygen	Close to 100% (>80% and < 120% saturation at all times)	IE, FI
95%ile Non-ionised Ammonia (mg/l N)	Compliant with the Freshwater Fish Directive National Regulations	IE, FI
Annual Mean total Ammonium (mg/l N)	Compliant with the Freshwater Fish Directive National Regulations for total ammonium	IE, FI
95%ile Total ammonium (mg N/l)	<0.04 mg/l	IE, FI, SE
Annual Median ortho-Phosphate	<0.015 mg P/l	IE, UK, SE
Annual Mean ortho-Phosphate	<0.03 mg P/l	IE, UK, SE
Annual mean total P	R-N1 < 20 ug/l R-N3 < 30ug/l R-N4 < 18 ug/l R-N5 <18 ug/l	SE, FI,
Annual Mean Nitrate (mg N/l)	< 1.6 mg N/l	SE, IE, UK, FI
Annual Mean Total N (mgN/l)	<1.8 mg N/l	FI, SE

Table 2.1.7 indicates which of the N GIG defined reference criteria were used for the screening exercise and what sources of information were available to each MS for this purpose. Member States were also asked to indicate if they used more stringent criteria (or different but equivalent ones).

Table 2.1.7: Criteria used by Member States for phytobenthos reference site selection in the Northern GIG. Key: 0: missing info; 1: not used; 2, Yes, Measured; 3, Yes, Estimated; 4, Yes, Field inspection; 5, Yes, Expert judgement.

	Landuse	BOD_5	O_2	$N\text{-}NH_4$	P- fraction	$N\text{-}NO_3$	Comments
FI	5	0	0	2	2	2	water chemistry not available for all sites
IE	2	2	2	2	2	2	See paragraph below
SE	2	1	1	1	2	1	
UK	2	0	0	2	2	2	

The following paragraphs give a more detailed description of the screening exercise for reference sites as undertaken by each MS:

Finland:
The main pressure criteria are: no major point sources, agriculture and forestry in catchment upstream of reference sites of low intensity (< 10% agriculture in total catchment area, no large clear cuts, mainly judged from visual observation of GIS land-use), Total P median concentration < 20 µg l-1. Experts from the regional environmental centres were used in the final determination.

Ireland:

Reference screening in Ireland was carried out by selecting reference sites for which maximum catchment land cover limits were below an agreed percentage, as carried out for the NGIG invertebrate intercalibration exercise. The CORINE Land Cover dataset was used to provide an estimate of the upstream land cover using ESRI's Arc View 3.2a GIS software. Water chemistry results for these selected sites were extracted from the Agency's water quality database, for sites where suitable water chemistry existed. Sites that did not meet the criteria for reference site water quality set out in Table 2.1.6 were removed from the list. Potential reference sites were also compared against their rTDI score (national metric for phytobenthos) and Q-Value (national metric for invertebrates). The final selection was found to have an rTDI score indicative of high status and a Q-Value of 4.5 – 5, also indicative of high status.

Sweden:

For the N-GIG, we used the following screening factors for a reference stream:
1a) < 10 µg/l Tot-P
1b) IF colour was high (> 100 mg Pt/l), then < 20 µg/l Tot-P
2) pH > 6

United Kingdom:

A database of SEPA-monitored diatom sites (which comprise the majority of N GIG sites in the UK) was used as the basis for reference site selection in the N GIG phytobenthos Intercalibration exercise. Sites were initially assigned to N GIG river types following the descriptors outlined in Table 2.1.1. Expert judgement was used in a minority of situations to make allowances for sites that were marginally outside the upper and lower threshold limits for N GIG river type descriptors. Colour data was not available to distinguish between the two water colour classes.

Screening for physico-chemical and landuse characteristics was carried out for all sites in the SEPA database in the initial stages of the selection process. The full process of reference site selection and validation is described as follows:

1. *Landcover 2000* data obtained for the SEPA database of sites was used as the basis for the landuse screening exercise. With the exception of forestry, the maximum landuse threshold limits used followed the guideline threshold limits for N GIG defined in Table 2.1.6; these were as follows:
   - Arable: 10%
   - Permanent crops: 15%
   - Pasture: 30%
   - Forestry*: 30% (Central Baltic GIG threshold substituted)
   - Urban fabric: <0.8% of catchment

   **Landcover 2000* does not distinguish between (semi-)natural woodland and plantations. A threshold value of 30% forestry was used as a proxy for the N GIG guideline of <5% clear-felled/planted forest.
2. The maximum chemical threshold values for screening of reference sites were as follows (*cf.* Table 2.1.7):
   - Soluble Reactive Phosphorus: 30 ug l^{-1}
   - Nitrate-N: 1.6 mg l^{-1}
3. Following the landuse/physio-chemical screening, expert judgement was used to review the list of proposed reference sites. In addition, the characteristics of each site was validated using the *SEPA GIS interactive Map* to check the proximity of potential sources of point/diffuse inputs, morphological alterations and biological/recreational pressures; any additional information logged against site locations was also taken into account.
4. Sites known to be influenced by acidification and with pH<6 were also eliminated from the selection.

5. The final step in the validation of the N GIG reference sites was on the basis of the revised TDI calculation. Any potential reference sites with revised TDI scores > 50 were removed to ensure that the final selection of sites did not include those influenced by elevated nutrient concentrations.

2.1.5 Development of Common Metric

In order to compare status class boundaries developed in each MS, national metrics first had to be converted to a common scale. The mechanism for doing this was to develop an 'intercalibration common metric' (ICM) (corresponding to Option 2 outlined in the Boundary Setting Protocol) similar to that developed for the CB GIG invertebrate IC exercise (Buffagini *et al*., 2005). This ICM should have a statistically-significant relationship with each national metric so that EQR values computed using national metrics can be quoted as the corresponding value of the ICM. In the case of N GIG phytobenthos, there was a high degree of congruence between national methods with common sampling and analysis methods (CEN, 2003, 2004; Kelly *et al*., 1998), and relying on the fact that both metrics used for the exercise are based on the weighted average (WA) equation of Zelinka and Marvan (1961).

Evaluation of Candidate Metrics

N GIG used a slightly different ICM to that used in CB GIG, although it is based on identical principles. The N GIG ICM is composed of two metrics developed in Austria: Trophien Index (TI) and Saprobien Index (SI). The N GIG ICM had two advantages over the CB GIG ICM:

1. Neither component metric is used by any participant in N GIG, so the ICM is independent of national methods (something that CB GIG were unable to achieve).

2. When tested against the national metrics, the N GIG ICM also had a better relationship with the IE and UK national metrics than the CB GIG ICM (composed of the TI and IPS).

Two variants of the N GIG ICM were tested – one based on the mean of the two component metrics (TISI-mean) and the other based on the minimum (TISI-min). Relationships between national metrics and the ICMs (TISI-mean and TISI-min) were evaluated using identical criteria to those used in CB GIG. These were as follows:

a. Nationally agreed assessment system and boundary values;

b. At least six reference samples (representing at least four sites);

c. A statistically-significant linear relationship with the ICM. More particularly:

   - Root mean square error (RMSE) ≤ 0.15
   - Coefficient of determination (r^2) ≥ 0.5; and,
   - Slope ≥ 0.5 and ≤ 1.5.

The coefficient of determination (r^2) measures association between two variables and gives little indication of the predictive power of that relationship. It is also dependent, to some extent, on the length of the gradient over which the coefficient is applied (see **Fig. 2.1.5.3**). RMSE, on the other hand, gives a better indication of the predictive power of the relationship, regardless of gradient length Using both, along with visual examination and slope, provides a robust basis for evaluating relationships between national metrics and the ICMs.

The properties of the relationships are shown in Table 2.1.8. FI and SE metrics showed a stronger relationship with TISI-mean whilst UK and IE had a stronger relationship with TISI-min.

Table 2.1.8: Regression properties for national metrics versus ICMs for the four national datasets used in the N GIG phytobenthos intercalibration exercise.

		TISI-mean			TISI-min		
	n	**r^2**	**RMSE**	**slope**	**r^2**	**RMSE**	**slope**
FI	112	0.601	0.0945	1.31	0.6292	0.115	1.692
IE	197	0.3716	0.129	0.4865	0.4063	0.157	0.64
SE	122	0.84	0.053	0.7	0.846	0.052	0.54
UK	920	0.562	0.141	0.72	0.612	0.133	0.834

Evaluation of the Intercalibration Common Metric

Table 2.1.9 shows the relationship between ICM-min and ICM-mean and nitrogen and phosphorus fractions. Note that the primary purpose of an ICM is to allow values of national metrics to be compared, so the performance characteristics in Table 2.1.8 are more instructive for the purposes of selecting an ICM but Table 2.1.9 helps to illustrate the relationship between the ICMs and the underlying nutrient / organic gradient.

Table 2.1.9: Correlation coefficients between nutrients and the minimum ('min') and mean ('mean') intercalibration metric (TISI) in the Northern GIG phytobenthos Intercalibration exercise. 'SRP' = soluble reactive phosphorus (≈ PO_4-P); 'NOx' = nitrogen oxides (≈ NO_3-N + NO_2-N).

Member State	**Determinand**	**Data Type**	**TISI-min**	**TISI-mean**
FI	Log Total N	Median	-0.466***	-0.505***
FI	Log Total P		-0.466***	-0.505***
IE	Log NOx	Spot	-0.5405*	-0.5211**
IE	Log PO_4-P		-0.3597*	-0.3391
SE	Log NH_4-N	Mean	-0.43***	-0.47***
SE	Log Total N		-0.75***	-0.76***
SE	Log NOx		-0.74***	-0.75***
SE	Log Total P		-0.81***	-0.83***
SE	Log PO_4-P		-0.81***	-0.83***
UK	Log NO_3-N	Mean	-0.604 ***	-0.610 ***
UK	Log NOx		-0.515 ***	-0.508 ***
UK	Log SRP		-0.659 ***	-0.648 ***

Significance level: P < 0.05: *; P < 0.01: **; P < 0.001: ***

Conversion of national metrics to the ICM

For each MS, the N GIG ICM was calculated as follows:

a. EQR values based on SI and TI values were calculated using MS data.

b. The expected value for each EQR value is the median of reference values for the MS.

c. Two ICMs were calculated: one as minimum of TI and SI (TISI-min) and one as the mean of TI and SI (TISI-mean)

d. The regression between the ICMs and the national metric was plotted – based on all sites in H, G and M only (some national datasets had non-linear relationships with the dataset and using just H, G and M confined the relationship to the linear portion). The regression equation and associated statistics (r^2, root mean square error, slope) were calculated (Table 2.1.8).

e. Once the linear relationship was confirmed, values of the national metric representing the High / Good and Good / Moderate boundaries were converted to corresponding values of the ICM for both ICMs. The procedure for doing this is identical to that used in the CB GIG invertebrate IC exercise and is based on a linear regression equation:

   ICM = a + b(national metric as EQR)

   Where: a = constant; b = slope.

   Figure 2.1.1 shows a regression between the EQR values of a national metric and the ICM for a hypothetical national dataset and illustrates the process of converting the national value of the Good/Moderate boundary to the ICM.

   A single relationship was computed for each national dataset and this relationship was used to convert boundary values for each national type to the ICM.

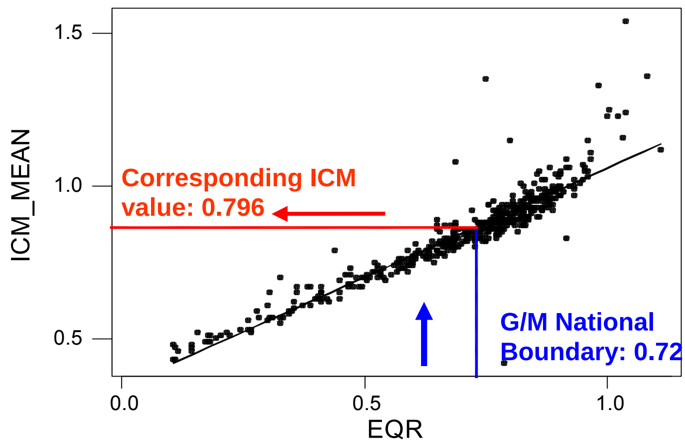


Figure 2.1.1: Conversion of the Good/Moderate national boundary value for a hypothetical national dataset into an ICM value using the regression formula: ICM = a + b(national metric as EQR).

2.1.6 Comparison of boundaries and harmonisation

Overview of results

The acceptable range of boundary values was calculated by identical criteria to those used in CB GIG, as the median boundary value ± 0.05 EQR units for all MS who fulfilled the statistical criteria described in Section 2.1.5. However, as only four countries are included in the exercise, the statistical power of the exercise is relatively low, and results are presented with an acceptable band based on boundary values for all four MS as well as with an acceptable band based on just those that fulfil the statistical criteria.

Table 2.1.10 shows a detailed breakdown of results for the high/good and good/moderate boundary for both ICMs. Table 2.1.11 presents the results of the intercalibration in terms of the relationship between national boundaries and the limits of the 'acceptable band'. SE boundaries are high for all tests performed using TISI-min (but, as these lie above the 'acceptable band' there are no implications for harmonisation). FI and IE were both marginally below the acceptable band for one of the comparisons. Experience from CB GIG suggests that both differences lie within the statistical limits of the exercise; again, there are no implications for harmonisation. Each of these cases is considered in more detail below.

Table 2.1.10: Boundary values for national methods involved in the N GIG phytobenthos intercalibration exercise.

	H/G			G/M		
	National metric	**TISI-mean**	**TISI-min**	**National metric**	**TISI-mean**	**TISI-min**
FI	0.912	0.892	0.804	0.804	0.751	0.622
IE	0.93	0.846	0.762	0.78	0.773	0.666
SE	0.89	0.905	0.930	0.74	0.800	0.850
UK	0.93	0.898	0.804	0.78	0.790	0.679
Acceptable bands						
	All MS			**All MS**		
	Median	0.895	0.804		0.782	0.673
	Upper limit	0.945	0.854		0.832	0.723
	lower limit	0.845	0.754		0.732	0.623
	Excluding IE			**Excluding IE and FI**		
	Median	0.898	0.804		0.795	0.679
	Upper limit	0.948	0.854		0.845	0.729
	lower limit	0.848	0.754		0.745	0.629

Table 2.1.11: Implications for harmonisation in the N GIG phytobenthos intercalibration exercise.

	H/G		G/M	
	TISI-mean	**TISI-min**	**TISI-mean**	**TISI-min**
Acceptable band based on all MS				
Inside acceptable band	All	FI, IE, UK	All	FI, IE, UK
Above		SE		SE
Acceptable band based on those MS that fulfill statistical criteria				
Inside acceptable band	FI, SE, UK	FI, IE, UK	All	IE, UK
Above		SE		SE
Below	IE			FI

Detailed comments

FI: Finland

The national assessment methods for diatoms are under development. The Finnish classification is, therefore, only preliminary and IPS values for class boundaries will be re-evaluated. There was also a wide variation in IPS values among reference sites, indicating that stratification for natural background variability might also be needed. Preliminary results have shown that the stream typology used for macroinvertebrates may not be useful for diatoms. Alternative typologies should thus be considered. Also metrics other than IPS should be tested in near future.

IE: Ireland

Ireland has a low coefficient of determination in the regression between the national EQR and ICM; with the ICM based on TISI-min ($r^2 = 0.4063$) being slightly better than that observed for the TISI-mean ($r^2 = 0.3716$). These regression statistics are lower than that obtained for other member states, including the UK, with whom IE shares a common national metric. These lower regression statistics are probably influenced strongly by several aspects inherent in the IE dataset. The IE dataset is heavily weighted towards the higher quality classes (see Table 2.1.5) with approximately 84% of the sites in high and good status. The number of alkalinity values necessary for the calculation of the national EQR was limited, and when estimated from conductivity for lower alkalinity sites some error in the EQR would be expected. Default rather than measured alkalinity values were also used in the EQR calculation for a large proportion of the sites.

Low correlation coefficients between the ICMs and nutrients were also observed, again with TISI-min giving a slightly better relationship. The relatively small number of sites used in this analysis, coupled with the chemistry results for some of these sites being from different years to that of the biological samples, and again the lack of dynamic range because most are of high or good status explains the low correlation coefficients in this instance.

When the acceptable bands are calculated (excluding IE due to poor regression statistics), Ireland is inside the acceptable band for the H/G and G/M boundary for the ICM based on TISI-min, and the G/M boundary for the TI/SI-mean. Ireland falls just outside the lower boundary of the TISI-mean for the H/G boundary, but only at the third decimal place.

SE: Sweden

The position of the SE boundaries is consistent with the results of the CB-GIG exercise, with both high/good and good/moderate boundaries falling within the ‘acceptable band’, when using the TISI mean.

UK: United Kingdom

The position of the UK boundaries is consistent with the results of the CB-GIG exercise, with both high/good and good/moderate boundaries falling within the ‘acceptable band’.

2.1.7 Conclusions/Recommendations

General issues associated with phytobenthos intercalibration exercises are addressed in the report on the CB GIG intercalibration exercise. The conclusions and recommendations listed in that report are all equally valid for the N GIG exercise. This section highlights a few points that are unique to the N GIG exercise.

The CB GIG exercise involved 12 Member States; whilst the N GIG exercise is much smaller, with just four participants. An important implication is that the exercise has lower statistical power and it is not always clear if those MS that fall outside the 'acceptable band' do so because there are issues that those MS need to address or because the 'acceptable band' is itself based on a small (and potentially atypical sample). On the other hand, however, the 'acceptable band' should not be equated with 'best practice'. MS that comply with the minimum requirements of the exercise are included in the acceptable band and the position of this band, therefore, reflects the consensus of those.

This must affect how results from N GIG and other smaller intercalibration exercises are judged. In particular, a 'Type 1 error' (i.e. erroneous rejection of the [null] hypothesis that boundaries are the same) may lead to the conclusion that a MS needs to adjust boundaries when, in fact, the median value of the ICM (which anchors the acceptable band) is unlikely to be stable with such a small sample size.

The approach adopted here was, therefore, to perform a suite of tests using different permutations of the statistical criteria and to make final judgements about the need (or otherwise) to adjust boundaries based on the weight of evidence. Whilst the CB GIG exercise evaluated two versions of the ICM (one based on the mean of component metrics, the other based on the minimum), the N GIG exercise used both versions. TISI-min favoured IE and UK, both of whose national metric was the TDI, which correlates more strongly with the nutrient-sensitive TI, whilst TISI-mean favoured FI and SE whose national metric was the IPS, which correlated more strongly with the SI. Whilst TISI-mean is not biased by a low value of one or other metric, TISI-min better embodies the 'one out, all out' principle used when comparing biological elements as part of status assessments.

Three of the four MS taking part in this exercise were also involved in the CB GIG exercise. Boundaries calculated in this exercise are broadly consistent between the two exercises. For H/G, IE, SE and UK were all inside the acceptable band for the CB GIG exercise whilst, for N GIG, UK were inside whilst SE was above the acceptable band for TISI-min but inside for TISI-mean and IE was marginally below for TISI-mean. For G/M, UK and SE were inside the acceptable band whilst IE was above. For the N GIG exercise, IE and UK were inside the acceptable band on all occasions whilst SE was again above the acceptable band when TISI-min was used. In the case of IE, the relatively small size of the dataset plus the low number of poor quality sites may be responsible for the differences in regression equations.

Whilst SE were above the acceptable band on two out of four occasions for each of H/G and G/M comparisons, it is only those MS that fall below the acceptable band that need to consider harmonisation. In this exercise, both IE and FI fell below the acceptable band on one out of four occasions, both were only marginally below the acceptable band on these occasions and we believe that there is no case for either MS to adjust their boundaries.

2.1.8 Appendix

Table A.1: REFCOND and N GIG guidance with regard to the description of reference sites to be included in the rivers intercalibration exercise. See Table 4.2 for physico-chemical thresholds.

	REFCOND	**N GIG Definition**
General statement	High status or reference conditions is a state in the present or in the past corresponding to very low pressure, without the effects of major industrialisation, urbanisation and intensification of agriculture, and with only very minor modification of physico-chemistry, hydromorpology and biology.	High status or reference conditions is a state in the present or in the past corresponding to very low pressure, without the effects of major industrialisation, urbanisation and intensification of agriculture, forestry, aquaculture and with only very minor modification of physico-chemistry, hydromorpology and biology.
Diffuse source pollution	**REFCOND**	**NGIG Definition**
Land-use intensification: Agriculture, forestry	Pre-intensive agriculture or impacts compatible with pressures pre-dating any recent land-use intensification. Pressures pre-dating any recent intensification in airborne inputs that could lead to water acidification.	**Agriculture and Forestry:** Agriculture and forestry in catchment upstream of reference sites of low intensity. Maximum percentage area for screening sites with respect to land cover in catchment upstream of a point at which reference conditions are believed to exist is as follows using CORINE terminology: (Figures are tentative and may vary from region to region. In larger reference catchments proximity of pressure to the proposed reference site may be taken into account. Where CORINE datasets are not available similar land use cover data may be used.) **Agriculture:** Arable land – less than 2 – 10 % Pastures- less than 30% Permanent crops– less than 15% **Forestry:** Forests - clear-felled area/planted area within last 5 years - < 5% **Diffuse Urban Pressures:** Urban fabric – <0.8% of catchment (close to zero)
Point source pollution	**REFCOND**	**NGIG Definition**
Specific synthetic pollutants	Pressures resulting in concentrations close to zero or at least below the limits of detection of the most advanced analytical techniques in general use (A Selection process for relevant pollutants in a river basin is presented as an example of best practice in section 6 of the guidance document from Working Group 2.1, IMPRESS).	• Pressures resulting in concentrations close to zero or below the limits of detection in water of the analytical techniques in general use. Concentrations should be below the NEC level or established national EQS values where available. • No significant point sources. • Airborne pollutants in water at background concentration.
Spec. non-synthetic pollutants	Natural background level/load (see reference above)	• At natural background concentrations or below EQS where available.
Other effluents/discharges	No or very local discharges with only very minor ecological effects.	• No or very local discharges with only very minor ecological effects. • No effects from IPPC controlled industrial plants • No other major discharges controlled by other statutory pollution control licences
Morphological alterations	**REFCOND**	**NGIG Definition**

River morphology	Level of direct morphological alteration, e.g. artificial instream and bank structures, river profiles, and lateral connectivity compatible with ecosystem adaptation and recovery to a level of biodiversity and ecological functioning equivalent to unmodified, natural water bodies	Level of direct morphological alteration, e.g. artificial instream and bank structures, river profiles, and lateral connectivity compatible with ecosystem adaptation and recovery to a level of biodiversity and ecological functioning equivalent to unmodified, natural water bodies. No major dams or control structures upstream of reference condition site. The river should not have been subject to any arterial drainage schemes that affect lateral connectivity or cause changes in the natural time of residence. River substratum should be appropriate to the catchment geology and river slope at the point of substratum assessment.
Water abstraction	**REFCOND**	**NGIG Definition**
water abstraction	Levels of abstraction resulting in only very minor reductions in flow levels or lake level changes having no more than very minor effects on the quality elements.	• Abstraction of water from the river upstream of a site regarded as being at reference condition should not **reduce** the 95 percentile discharge flow (m^3/s) by more than 10%. (The 95 percentile flow or discharge is that which is exceeded 95% of the time over the hydrological year).
Flow regulation	**REFCOND**	**NGIG Definition**
River flow regulation	Levels of regulation resulting in only very minor reductions in flow levels or lake level changes having no more than very minor effects on the quality elements.	• Levels of regulation resulting in only very minor reductions in flow levels having no more than very minor effects on the quality elements. As a guideline low flow alteration should be less than 20% of monthly minimum flow. • There should be no major dams or control structures upstream of the reference condition site. Dams located downstream should not affect the flow regime at the reference site and should not impede the passage of migratory fish.
Riparian zone vegetation	**REFCOND**	**NGIG Definition**
	Having adjacent natural vegetation appropriate to the type and geographical location of the river.	• Having adjacent natural vegetation appropriate to the type and geographical location of the river.
Biological pressures	**REFCOND**	**NGIG Definition**
Introductions of alien species	Introductions compatible with very minor impairment of the indigenous biota by introduction of fish, crustacea, mussels or any other kind of plants and animals. No impairment by invasive plant or animal species.	• Introductions compatible with very minor impairment of the indigenous biota by introduction of fish, crustacea, mussels or any other kind of plants and animals. • No impairment by invasive plant or animal species. • No recent introductions (<15 years) that are still causing major ecological changes within a river ecosystem.
Fisheries and aquaculture	Fishing operations should allow for the maintenance of the structure, productivity, function and diversity of the ecosystem (including habitat and associated dependent and ecologically related species) on which the fishery depends Stocking of non indigenous fish should not significantly affect the structure and functioning of the ecosystem.. No impact from fish farming.	• There should be no commercial fishing operations or fish farming which affects the biological quality elements or water quality of the river system. No significant stocking of non-native species or stocking of 'put and take' fish for angling purposes.
Biomanipul-ation	No biomanipulation.	• No biomanipulation or liming of the system in response to acidity pressures.
Other pressures	**REFCOND**	**NGIG Definition**
Recreation uses	No intensive use of reference sites for recreation purposes (no intensive camping, swimming, boating, etc.)	• No intensive use of reference sites for recreation purposes (camping, swimming, boating, etc.) causing physical, chemical or biological disturbance

2.2 Central-Baltic GIG

2.2.1 Intercalibration approach

The Central GIG includes (parts of) Austria, Belgium, Czech Republic, Denmark, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and UK. The Baltic countries – Estonia, Lithuania and Latvia – are included with the Central GIG countries, although it is recognised that rivers (and lakes) in these regions are often quite different from the rest of the Central regions, with very high values for alkalinity and organic matter. The river expert network, however, recommended the merging of the Central and Baltic GIGs (CB GIG) for rivers after an analysis of the metadata submitted for the draft IC network in 2004.

Twelve MS belonging to CB GIG are taking part in the phytobenthos IC exercise: Austria (AT), Belgium (BE), Estonia (EE), France (FR), Germany (DE), Ireland (IE), Luxembourg (LU), Netherlands (NL), Poland (PL), Spain (ES), Sweden (SE) and the United Kingdom (UK). Czech Republic, Denmark, Italy, Latvia and Lithuania are also part of CB GIG but have not been involved in this exercise. The two administrative regions of Belgium, Flanders (BE-F) and Wallonia (BE-W) have different methods for assessing ecological status and are treated separately here.

Six common river types were identified for the CB GIG (Table 2.2.1) and are characterised by the following descriptors:

- catchment area, following System A typology.
- altitude - two classes: lowland (altitude <200m), mid-altitude (from 200 – 800 m).
- geomorphology – for each of the types a description is given, taking substrate and width into account.

Table 2.2.1: **Central-Baltic GIG common intercalibration rivers types**

Type	River characterisation	Catchment area (of stretch)	Altitude & geomorphology	Alkalinity (meq/l)
R-C1	*Small lowland siliceous sand*	10-100 km^2	lowland, dominated by sandy substrate (small particle size), 3-8m width (bankfull size)	> 0,4
R-C2	*Small lowland siliceous - rock*	10-100 km^2	lowland, rock material 3-8m width (bankfull size)	< 0,4
R-C3	*Small mid-altitude siliceous*	10-100 km^2	mid-altitude, rock (granite) - gravel substrate, 2-10m width (bankfull size)	< 0,4
R-C4	*Medium lowland mixed*	100-1000 km^2	lowland, sandy to gravel substrate, 8-25m width (bankfull size)	> 0,4
R-C5*	*Large lowland mixed*	1000-10000 km^2	lowland, barbel zone*, variation in velocity, max. altitude in catchment: 800m, >25m width (bankfull size)	> 0,4
R-C6	*Small, lowland, calcareous*	10-300 km^2	lowland, gravel substrate (limestone), width 3-10m (bankfull size)	> 2

*mixed cyprinids, with some salmonids

- alkalinity was used as a proxy for siliceous/calcareous geology, with three classes – low (< 0,4 meq/l), medium (0,4 – 2 meq/l), and high (>2 meq/l).

CIS guidance suggested that river types should be split up if necessary to ensure that IC will compares like with like.

2.2.2 National methods that were intercalibrated

Compliance with normative definitions

Annex V of the WFD treats 'macrophytes and phytobenthos' as a single biological element for the purpose of ecological status assessment and identifies four characteristics of this biological element (taxonomic composition, abundance, likelihood of undesirable disturbances and presence of bacterial tufts) that need to be considered when setting status class boundaries. Most MS in CB GIG have chosen to develop separate methods for macrophytes and phytobenthos and, in addition, to use diatoms as proxies for phytobenthos. There are, however, differences in national concepts of 'macrophytes' with some MS including larger algae such as *Cladophora* in macrophyte methods whilst others treat these as part of the phytobenthos.

All MS participating in phytobenthos IC were asked to justify their methods in terms of the normative definitions (NDs) and their responses will be considered below. It should be borne in mind that a phytobenthos assessment method does not necessarily need to consider all properties defined in the NDs either because these are considered in a macrophyte method that will be used in parallel with the phytobenthos method or because the MS can demonstrate a relationship between properties defined in the NDs which means that measurement of one property provides an indication of the state of another. In such cases, MS can use a cost-effective method for routine estimation of ecological status whilst, at the same time, demonstrating *de facto* compliance with the NDs.

Table 2.2.2 shows the extent to which the four properties listed in the NDs are incorporated into the national assessment methods. All methods assess taxonomic composition of diatoms alone although in two MS (AT, DE) there is also parallel assessment of non-diatoms. Of the remaining MS, only two (IE, UK) have evaluated the relationship between diatoms and non-diatoms (Kelly *et al*., 2006; Kelly, 2006).

Abundance is problematic. All but two MS (IE, UK) report that abundance is assessed, but the measurement is of relative, rather than absolute, abundance of diatom taxa alone. Relative abundance is assessed by IE and UK but neither regard this as an assessment of abundance within the meaning of the NDs. The relationship between taxonomic composition, abundance and ecological status was assessed by IE and UK as part of a joint project. The results of this project revealed a relationship between EQR and the upper 90th percentile of biomass measurements, suggesting that the trophic gradient determined the upper limit of biomass at a site but that other factors acted locally to reduce this (Kelly *et al*., 2006). These findings are broadly in line with those found in other studies (Bernhardt and Likens 2004, Pan *et al.,* 1999, Biggs & Close, 1989; Biggs, 1996) and suggest that routine evaluation of absolute abundance may not yield significant extra information about ecological status.

This suggests that the requirement for assessment of abundance as outlined in the NDs might be better served by macrophyte survey methods, particularly where these include macroalgae. Phytobenthos biomass is very spatially and temporally heterogeneous and therefore quantitative assessment is unlikely to yield detailed insights about ecological status at low or moderate pressure levels. However, at higher pressure levels, visually-obvious growths of macroalgae such as *Cladophora* are likely to be conspicuous, often at the expense of macrophyte diversity more generally, and routine assessment of such growths using straightforward survey techniques may well yield more useful information than quantitative assessment of phytobenthos abundance.

'Undesirable disturbances' are not defined any further in the WFD itself, but ECOSTAT (2005) defines an undesirable disturbance as: 'a direct or indirect anthropogenic impact on an aquatic ecosystem that appreciably degrades the health or threatens the sustainable human use of that ecosystem.' Only BE-F and NL consider this to be assessed as part of their national methods. Several of the examples of 'undesirable disturbances' listed in ECOSTAT (2005) relate to the effects of macrophytes and phytobenthos on other biological elements, however, it is difficult to differentiate between direct effects of the pressure gradient on these biological elements and interactions with other biological elements.

Similarly, assessment of 'bacterial tufts' are not included directly in any of the assessment systems evaluated here although four MS (BE-F, IE, PL, SE) include these growths in other parts of their overall assessment methods. Again, a precautionary approach to boundary setting should ensure that the probability of such growths should be minimal when ecological status is good or better.

The view of the CB GIG phytobenthos expert group is that if a precautionary approach to boundary setting is taken using other properties (e.g. taxonomic composition), then the probability of undesirable disturbances and bacterial tufts should be minimal when ecological status is good or better.

Table 2.2.2: Phytobenthos methods: compliance with WFD normative definitions. 1 = assessed as part of national metric; 0 = not included in national metric; -1 = assessed but not included in national metric.

MS	Taxonomic composition	Abundance	Undesirable disturbances	Bacterial tufts
AT	1	1	0	0
Comment	All algal groups or (in special cases) diatoms only	Only relative abundances in both cases. Non-diatom taxa in % of total algal coverage (in sum 100%) and diatom taxa in % of 500 enumerated valves (in sum also 100%).		
Be-F	1	1	1	-1
Comment	But only diatoms. Filamentous algae are considered in macrophyte assessment	But only as relative abundance.	Good/moderate boundary relates to occurrence of average BOD values above 4 mg l^{-1}, indicating that self-purification capacity is exceeded.	Negative appreciation included in macrophyte assessment.
Be-W	1	1	0	0
Comment	National method is based on species composition characterized by polluosensitivity degree of each taxa	Relative abundance of diatom taxa.		
DE	1	1	0	0
Comment	All algal groups Metrics for Diatoms and non-diatoms	Relative abundances in case of Diatoms. Diatom taxa in % of 400 enumerated objects (in sum 100%). Non-diatom taxa as abundance class estimation according to Kohler(1978)		

MS	Taxonomic composition	Abundance	Undesirable disturbances	Bacterial tufts
EE	1	1	0	-1
Comment	Diatoms only.	The absolute number of counted taxa is converted to relative abundance.		Included in field inspection. Macroalgae are included in the macrophyte method.
ES	1	1 (relative abundance)	0	0
Comment	Estimated with a suite of metrics. See Table 2.2 for more details.	Estimated with percentage of sensitive species and some of the above metrics. See Table 2.2 for more details.		
FR	1	1	0	0
Comment	Species level of identification (diatoms only).	Only relative abundance	Each taxa included inside the national routine index (IBD) as in IPS gets a quality profile in 7 classes based on the sensitiveness-tolerance to undesirable disturbance (mostly organic, -trophic, salinity)	- Not included inside the diatom index. - Some diatoms (ex : Nitzschia umbonata) are as informative on the worst organic / trophic pollution levels as bacterial or fungal tufts - Anyway, included inside French macrophyte assessment tool (IBMR)
IE	1	0	0	-1
Comment	See comments for UK.	See comments for UK.		Included in field inspection.
LU	1	1	0	0
Comment	IPS formula is based upon each diatom polluo-sensibility and valence so as to characterize the sampled taxonomic composition.	IPS formula is also based upon relative abundances	IPS is sensitive to any disturbance	IPS is sensitive to any disturbance
NL	1	1	1	0
Comment	The national system is based on the presence of negative indicator species. Their relative abundance increases with increasing level of disturbance. Disturbance is the criterion for differentiating between status classes.	The method takes into account only relative and not absolute abundance.	See taxonomic composition.	

MS	Taxonomic composition	Abundance	Undesirable disturbances	Bacterial tufts
PL	1	1	0	-1
Comment	Species composition, namely the presence of sensitive or tolerant taxa is reflected by a value of a national metric.	Relative abundance of indicative species influences the national metric value.		Results of field inspection assessing bacterial tufts (if present) are noted in sample protocols.
SE	1	1	0	-1
Comment	Diatoms only.	Relative abundance of diatom taxa. Percent cover of all benthic algae noted on field protocol, and used in expert assessment of status class.		Noted in field protocol, used in expert assessment of status class.
UK	1	0	0	0
Comment	Diatoms only. The relationship between diatoms and other algae has been tested (Kelly et al., 2006b; Kelly, 2006). Macroalgae are included in the UK macrophyte method.	There is a negative relationship between EQR and abundance (as chlorophyll a concentration) but abundance is not measured routinely and was not used to set status class boundaries – see Kelly et al. (2006b).	Undesirable disturbances have not been considered.	Bacterial tufts have not been considered.

Evaluation of taxonomic composition

The main focus of all metrics used within CB GIG is taxonomic composition and a variety of approaches have been adopted (Table 2.2.2). Several national systems base their assessment systems on existing metrics based on weighted averaging although a few MS have developed new methods for the WFD based on the relative abundance of positive and negative indicator species, often determined type-by-type and sometimes in conjunction with parallel assessment based on weighted average metrics. Table 2.2.3 summarises the national assessment methods being used by each MS; a comprehensive description of each MS national metric is stored in a restricted access folder on the EEWAI CIRCA website.

Table 2.2.3: Member State (MS) national metric/assessment methods for phytobenthos intercalibration.

MS	National metric
AT	Multimetric method consisting of 3 modules/metrics: A) trophic status module (based on TI: Rott et al. 1999) B) saprobic status module (based on SI: Rott et al. 1997) C) reference species module (portion of defined reference and bioregion-specific species in total abundance and species number) Ecological status is evaluated separately for each of the modules and overall phytobenthos classification is equivalent to the worst of the three results (worst-case-scenario).
BE-F	Proportions of Impact-Sensitive and Impact-Associated Diatoms (PISIAD) (Hendrickx & Denys, 2005)
BE-W	IPS (Coste, in CEMAGREF, 1982 ; Lenoir & Coste, 1996)
DE	Diatom Module: WFD Diatom Index = Average of the sum of abundances of type specific reference species (following Schaumburg et al. 2005) and Trophic Index (Rott et al., 1999) or (in one special case) Saprobic Index (Rott et al., 1997). Additional metrics are available for cases of acidification or salinisation. Non Diatom Module: WFD Reference species Index depends on type specific taxa and abundances (following Schaumburg et al. 2005) Macrophyte Module: WFD Reference species Index depends on type specific taxa and abundances (following Schaumburg et al. 2005). Additional metrics are available for cases of mass growth stands of special taxa. Ecological status is calculated and classified from the average of the three module scores. If a module is absent, status class can be calculated with two moduiles or, exceptionally, with a single module. For this reason every module is classified separately and can be considered separately for intercalibration purposes. The national classification system needs all modules of the benthic flora occurring in a monitoring section of a water body.
EE	IPS (Lenoir & Coste, 1996)
ES	MDIAT (Diatom multimetric). composed by simple average addition of six indices calculated using OMNIDIA (SHE +SLAD+IDG+TDI+IPS+L&M) and 2 sensitive taxa metrics constructed with the reference diatom community of small and medium rivers in Galicia (NWSpain) (FPSS+PABSS). Note: SLAD: Slàdecek (1986); SHE: Schiefele & Schreiner (1991); IDG: Coste & Ayphassorho (1991); IPS: Coste in CEMAGREF (1982); L&M: Leclercq and Maquet (1997); TDI: Kelly & Whitton (1995); FPSS: % richness of sensitive taxa (Developed for Galicia,); PABSS: % abundante of sensitive taxa (Developer for Galicia)

MS	National metric
FR	IBD (national routine index: Lenoir & Coste, 1996, french-normalized AFNOR NF T90-354, 2000) IPS (one of the reference indices included inside the ICM (Coste, in CEMAGREF, 1982)
IE	Revised form of Trophic Diatom Index (TDI) (Kelly et al., 2006)
LU	IPS (Coste, in CEMAGREF, 1982)
NL	EKR (Van der Molen, 2004)
PL	Average of Trophic Index (Rott et al., 1999) and Saprobic Index (Rott et al., 1997)
SE	Swedish assessment method, Swedish EPA regulations (NFS 2008:1) based on IPS (Coste, in CEMAGREF, 1982).
UK	Revised form of Trophic Diatom Index (TDI) (Kelly et al., 2006)

2.2.3 Reference conditions and class boundary setting

Placement of status class boundaries

The metrics used by MS convert the response to a pressure gradient into a continuous variable which then has to be converted into an EQR, computed from Observed (O) and Expected (E) values. MS adopted a variety of approaches to split this EQR scale into separate status classes. Table 2.2.4 summarises these approaches.

The NDs define high, good and moderate status in terms of their deviation from the biota expected at the reference state and, therefore, a national method, if it is to be compliant with the NDs, has to be able to express each status class in terms of change from the reference state.

Table 2.2.4: **Rationales for defining Member State status class boundaries.**

	High / Good Boundary	Good / Moderate Boundary
AT	25th percentile of high class TI values (all values of Austrian WFD dataset lying within the defined type-specific trophic reference class - based on TI classes according to Rott et al. (1997) recalculated to give an EQR.	Measure of deviation from reference state are the trophic classes according to Rott et al. (1997). So the Good / Moderate Boundary corresponds to the upper TI boundary of the next worse trophic class following the type-specific trophic reference class - recalculated to EQR. For example, in type R-C3 where the trophic reference state is 'meso-eutrophic' (TI < 2,25), a good status sample must lie within the eutrophic status class (TI < 2,65 or EQR > 0,41 respectively).
Be-F	Relative abundance of impact-sensitive diatoms is not reduced from what can be expected for the type in unimpacted	Relative abundance of impact-associated diatoms is not higher than what can be expected for the type with slight human

	High / Good Boundary	Good / Moderate Boundary
	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-F region and for which the relative abundance decreases distinctly if at least one of the pressures affecting the respective water type increases (acidification, alkalinisation, eutrophication, organic pollution, salinisation, impoundment,...). The minimum relative representation of these taxa corresponding to high status for each water type is set initially by expert judgment. Impact-sensitive and indifferent taxa dominate at high status, whilst the abundance of impact-associated taxa remains very limited.	impact. Impact-associated taxa are listed for each water type, separately. Each list includes those diatom taxa that have been reported from water courses in the BE-F region and for which the relative abundance increases distinctly if at least one of the pressures affecting the respective water type increases. The maximum relative representation of these taxa corresponding to good status is estimated from its relation to pressure-related variables. Impact-sensitive and indifferent taxa dominate at good status and the abundance of impact-associated taxa remains limited (current boundary 20 %).
	The difference between the proportion of impact-sensitive taxa at high status and the lower limit of good status may be substantial.	
Be-W	17/20	13/20
	We carried out correlations between different values of high/good and good/moderate boundaries and chemical quality using SEQ-Eau index, we noted that in fact the values of 13/20 and 17/20 for good/moderate and high/good status boundaries showed strong correlations with SEQ-Eau indices measuring organic pollution and nutrient enrichment.	
DE	Type specific lists of species (reference, degradation including nutrient loading, tolerant) were made. The indices described for DE in Table 2.2 were sub-divided according to the NDs so that reference species dominated at high status whilst degradation indicators were either absent or occurred in very low numbers.	In good status reference and tolerant species are abundant, degradation indicators occur. In moderate status degradation indicators dominate over reference species.
EE	The high/good boundary is set at an EQR value corresponding to 90% of the EQR of reference sites.	The high/good boundary is set at an EQR value corresponding to 70% of the EQR of reference sites.
ES	The crossover between SHE, SLAD, IPS, IDG & L&M and the sensitive species metrics (PABSS and PFSS) is the centre of the high status classes, and the 0.93 marks the boundary of the high status class.	The crossover between the TDI and the sensitive species metrics (PABSS and PFSS) is the centre of the good class, and the crossover between the TDI and the other metrics (SHE, SLAD, IPS, IDG & L&M) is the centre of the moderate class. The good / moderate boundary is equidistant between these points (0.70).
FR	25th percentile of reference values for IBD or IPS (for every diatom-derived biotype covering all the national river	H/G boundary – [(H/G – minimum note)/ 4] +1 (for every diatom-derived biotype covering all the national river

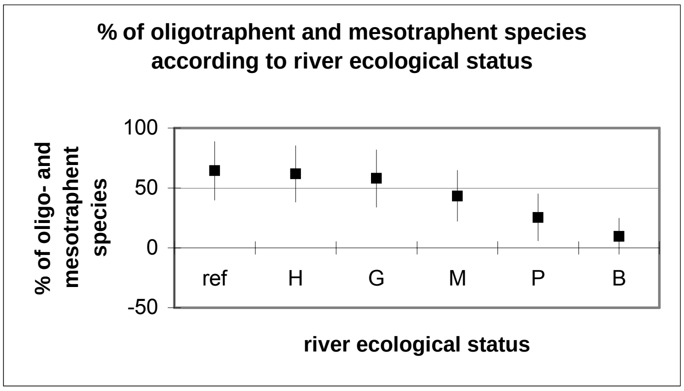
	High / Good Boundary	Good / Moderate Boundary
	types)	types)

The good/moderate boundary was calculated using a two step procedures (this procedure based on diatom-derived biotypes to define the provisional threshold values of the good ecological status of French river (ministerial circular DE/MAGE/BEMA 05 n°14 of the 28th July 2005):
1: For each type, the remaining range below the H/G boundary and the IBD minimum value was split into 4 equal classes to derive a preliminary G/M boundary, following a procedure proposed in the REFCOND guidance.

2: This preliminary boundary was then increased by 1 point on the IBD scale for all national types.

This procedure of boundaries calculation was chosen to be congruent with the French macroinvertebrates approach.

The IBD values obtained were then checked to verify their compliance with normative definitions: the graph below shows the percentage of sensitive species (‘oligotraphent’ + ‘mesotraphent’ species: van Dam *et al*., 1994) in reference conditions and along the ecological status gradient.



This graph shows:

-no significant difference in sensitive species % between reference conditions and high status;
-a very slight but significant decrease of sensitive species between high and good status;
-a drop in the percentage of sensitive species between good and moderate status.

IE	The high/good boundary is set at the 75th percentile of EQR values for reference sites within a particular type. See comments for UK	‘Crossover’ between nutrient-sensitive and nutrient-tolerant species (Pollard and van de Bund, 2005).
LU	85% of the median value of the evaluation criteria for reference sites (which sets the	70% of the median value of the

	High / Good Boundary	Good / Moderate Boundary
	EQR =1).	evaluation criterion for reference sites
	Those criteria have the advantage of not being influenced by occasional low values that can occur among reference site index values (Wallin *et al*., 2003).	
NL	Not more than 10% negative indicator species.	Not more than 30% negative indicator species.
	In the reference state, the presence of ca 5% negative indicators is considered to be more or less natural. An extra 5% is deemed to be almost equivalent to undisturbed conditions. The value of 30% negative indicators is considered to be a 'slight deviation' from high status.	
PL	As the 95th percentile of an average of TI and SI reference values, expressed as EQR = 0.814	As EQR = 0.600
SE	High/good boundary: IPS=17,5 High status: River/stream fulfils the national reference criteria, e.g. Tot-P < 10 µg/l or no eutrophication (arealspecific loss of Tot-P = class 1; in case of missing data for calculation of arealspecific loss: Tot-P < 20 µg/l if colour > 100 mg Pt/l), no acidification, land use: < 20 % farming, < 0,1 % urban area.	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 ca. 30 % (and the amount of sensitive species falls below ca. 30 %).
UK	The high/good boundary is set at the 75^{th} percentile of EQR values for reference sites within a particular type.	'Crossover' between nutrient-sensitive and nutrient-tolerant species (Pollard and van de Bund, 2005).
	Biological metrics tend to show gradual change as the level of nutrient/organic pressure increases, with no distinct discontinuities that could act as criteria for setting class boundaries. An alternative approach – based on the proportions of nutrient-tolerant, nutrient-sensitive and indifferent taxa within samples – was used to define status class boundaries in the UK, with the good/moderate boundary set at the point where the proportion of sensitive taxa falls below that of tolerant taxa. In ecological terms, the diatom flora at high and good status is characterised by a number of taxa, often with relatively broad niches (e.g. *Achnanthidium minutissimum, Fragilaria capucina*) which occur at different phases of a microsucession from colonisation of bare rock up to a mature biofilm (see Biggs *et al*., 1989). At high status, these are accompanied by other nutrient-sensitive taxa but as nutrient concentrations increase, the most sensitive of these taxa disappear whilst the numbers of nutrient tolerant taxa increases. The 'crossover' is, therefore, the point at which the taxa which form the 'association' characteristic of a site in the absence of pressure become subordinate to taxa which are favoured by a pressure (nutrients, in this case). The EQR gradient below the good/moderate boundary is then divided into three equally-spaced portions from which the moderate/poor and poor/bad boundaries are derived.	

Test datasets

All data required for the IC exercise was stored in a central relational database, managed by SEPA (UK). The database comprises three main components: raw diatom data, supporting chemical data and sample information. A summary of the number of sites available in each quality class (including reference sites) from each MS is presented in Table 2.2.5. In the CB GIG, six common river types were defined (Table 2.2.1). The CB GIG phytobenthos datasets incorporates data for all six common IC river types, even though the expert group took the decision not to intercalibrate using these common river types.

Table 2.2.5: Summary of the number of samples available in each quality class (including reference sites) from each MS.

Member State	H	G	M	P	B	Total
AT	19	279	167	51	3	**519**
BE-F		10	26	30	14	**80**
BE-W	26	250	120	47	24	**467**
DE	8	11	22	11	1	**53**
EE	55	4	2	1		**62**
ES	40	57	41	6		**144**
FR	32	57	140	64		**293**
IE	14	16	16	4	1	**51**
LU	97	34	41	24	6	**202**
NL	26	57	32	17	20	**152**
PL	8	4	9	5		**26**
SE	16	10	15	4	1	**46**
UK	314	211	377	139	10	**1051**
Total	**655**	**1000**	**1008**	**403**	**80**	**2095**

2.2.4 Results of the comparison

Standardisation of reference conditions
Introduction to Reference Conditions

The concept of 'type-specific reference conditions' is central to the WFD as ecological status is defined in terms of deviation from the biota expected under such conditions. Different interpretations of 'reference conditions' may lead to different values being used as the denominator in EQR calculations leading, in turn, to the same 'observed' biota having different ecological status assessments. On the other hand, the WFD also recognises that the 'expected' biota will vary from place to place depending on local factors such as climate, underlying geology and stream order and this too will have an effect on ecological status class boundaries. The challenge facing the IC exercise is to differentiate between those differences in national reference states that reflect genuine biogeographical variability across the GIG and those that reflect differences in approach by those responsible for implementation.

Evaluation of reference conditions and principles of setting classification boundaries within the GIGs assumes a cascade of effects, with alterations to catchments (removal of natural vegetation, replacement by agriculture or urban development) leading to increases in pressure variables in surface water which, in turn, affect the biota. Ideally, evaluation of reference conditions focuses on changes to the catchment, and incorporates data on land use and supports this with data on pressure variables (nutrients, BOD etc). The final approach – use of the biota to define reference conditions – is not encouraged as the NDs define ecological status in relation to the biota expected under undisturbed conditions (Annex V, article 1.2) and the use of land use and pressure data to define 'undisturbed conditions' ensures rigour and objectivity in the definition of the 'expected' value.

Member States adopted one of two approaches to define the 'expected' value: either using the median or the 95^{th} percentile of the metric values of reference samples. The former was more common and is a more stable property when the population of reference sites is small; however, one consequence is that a number of high status sites will have EQR >1. The latter approach is robust if the population of reference sites is large and means that the number of situations where EQR > 1 is smaller. Both are acceptable approaches and, in both cases, EQR values >1 can be automatically set to 1 for reporting.

Two MS (BE-F and NL) had no reference sites, due to an absence of streams in pristine condition. Both BE-F and NL have a reference concept based on theoretical, rather than actual, reference conditions (Denys, 2006).

The purpose of this chapter is to perform a multilateral comparison of all reference site data in order to determine whether reference conditions comply with the NDs and criteria set by REFCOND and CB GIG and to examine the extent to which differences in the reference state may influence the comparison of boundaries. An additional objective was to see whether the IC typology (Table 2.2.1) had any ecological validity.

Reference Data

Member States participating in the phytobenthos IC were asked to supply the raw biological data for all reference samples in their IC datasets (Table 2.2.6), along with information on how candidate reference sites were screened in relation to criteria established by REFCOND (Working Group 2.3 - REFCOND Guidance Document No 10.) and CB GIG (Table 2.2.7).

Data were analysed in two ways:

- Four widely-used weighted average metrics ('candidate metrics) were calculated for all samples, which were then plotted by IC type and by MS. The former indicated whether or not there were significant differences in the baseline conditions of streams between types whilst the latter allowed comparisons of national concepts.

- The biological data for all reference sites were submitted to the ordination technique Detrended Correspondence Analysis (DCA: Hill, 1979) after taxonomic differences within the national datasets were resolved. Again, outputs were plotted by both IC type and by MS.

Table 2.2.6: Number of reference samples by Member State and intercalibration river type.

Member State	R-C1	R-C2	R-C3	R-C4	R-C5	R-C6	Total
AT			7				7
BE-F							0
BE-W			14	9	13	3	39
DE			7	1			8
EE				12			12
ES			10				10
FR	4	1	13	7		7	32
IE		4		8			12
LU			44			10	54
NL							0
PL			6				6
SE	5	8	2	1			16
UK	22	5	4		1		32
Total	31	18	107	38	14	20	228

The phytobenthos expert group adopted an approach that is consistent to the CB GIG macro-invertebrate working group with regards to defining what is meant by reference conditions. Reference sites were initially chosen by Member States using REFCOND guidance (Working Group 2.3 - REFCOND Guidance Document No 10.). A list of the more detailed criteria and type-specific concentrations ("reference thresholds") of key chemical parameters were developed by the macro-invertebrate working group.

Table 2.2.7 below outlines the chemical reference thresholds used for reference screening in CB rivers GIG. Both mean values and 90-percentile values have been proposed for some parameters. The mean is the most robust statistic when few data are available, as is frequently the case for new reference sites. The 90th percentile should be used only when sufficient data are available (at least 12 monthly chemical samples).

The proposed reference thresholds allow the same criteria to be applied to the selection of all reference samples used in the IC exercise in CB rivers GIG and were intended for use in conjunction with other general pressure criteria. The thresholds aim to interpret the WFD requirement of "very minor anthropogenic impact".

The thresholds were principally derived from datasets linking invertebrates to general chemical elements, but other values taken from national water quality classifications, diatoms datasets (in the case of nutrients), specific studies and expert opinions were also considered. In general, the available information was not sufficient to derive type-specific reference thresholds for all types. Wasson (2006) outlines the different methods used to establish chemical threshold values that correspond to "no or very minor impact on biological quality elements".

Table 2.2.7: Chemical reference thresholds defined by CB GIG for reference screening.

	R-C1	R-C2	R-C3	R-C4	R-C5	R-C6
BOD (m/l)						
Mean	2.4	2.4	2	2.4	2.4	2.4
90th percentile	3.6	3.6	2.75	3.6	3.6	3.6
Dissolved Oxygen (% saturation)						
Mean	95-105	95-105	95-105	95-105	95-105	95-105
10th-90th percentile	85-115	90-110	90-110	85-115	85-115	85-115
N-NH4 (mg/l)						
Mean	0.1	0.05	0.05	0.1	0.1	0.1
90th percentile	0.25	0.12	0.12	0.25	0.25	0.25
P-PO_4 or SRP (µg/ l)						
Mean	40	30	20	40	40	40
N-NO_3 (mg / l)						
Mean (invertebrates)	6	6	2	6	6	6
Mean (phytobenthos)	4	4	2	4	4	4

Reference screening procedures

Representatives from each MS were asked to screen reference sites, chosen using REFCOND guidance (refer to **Appendix A**), against agreed catchment land use and chemical reference thresholds. Table 2.2.8 indicates which of the GIG defined reference criteria were used for the screening exercise and what sources of information were available to each MS for this purpose. Member States were also asked to indicate if they used more stringent criteria (or different but equivalent ones).

A reference screening flow chart (refer to Appendix A) illustrates the screening process and how reference and rejections thresholds (only available for some parameters) should be interpreted by MS. Instructions issued to MS indicated that not all of the reference criteria had to be fulfilled for each reference site, but all the pressures acting on a site should be taken into account for at least one of the criteria. In cases where some (<10%) of the reference criteria exceeded the equivalent reference thresholds an evaluation of cumulative pressures by expert judgement was used to validate reference sites. A database of MS reference samples is stored in a restricted access folder on the EEWAI CIRCA website.

Table 2.2.8: Criteria used by Member States to select reference sites. Key: 0: missing info; 1: not used; 2, Yes, Measured; 3, Yes, Estimated; 4, Yes, Field inspection; 5, Yes, Expert judgement. See Appendix A for reference to details on national screening procedures.

	Landuse data (e.g. CORINE)	BOD_5	O_2	N-NH_4	Phosphorus fractions	N-NO_3	Comments

AT	1	2	1	1	2	1	
BE-F	2	2	2	2	2	2	
BE-W	1	2	2	2	2	2	
DE	3	2	2	2	2	2	Hydro morphological degradation, biological data, expert judgement
EE	0	1	1	2	2	1	TN
ES	2	2	2	2	2	2	REFCOND criteria used for invertebrate exercise
FR	2	2	2	2	2	2	
IE	3	2	2	2	2	2	2
LU	3	2	2	2	2	2	A land use Index was set from ministry of environment CORINE data
NL	5	5	5	5	5	5	5
PL	3	2	1	0	2	0	2
SE	2	1	1	1	2	1	Assessment of acidification
UK	1	3	3	3	2	2	2

Evaluation of IC typology
Raw values of candidate metrics showed significant differences between the IC types (Fig. 2.2.1), with R-C2, in particular, showing higher values of IPS and lower values of the other three metrics, compared with other types. R-C6 showed the opposite trend, though less pronounced, whilst there were too few sites within R-C5 to draw meaningful conclusions. However, the IC-types are not evenly represented between MS and Fig. 2.2.1 needs to be considered alongside Figs 2.2.4 – 2.2.7, which compares metric responses between MS. R-C6, in particular, has 18 samples divided between three MS (BE-W, FR, LU – see Table 2.2.6) and it is, consequently, difficult to separate elements of the response of R-C6 samples that are type-specific from those that are due to national interpretations of the typology.

An ordination of these data shows no clear separation of any of the IC types, based on the diatom flora at reference (Fig. 2.2.2). Although R-C2 samples tend to have relatively low scores on Axis 1, the distribution of R-C2 samples in the ordination space overlaps with those of the other types. R-C3, the type with the largest number of samples, is distributed throughout the ordination plot. The first axis of the ordinationis strongly correlated with the candidate metrics (Table 2.2.9), suggesting that there is a strong nutrient / organic gradient within the reference community. . Sites with low scores on axis 1 have higher scores of trophic metrics (TDI, TI) in particular (Fig. 2.2.3). IPS, SI and TI are also correlated with axis 2. However, the ordination explained a relatively small part of the total variation within the diatom assemblage, suggesting that other factors were also responsible for shaping diatom assemblages in rivers. Overall, these data suggest that the IC typology has no meaning for phytobenthos and subsequent analyses ignore the IC types.

Table 2.2.9: Summary statistics for Detrended Correspondence Analysis of all reference sites, along with correlations with candidate intercalibration metrics.

	Axis 1	Axis 2	Axis 3	Axis 4
Eigenvalue	0.727	0.452	0.401	0.392
Decorana value	0.853	0.536	0.470	0.421
Axis length	6.608	4.734	4.286	4.569
Variance explained (%)	4.780	3.002	2.633	2.356
Correlations				
TDI	-0.757***	-0.077	-0.390***	-0.297***
IPS	0.521***	-0.303***	0.086	0.272***
TI	-0.498***	0.247**	-0.116	-0.396***
SI	-0.704***	0.207**	-0.216**	-0.214**

Significance level:
P<0.05: *, P<0.01: **, P<0.001: ***

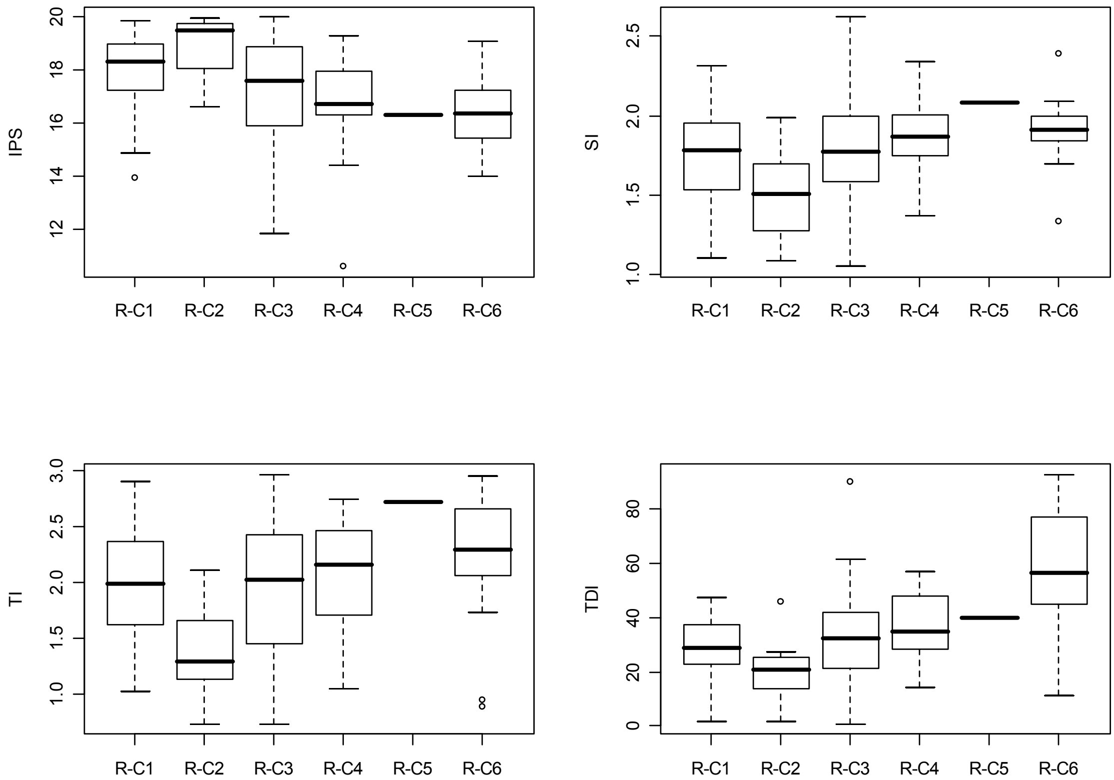


Fig. 2.2.1: Variation in values of candidate metrics between intercalibration (IC) types. Differences between IC types are significant at $p < 0.001$ for IPS, TI and TDI and at $p < 0.01$ for SI.

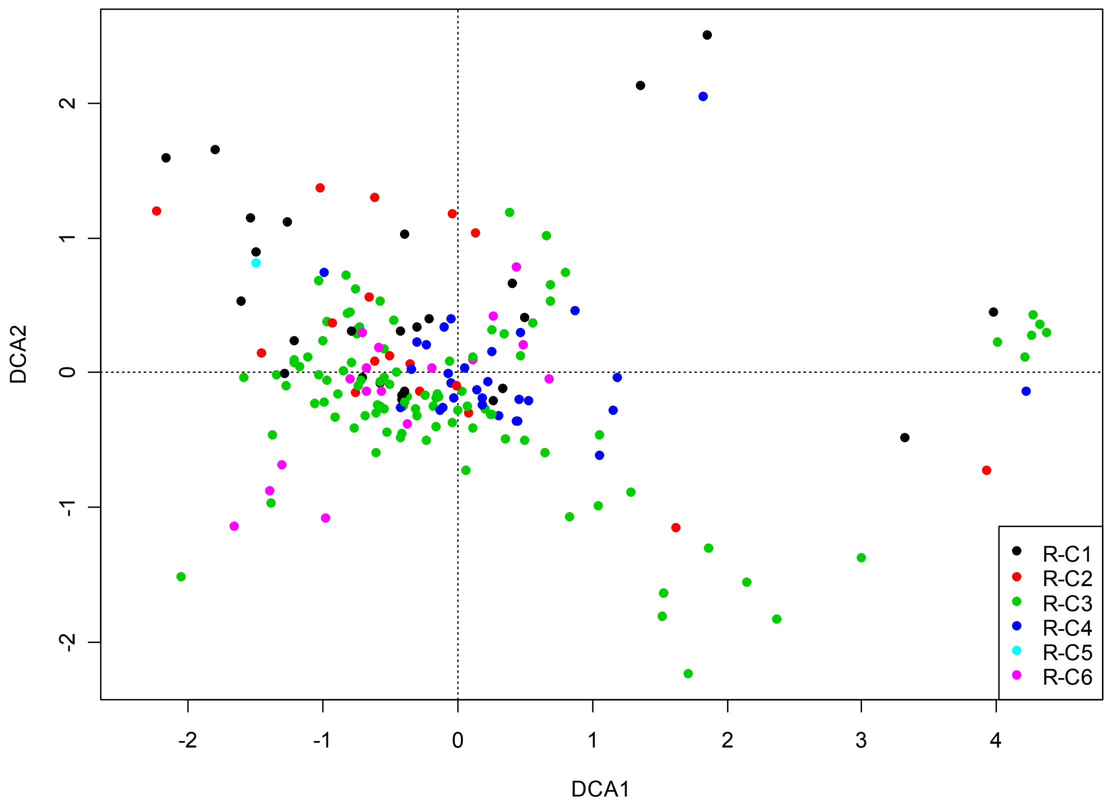


Fig. 2.2.2: Detrended correspondence analysis (DCA) of 190 phytobenthos reference samples from Central Baltic GIG, plotted by intercalibration type.

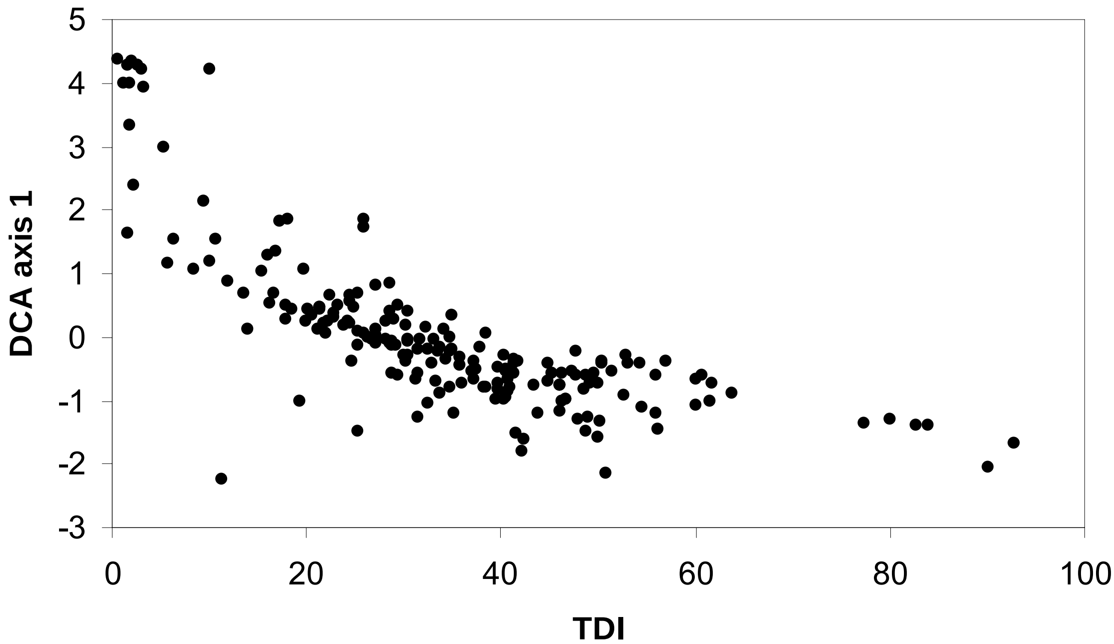


Fig. 2.2.3: Relationship between trophic diatom index (TDI) and the first axis of the DCA illustrated in **Fig. 2.2.2.**

National differences in reference sites

Figs. 2.2.4-2.2.7 show the variation in candidate metric values between Member States, ignoring both IC and national typologies. BE-W, ES, PL and SE tended to have lower values for SI, TI and TDI (higher values for IPS) than other Member States, whilst EE and LU tended to have higher values for SI, TI and TDI (lower values for IPS). Other Member States were neither consistently high nor consistently low.

The mean value of the TI was 1.92, which means that variation in reference samples alone extends across about 50 per cent of the entire metric scale. The TI was designed to be particularly sensitive to inorganic nutrients, and the mean value of the IPS, a metric which operates across a longer nutrient/organic gradient was 17.4, although LU had a mean value of 15.8 and one LU reference sample had an IPS value of 11.9.

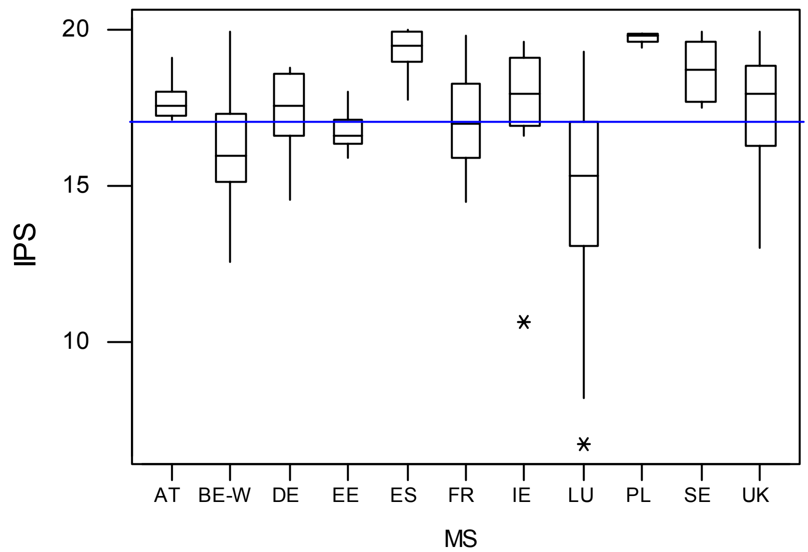


Fig. 2.2.4: Variation in Indice de Polluosensibilité (IPS) values for reference samples between Member States participating in the phytobenthos intercalibration exercise. Mean value (blue line): 17.3.

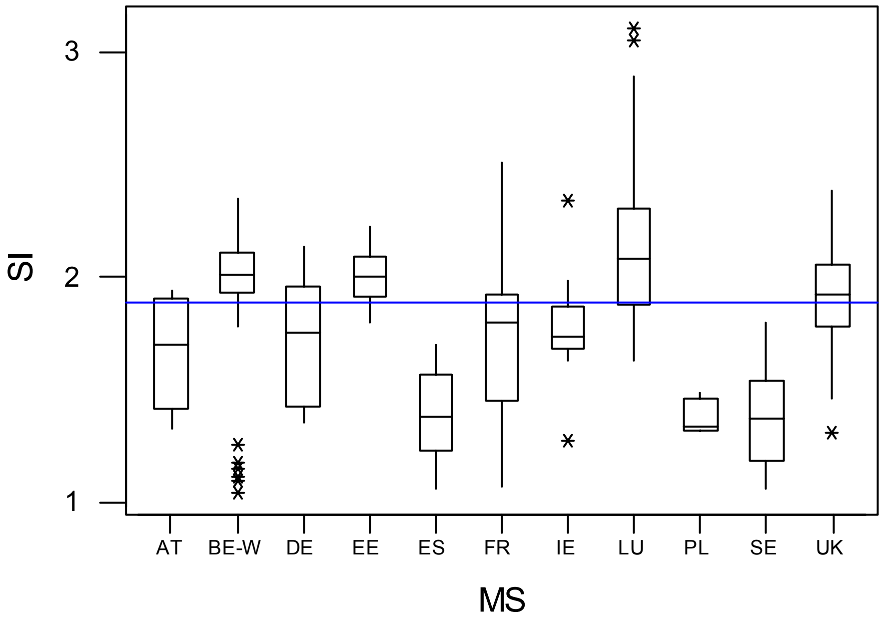


Fig. 2.2.5: Variation in Saprobienindex (SI) values for reference samples between Member States participating in the phytobenthos intercalibration exercise. Mean value (blue line): 1.9.

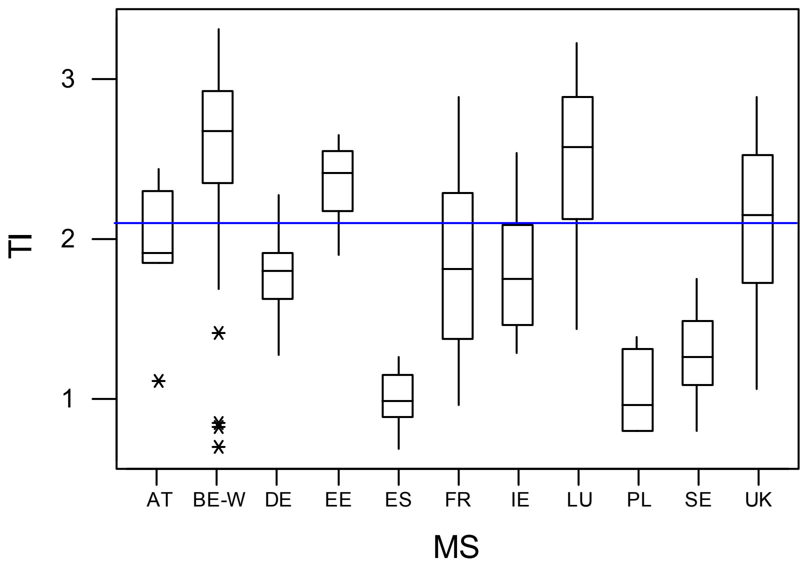


Fig. 2.2.6: Variation in Trophienindex (TI) values for reference samples between Member States participating in the phytobenthos intercalibration exercise. Mean value (blue line): 2.1.

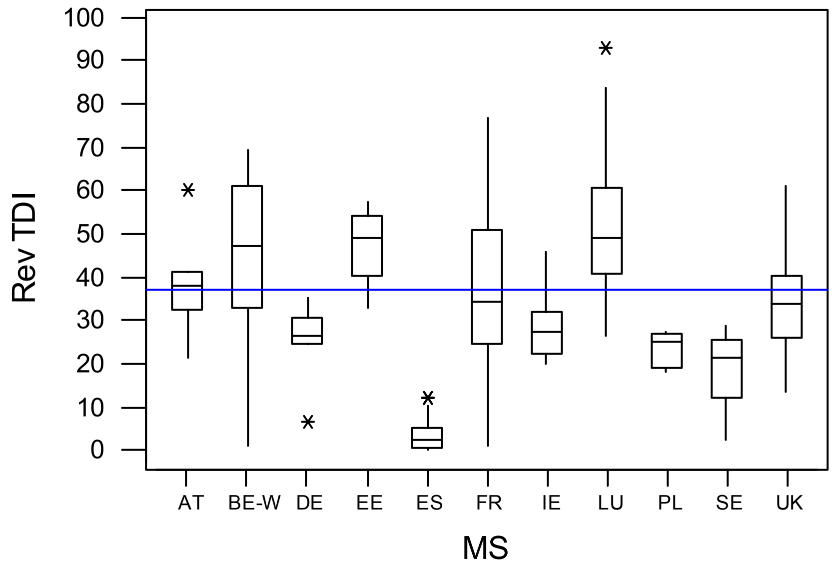


Fig. 2.2.7: Variation in Trophic Diatom Index (TDI) values for reference samples between Member States participating in the phytobenthos intercalibration exercise. Mean value (blue line): 36.9.

Conclusions

- Member States used a variety of approaches to screen candidate reference sites in order to ensure an absence of pressures.

- The IC typology failed to discriminate between reference sites on the basis of their diatom floras and, for this reason, the phytobenthos IC exercise has been performed without any differentiation into types.

- There was considerable variation in the values of four metrics computed at reference sites between MS. It is not clear from this exercise whether these differences are due to underlying differences in the unimpacted state between MS or whether they reflect failures to screen data adequately.

- There is considerable scope for refining this exercise in the future. In particular:

    i. A means of validating national screening procedures needs to be introduced;

    ii. There is scope for developing a more realistic typology for phytobenthos which will improve the resolution of future IC exercises.

Both of these steps were not possible during the present exercise due to resource constraints. There were, in particular, difficulties in obtaining environmental data in comparable formats (annual means vs spot measurements for chemical determinands, total versus ‘available’ fractions of nutrients). For the present exercise, all reference data were accepted at face value, and the implications will be addressed in more detail in the Section 2.2.6.

Development of Common Metric

Evaluation of Candidate Metrics

In order to compare status class boundaries developed in each MS, national metrics first had to be converted to a common scale. The mechanism for doing this was to develop an 'intercalibration common metric' (ICM) (corresponding to Option 2 outlined in the Boundary Setting Protocol) similar to that developed for the CB GIG invertebrate IC exercise (Buffagini *et al*., 2005). This ICM needs to have a statistically-significant relationship with all of the national metrics so that EQR values computed using national metrics can be quoted as the corresponding value of the ICM. In the case of phytobenthos, there was a high degree of congruence between national methods, sharing common sampling and analysis methods (CEN, 2003, 2004; Kelly *et al*., 1998) and, in most cases, based wholly or partly on the weighted average (WA) equation of Zelinka and Marvan (1961).

Results of an initial comparison between national metrics and the four candidate metrics in common use within the GIG are shown in Fig. 2.2.8. These metrics are the Indice de Polluosensibilité (IPS: Coste, in CEMAGREF, 1982); Trophienindex (TI: Rott et al., 1999) and Saprobienindex (SI: Rott *et al*., 1997) and the Trophic Diatom Index ([revised]-TDI: Kelly & Whitton, 1995; Kelly *et al*., 2001). What is striking is that all four candidate ICMs showed high correlations with some of the national metrics, but also that there were instances where each of the candidate ICMs had very low correlations with national metrics. The conclusion of this preliminary exercise was that no single metric was likely to fulfil the requirements of an ICM.

The four candidate metrics showed two types of response along the pressure gradient, with two (TI, TDI) being particularly responsive at low levels of nutrient / organic pressure (moderate to high EQRs) and the other two (IPS, SI) being more responsive at higher pressure levels (low to moderate EQRs). Rather than use any metric in isolation, a simple multimetric, composed of two of the candidate ICMs was tested. The TI was chosen over the TDI as the 'sensitive' metric as this had a slightly better performance when compared to ambient nutrient concentrations, whilst the IPS was chosen over the SI as the complementary metric as this metric was already widely used as a national metric within the GIG. The metrics were converted to EQRs as follows:

IPS: this metric measures 'general water quality', with low values corresponding to high pressure levels and, therefore, low EQRs. Therefore:

EQR_IPS = Observed (O)/Expected (E),

where: Expected = median IPS value of reference sites for a national dataset. Different reference values for each national type could be used, if appropriate, and the two MS without reference sites used expert judgement to select reference sites from neighbouring countries (the latter will not be included in the calculation of a mean reference value based on all MS data).

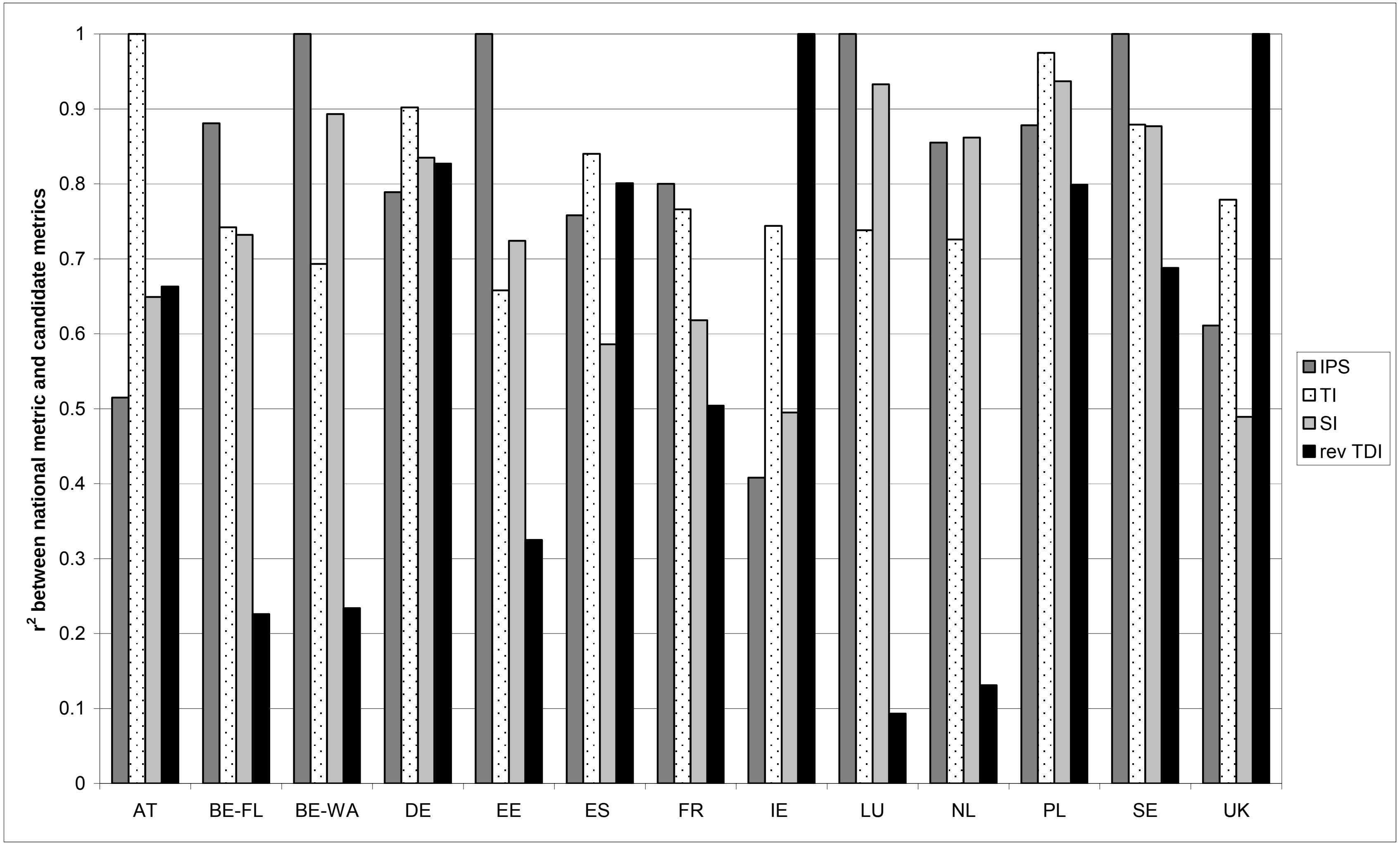


Fig. 2.2.8: Performance of four candidate ICMs against national metrics expressed in terms of the coefficient of determination, r^2. Where a candidate ICM is also a national metric, $r^2 = 1$.

TI: as this is a trophic index it needs to be adjusted so that high values represent high EQR values, therefore,

EQR_TI = (4-O)/(4-E)

(4 is the maximum possible value of the TI). Expected values were calculated as for EQR_IPS.

Two options for combining the metrics were considered: where metrics indicating the same stressor are combined in a multimetric index, then the average of these metrics is the most appropriate value to use (based on the assumption that it shows the stronger relationship across the entire gradient). However, if the metrics indicate different stressors, then the minimum value of the two metrics would be appropriate. The response of the TI and IPS to a nutrient / organic gradient is assumed to be a composite of a number of ecophysiological processes, with interspecific competition for inorganic nutrients prevailing at low pressure levels (high EQRs) whilst factors such as tolerance to ammonia toxicity, capacity for heterotrophic growth and survival in environments with low oxygen concentration and redox potential prevailing at high levels of nutrient / organic stress (low EQRs). If this is the case, then the nutrient / organic gradient could be viewed as a combination of stressors, and the minimum of EQR_TI and EQR_IPS might be an appropriate measure. On the other hand, it is difficult to separate the effects of these stressors, in which case, the nutrient/organic gradient could be regarded as a single stressor, and it would be more appropriate to use the average of the two metrics. Both options are considered here.

Evaluation of the Intercalibration Common Metric

The performance of the ICM was evaluated using linear regression models. The objective is to predict values of the ICM from values of each national metric. This can be regarded as a conventional Model I regression with a dependent and independent variable. However, when both the response and explanatory variables of the model are random (*i.e.* not controlled by the researcher), there is error associated with the measurements of both *x* and *y* and Model II regression is more appropriate for the estimation of parameters associated with the regression itself (see Fig. 2.2.9). Both types of regression model were evaluated during this exercise (Model II regression using Reduced Major Axis regression routines in the R statistical package: R Development Core Team, 2005; Warton, 2005) with Model I being chosen for the final analyses to ensure compatibility with other intercalibration exercises.

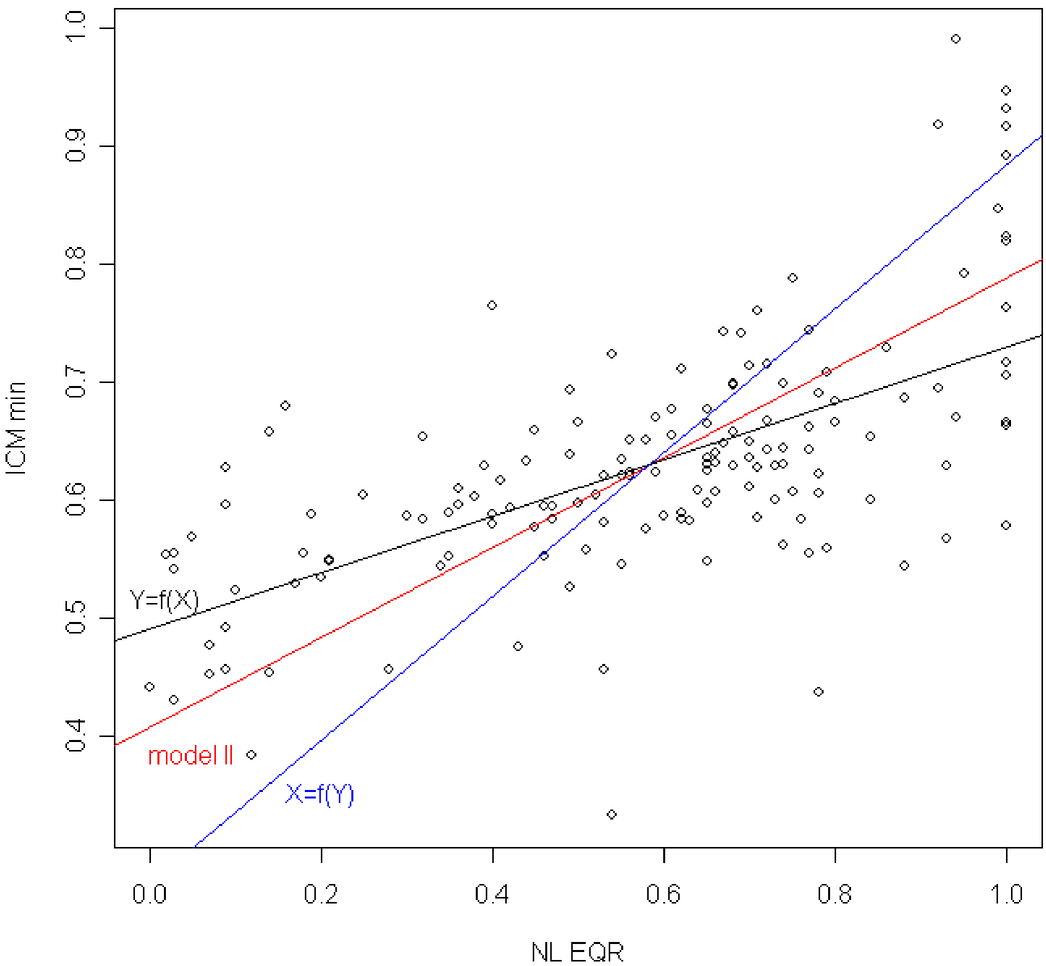


Fig. 2.2.9: Model I and model II regression equations for the Dutch dataset used in the phytobenthos intercalibration exercise. In model I regression, the sum of verticals squared deviations is minimised when regressing Y on X; the same thing occurs horizontally when regressing X on Y. In model II regression, the sum of squared Euclidian distances to the regression line is minimised. Fitted results are the same from X to Y as well as from Y to X with a model II regression; this is not the case for model I (generally).

Table 2.2.10 shows the performance characteristics for ICMs based on the minimum and mean values of EQR_TI and EQR_IPS. Four properties were used to evaluate the relationship:

- A visual examination of scatterplots to check for a linear response between the ICM and national metrics;
- The root mean square error (RMSE ≡ residual standard error: a measure of prediction error - Wallach & Goffinet, 1989);
- The coefficient of determination (r^2); and,

Table2.2.10: Performance characteristics of linear regressions between national metrics and the minimum ('min') and mean ('mean') intercalibration metric (ICM) (based on EQR_TI and EQR_IPS). (* = non-linear responses – see 4.2). Based on data available in July 2006.

Member State	RMSE		Slope		r^2	
	min	mean	min	mean	min	mean
AT	0.072	0.056	0.901	0.654	0.616	0.506
BE-F	0.130	0.111	0.840	0.886	0.591	0.686
BE-W	0.065	0.083	0.640	0.645	0.792	0.705
DE	0.091	0.086	0.694	0.885	0.687	0.803
EE *	0.037	0.083	1.021	1.197	0.888	0.685
ES	0.0116	0.083	1.034	0.874	0.673	0.743
FR	0.105	0.122	0.668	0.826	0.621	0.653
IE	0.123	0.096	0.527	0.401	0.528	0.514
LU	0.110	0.079	0.622	0.719	0.540	0.752
NL	0.119	0.096	0.490	0.541	0.550	0.696
PL	0.037	0.062	1.067	1.030	0.983	0.951
SE	0.098	0.093	1.974	1.865	0.824	0.825
UK	0.095	0.061	0.379	0.233	0.349	0.323

- The closeness of the slope of a Model I regression of the ICM against the national metric to 1 (to maximise sensitivity of predictions across the entire EQR scale).

The coefficient of determination (r^2) measures association between two variables and gives little indication of the predictive power of that relationship. It is also dependent, to some extent, on the length of the gradient over which the coefficient is applied (see Fig. 2.2.10). RMSE, on the other hand, gives a better indication of the predictive power of the relationship, regardless of gradient length Using both, along with visual examination and slope, provides a robust basis for evaluating relationships between national metrics and the ICMs.

Overall, RMSE was lower using ICM (mean) though ICM (min) gave slopes closer to unity and higher r^2. However, examination of scatterplots showed fewer obvious deviations from linearity using ICM (mean).

Table 2.2.11 shows the relationship between ICM (min) and ICM (mean) and nitrogen and phosphorus fractions. Note that the primary purpose of an ICM is to allow values of national metrics to be compared, so the performance characteristics in Table2.2.10 are more instructive for the purposes of selecting an ICM but Table 2.2.11 helps illustrate the relationship between the ICMs and the underlying nutrient / organic gradient.

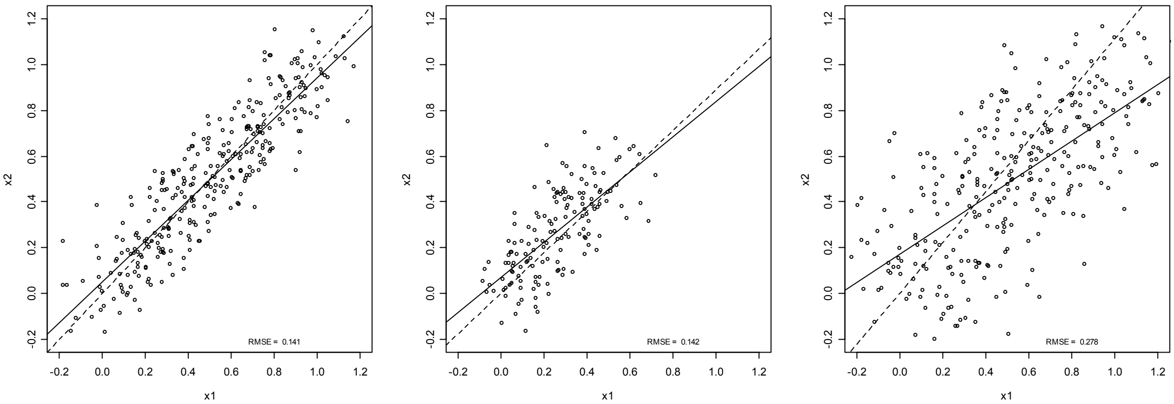


Fig. 2.2.10. Model I and Model II regressions for 3 hypothetical datasets. The left and central plot have the same amount of error and the same RMSE of 0.14, even though they cover different ranges (their product moment correlations are very different – 0.9 for left plot, 0.64 for centre). The right plot shows data with the same trend but double the error. In this case the RMSE = 0.28, r = 0.66. The central figure has a clearly better agreement between x1 and x2 than the right-hand one but their correlations are very similar.

No clear preference for one ICM over the other emerges from this: an ICM based on minimum values shows stronger correlations with some datasets (e.g. ES, FR, IE) but the ICM based on mean values shows stronger correlations in some other instances (e.g. PL, UK). Please note that the correlation between ICM and chemical values for Sweden was calculated from very few streams, which probably explains the low correlation.

Figs. 2.2.11 and 2.2.12 show results of Principal Components Analyses (PCA), performed in order to check the interrelationships between metrics and to evaluate differentiation between these metrics. Note the separation of status classes along the first axis in Fig. 2.2.11 and the close alignment of the two ICMs in the factorial map (Fig. 2.2.12). The results indicate that nearly 90% of the total variation (inertia) can be summarized by the first two eigenaxes of the PCA. 78.11% of the total variation is represented by the first axis whilst the second axis represents only 11.58%. As expected, the quality gradient is clearly identified along the first axis (Fig 2.2.11) and explains most of the total variation. This main trend is more or less artificial (because all variables are strongly linked), and therefore the interpretation of the results should focus on side trends. Thus, the close alignment of the two ICMs along the first axis of the factorial map (Fig 2.2.12) indicates that they are both the best correlated metrics with the quality gradient. Others metrics are also linked with the second eigenaxis. The sites scores on the factorial map (Fig 2.2.12) display a curve along the second eigenaxis that can be explained by the differences between the other metrics. TI and TDI seem to be more associated with High status sites, whereas SI and IPS seem to be more associated with Bad status sites. Good and Moderate status are less influenced by the indices because they are close to the origin of the factorial map. Furthermore, a linear trend is displayed from Bad status sites to Good status ones; the High status sites are principally responsible for creating the curve. In conclusion, SI and IPS seem to be more efficient in distinguishing lower quality sites including Moderate and Good status sites whereas TI and TDI are more efficient in separating High status sites from the others. National EQR values correlate best with IPS values but this may be because several countries have chosen this index for deriving their EQR values.

Overall, these evaluations suggest that the ICM based on the mean values of the EQR_TI and EQR_IPS is slightly better for use in the IC exercise than the ICM based on minimum values as both show similar trends between national datasets.

Table 2.2.11: Correlation coefficients between nutrients and the minimum ('min') and mean ('mean') intercalibration metric (ICM) (based on EQR_TI and EQR_IPS). 'DI-N' = dissolved inorganic nitrogen; 'TIN' = total inorganic nitrogen (≈ NO_3-N + NO_2-N 'SRP' = soluble reactive phosphorus (≈ PO_4-P).

Member State	Determinand	Data Type	ICM (min)	ICM (mean)
AT	Log DI-N	Spot	-0.155**	0.114
AT	Log NO_3-N		-0.098	-0.214***
AT	Log PO_4-P		-0.168**	0.317***
AT	Log Total P		-0.185**	0.348***
BE-F	Log Total P	Mean	-0.451*	-0.357
BE-F	Log PO_4-P		-0.44	-0.352
BE-W	Log NO_3-N	Mean	-0.182***	-0.301***
BE-W	Log NH_4-N		-0.658***	-0.709***
BE-W	Log NO_2-N		-0.639***	-0.672***
BE-W	Log PO_4-P		-0.641***	-0.693***
DE	Log NO_3-N	Spot	-0.508***	-0.758***
DE	Log Total P		-0.799***	-0.488***
EE	Log NH_4-N	Spot	-0.224	-0.216
EE	Log NO_2-N		-0.07	-0.089
EE	Log NO_3-N		-0.204	-0.217
EE	Log Total N		-0.12	-0.15
EE	Log PO_4-P		-0.429**	-0.447***
EE	Log Total P		-0.455***	-0.463***
ES	Log NH_4-N	Spot	0.453***	-0.329**
ES	Log NO_2-N		0.691***	-0.497***
ES	Log NO_3-N		0.435***	-0.355**
ES	Log PO_4-P		0.600***	-0.445***
FR	Log NH_4-N	Spot	-0.382***	-0.350***
FR	Log NO_2-N		-0.336***	-0.313***
FR	Log NO_3-N		-0.135*	-0.122*
FR	Log TIN		-0.375***	-0.343***
FR	Log PO_4-P		-0.455***	-0.444***
IE	Log NOx	Spot	-0.24	-0.146
IE	Log PO_4-P		-0.550**	-0.517**
LU	Log NO_2-N	Spot	-0.562***	-0.572***
LU	Log NO_3-N		0.094***	0.070
LU	Log NH_4-N		-0.457***	-0.423***
LU	Log Total P		-0.409***	-0.430***
LU	Log PO_4-P		-0.425***	-0.453***

Table 2.2.11 (cont.)

Member State	Determinand	Data Type	ICM (min)	ICM (mean)
NL	Log Total N	Spot	-0.510***	-0.498***
NL	Log NH_4-N		-0.348**	-0.302**
NL	Log NOx-N		-0.435***	-0.474***
NL	Log Total P		-0.557***	-0.565
NL	Log Soluble P		-0.533***	-0.517***
PL	Log Phosphate	Spot	-0.647***	-0.699***
PL	Log PO_4-P		-0.645***	-0.697***
PL	Log Total P		-0.663***	-0.739***
SE	Log NH_4-N	Mean	0.395	0.257
SE	Log Total N		-0.034	-0.072
SE	Log NOx		0.191	0.115
SE	Log Total P		0.238	0.208
SE	Log PO_4-P		0.33	0.295
SE	Log NH_4-N	Spot	0.045	0.061
SE	Log Total N		0.082	0.074
SE	Log NOx		0.114	0.118
SE	Log Total P		-0.247	-0.234
SE	Log PO_4-P		-0.135	-0.175
UK	Log NH_4-N	Mean	-0.272**	-0.312***
UK	Log NOx		-0.535***	-0.661***
UK	Log SRP		-0.469***	-0.517***

Significance level: P < 0.05: *; P < 0.01: **; P < 0.001: ***

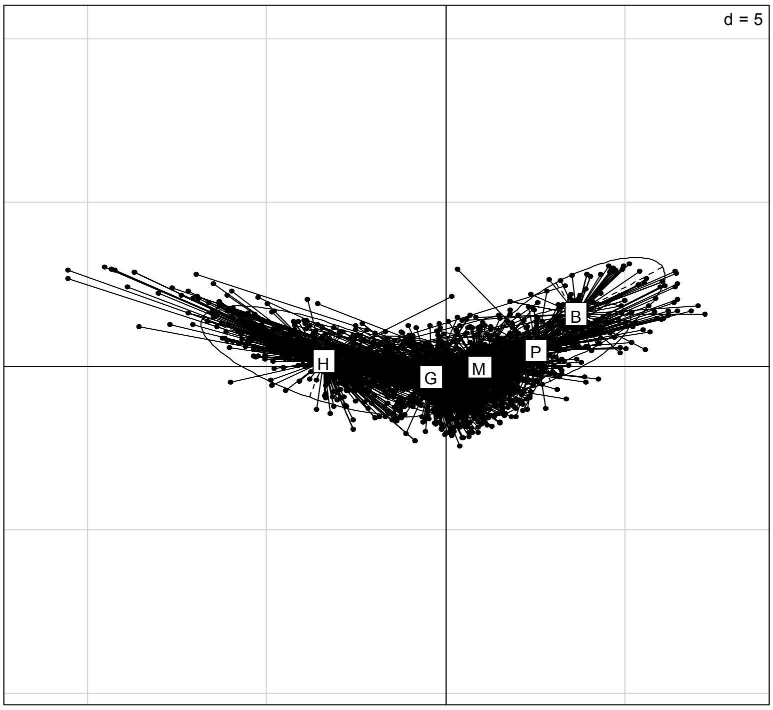


Fig. 2.2.11: Principal components analysis (PCA) of index values for all samples in the CB GIG intercalibration database (H: High, G: Good, M: Moderate, P: Poor, B: Bad).

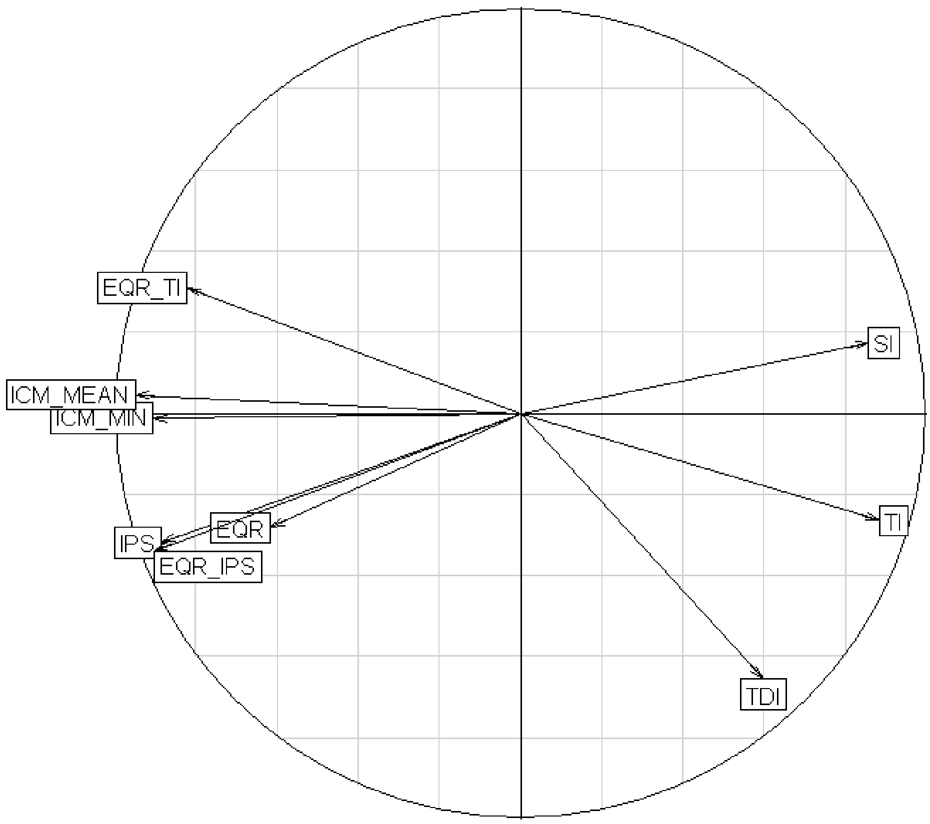


Fig. 2.2.12: Factorial map of a correlation circle, based on PCA presented in **Fig.2.2.11.**

Conversion of national metrics to the ICM

The procedure for converting values of the national metric representing the High / Good and Good / Moderate boundaries to corresponding values of the ICM is identical to that used in the CB GIG invertebrate IC exercise and is based on a linear regression equation:

ICM = a + b(national metric as EQR)

Where: a = constant; b = slope

For each MS, EQR values from the national assessment method were plotted against the corresponding EQRs from the ICM and the regression equation and associated statistics were calculated. Conspicuous outliers were removed prior to calculation of the regression equation. Figure 2.2.13 shows a regression between the EQR values of a national metric and the ICM for a hypothetical national dataset. Figure 2.2.14 illustrates the process of converting the national value of the Good/Moderate boundary to the ICM.

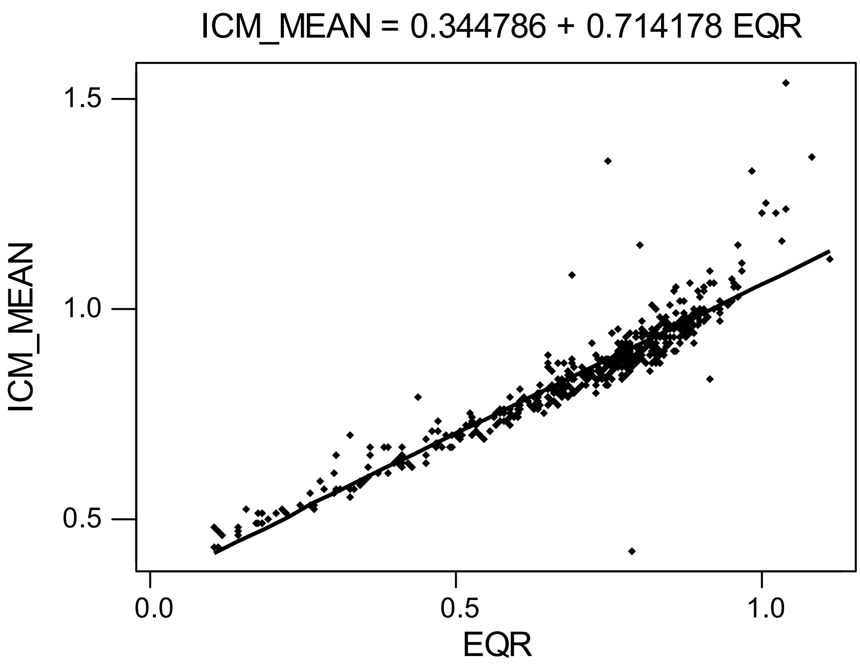


Fig. 2.2.13: Relationship between EQR of a national metric and the ICM for a hypothetical national dataset.

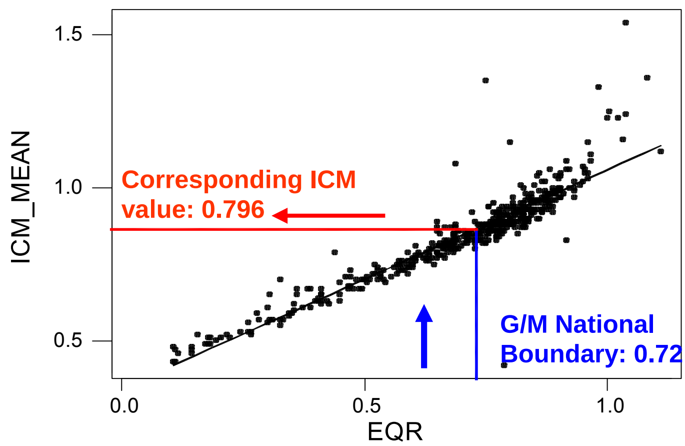


Fig. 2.2.14: Conversion of the Good/Moderate national boundary value for a hypothetical national dataset into an ICM value using the regression shown in **Fig. 2.2.13.**

A single relationship was computed for each national dataset and this relationship was used to convert boundary values for each national type to the ICM. Some MS had national types each with a different reference value. In these cases, EQR values were calculated for each type separately and then all data were pooled before the regression was calculated.

The EE test dataset had a curvilinear response to the ICM. A second order polynomial equation was fitted to this dataset (Fig. 2.2.15):

ICM = a + b_1(national metric as EQR) + b_2(national metric as EQR)^2

ICM values for the H/G and G/M boundaries are presented as the predicted value of the MS boundary ± the confidence limits of the regression line.

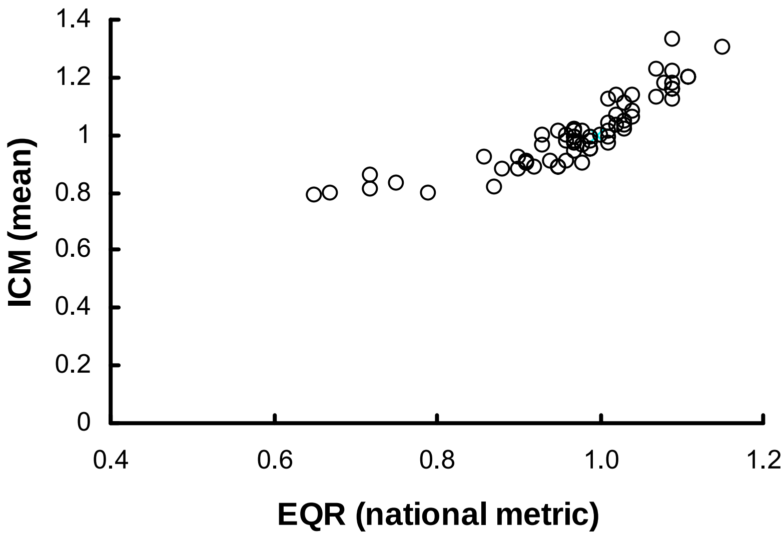


Fig. 2.2.15: Relationship between national metrics of Estonia and the intercalibration metric (ICM mean) (adjusted R^2 and RMSE for a second-order polynomial equation are 0.061 and 0.828).

2.2.5 Results of the harmonisation – Boundary EQR values

Overview of results

Status class boundaries for each MS, expressed as the ICM ± 95 % confidence limits of the prediction are presented in Figs. 2.2.16 (High/Good) and 2.2.17 (Good/Moderate). Several MS had a range of boundary values, depending on the national type; in these cases, the plotted value is the median of all the boundaries, along with the highest of the upper 95% confidence intervals of the predictions and the lowest of the lower 95% confidence intervals.

The acceptable range of boundary values was calculated as the median boundary value ± 0.05 EQR units for all MS that fulfilled an agreed list of criteria - the same approach being used for the invertebrate IC exercise. These criteria were as follows:

- Nationally agreed assessment system and boundary values;
- At least six reference samples (representing at least four sites) screened according to ECOSTAT and CB GIG guidelines;
- A statistically-significant linear relationship with the ICM. More particularly:

- Root mean square error (RMSE) ≤ 0.15
- Coefficient of determination (r^2) ≥ 0.5; and,
- Slope ≥ 0.5 and ≤ 1.5.

These criteria excluded PL (whose national assessment is not yet nationally recognised), BE-F and NL (who did not have any reference sites), BE-W who use a predicted reference value for their national EQR values, EE and UK (whose national metric had a curvilinear relationship with the ICM), and IE (whose data gave a low slope when the ICM was plotted against the national metric).

The band of acceptable values for the high/good boundary has been superimposed as a blue rectangle on Figs 2.2.16 and, for the good/moderate boundary, as a green rectangle on Fig. 2.2.17. Seven MS fall within the acceptable band for H/G and six for G/M. A few other MS are marginally above or below one or both boundaries (i.e. the upper or lower 95^{th} confidence limit overlaps with the acceptable band) while four MS for H/G and three for G/M fall outside the acceptable bands. .

Table 2.2.12 shows a detailed breakdown of results, taking national typologies into account.

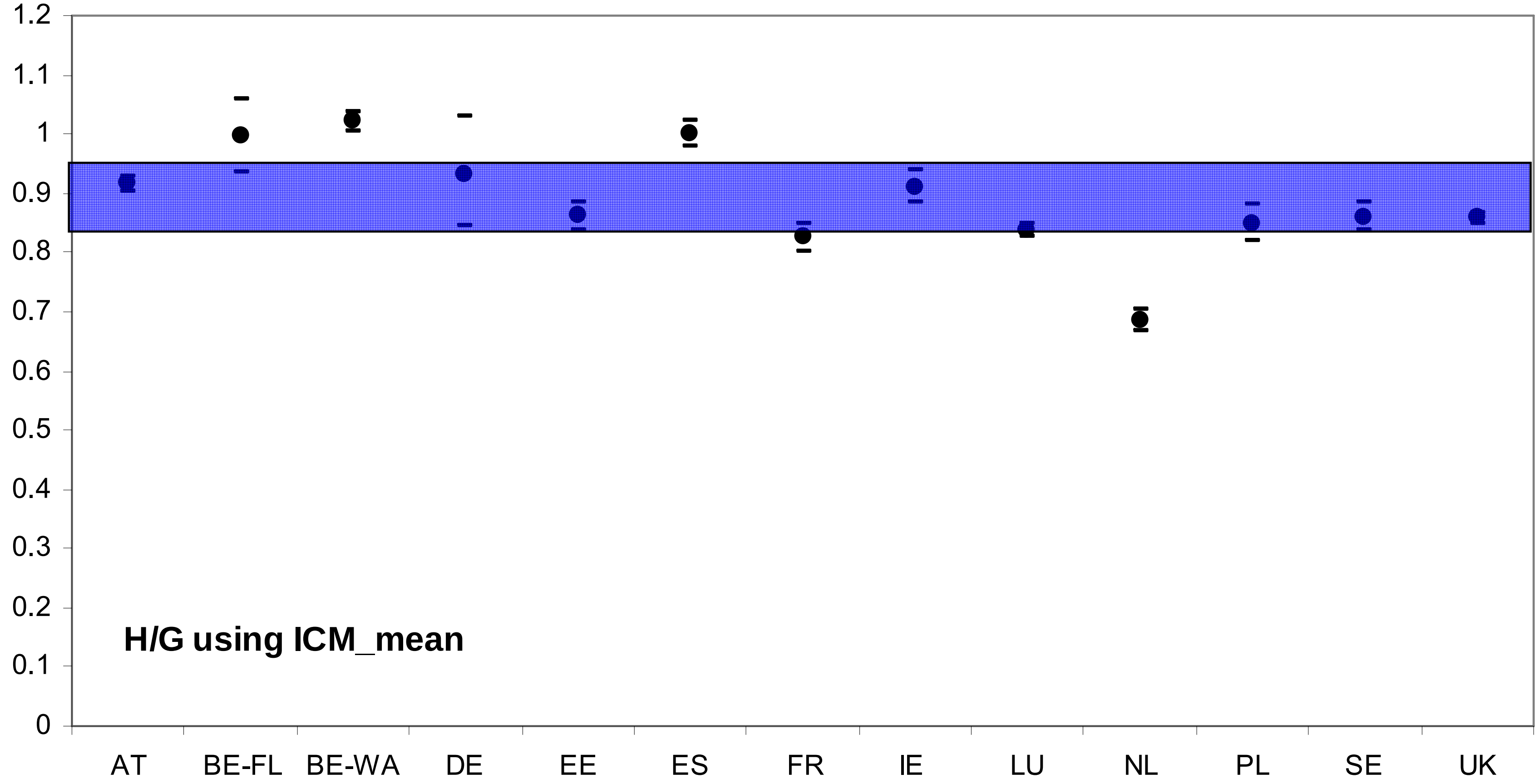


Fig. 2.2.16: High / good boundaries proposed by participants in the CB GIG phytobenthos intercalibration exercise. Data points show either the predicted boundary value ± 95% confidence limits (for those countries with a single H/G boundary value) or the median of all national boundary values, along with the highest and lowest confidence limits of the predictions (for those countries with >1 H/G boundary value. The blue rectangle shows the approximate limits of acceptable boundary values: 0.839 – 0.939.

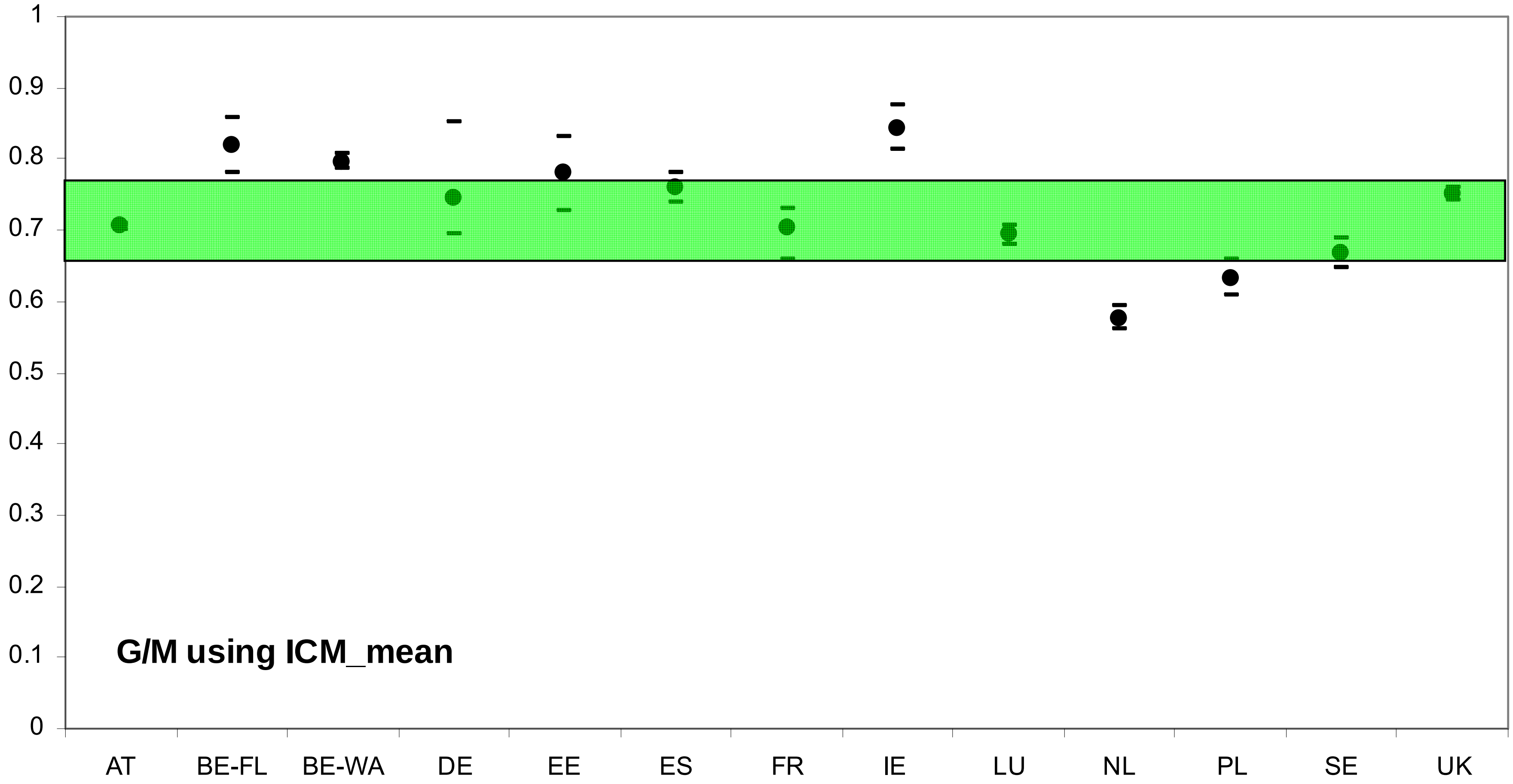


Fig. 2.2.17: Good / moderate boundaries proposed by participants in the CB GIG phytobenthos intercalibration exercise. Data points show either the predicted boundary value ± 95% confidence limits (for those countries with a single G/M boundary value) or the median of all national boundary values, along with the highest and lowest confidence limits of the predictions (for those countries with >1 G/M boundary value. The green rectangle shows the approximate limits of acceptable boundary values: 0.654 – 0.754.

Table 2.2.12: High/good and good/moderate boundaries expressed as ICM for all national datasets, sub-divided by national type, where appropriate. 'Lower' and 'upper' refer to the lower and upper 95% confidence limits of the predicted MS boundary values respectively, or the lower and upper limits of the acceptable band (median boundary value ± 0.05 EQR units) respectively in the case of the median boundary value. Codes: ✓ = predicted value for boundary falls within acceptable band; ↑ = marginally high (predicted value falls outside acceptable band but lower 95% confidence limit falls inside); ↑↑ = high (predicted value and lower 95% confidence interval both above acceptable band); ↓ = marginally low (predicted value falls outside acceptable band but upper 95% confidence limit falls inside); ↓↓ = low (predicted value and upper 95% confidence interval both fall outside acceptable band). The acceptable band for H/G is 0.839 – 0.939 and for G/M it is 0.654 – 0.754.

		Properties of regression			**High/good boundary**				**Good/moderate boundary**			
	Type	**R-squared**	**RMSE**	**slope**	**lower -**	**H/G**	**- upper**	**Code**	**lower -**	**G/M**	**- upper**	**Code**
AT (Austria)		0.683	0.068	0.758								
	< 500 m				0.904	0.917	0.929	✓	0.699	0.705	0.71	✓
	> 500 m				0.904	0.917	0.929	✓	0.699	0.705	0.71	✓
	mean					**0.917**				**0.705**		
BE-F (Belgium - Flanders)		0.686	0.112	0.886	0.936	0.997	1.058	↑	0.781	0.82	0.858	↑↑
BE-W (Belgium – Wallonia)		0.755	0.116	1.023	1.005	1.021	1.037	↑↑	0.785	0.796	0.806	↑↑
DE (Germany)		0.803	0.086	0.885								
	R-C1, R-C3				0.890	0.930	0.969	✓	0.694	0.707	0.741	✓
	R-C4				0.843	0.877	0.910	✓	0.694	0.707	0.741	✓
	R-C5				0.938	0.983	1.028	↑	0.795	0.824	0.852	↑↑
	mean					**0.930**				**0.746**		
EE (Estonia)		0.828	0.062	*	0.839	0.862	0.885	✓	0.727	0.779	0.83	↑

	Properties of regression			High/good boundary				Good/moderate boundary			
Type	**R-squared**	**RMSE**	**slope**	**lower -**	**H/G**	**- upper**	**Code**	**lower -**	**G/M**	**- upper**	**Code**
ES (Spain)	0.718	0.109	1.054	0.978	1.001	1.024	↑↑	0.738	0.759	0.779	↑
FR (France)	0.653	0.128	0.826								
Type 1				0.810	0.830	0.849	↓	0.685	0.699	0.714	✓
Type 2				0.810	0.830	0.849	↓	0.685	0.699	0.714	✓
Type 3				0.800	0.820	0.839	↓	0.655	0.669	0.684	✓
Type 4				0.810	0.830	0.849	↓	0.685	0.699	0.714	✓
mean					**0.828**				**0.703**		
IE (Ireland)	0.566	0.092	0.446	0.884	0.911	0.938	✓	0.813	0.844	0.844	↑↑
LU (Luxembourg)	0.869	0.820	0.961	0.826	0.838	0.849	↓	0.680	0.694	0.707	✓
NL (Netherlands)	0.696	0.096	0.541	0.666	0.685	0.705	↓↓	0.562	0.577	0.592	↓↓
PL (Poland)	0.951	0.060	1.030	0.819	0.849	0.880	✓	0.608	0.633	0.658	↓
SE (Sweden)	0.910	0.066	1.206	0.838	0.861	0.884	✓	0.647	0.668	0.689	✓
UK	0.563	0.121	*	0.849	0.858	0.867	✓	0.743	0.752	0.761	✓
Median Boundary				0.839	0.889	0.939		0.654	0.704	0.754	

*Curvilinear relationship between NM and ICM

The order of national boundaries (using the mean value for MS with >1 national type) for the high / good boundary was as follows:

BE-W > ES > BE-F > **DE > AT > IE > EE > SE > UK > PL** > LU > FR > NL

(MS in bold fall within the acceptable band).

For the good / moderate boundary, the order was:

IE > BE-F > BE-W > EE > ES > **UK > DE > AT > FR > LU > SE** > PL > NL

Detailed comments

AT: Austria

- Separate regressions computed for each national type.
- Strong linear response in each case.
- Both H/G and G/M boundaries fall within acceptable band.

BE-F: Belgium – Flanders

- Not included in calculation of acceptable band (no reference sites).
- Single regression computed for all national types.
- Strong linear response.
- H/G boundary occurs above the acceptable band (lower 95% confidence limits fall within acceptable bands) and G/M boundary is significantly above acceptable band.

Comment from Flanders:

The analysis shows that at least the same level of discrimination in ecological quality can be reached by a method based on the unweighted representation of indicator taxa, than by the more conventional methods used in biological water quality assessment applying weighted averaging, where water-type specificity is less transparent and the effects of multiple stressors may not always add up in the EQR. An analysis of reference samples from other MSs in the GIG was carried out to validate the reference concept, leading to a proposal of more realistic H/G boundary values for national types where data availability appears adequate (Denys, 2006). Application of these new boundary values will slightly lower the H/G boundary, bringing it more in line with the acceptable band. Accurate definition of the good/moderate boundary in relation to national types is hampered by limited data availability. At present most data are from more upstream sites and this boundary may be set somewhat too high for types representing more downstream reaches.

BE-W: Belgium – Wallonia

- Strong linear response.
- Both the H/G and GM boundaries occur above the acceptable bands.
- Dataset not included in calculation of the acceptable bands due to the use of a predicted reference value for national EQR values.

Comment from Wallonia:

Differences have been observed between the national reference value and the reference value used for the comparison exercise. The first reference value (IPS 18 used for national EQR values) came from a preliminary study. The reference sites were selected from river type RC3; the number of sites was lower and the criteria for screening reference conditions differed from the CB GIG criteria. The second value (IPS 16,1) that is used for the ICM calculation is the median value of the reference sites presented in the exercise and selected according to the common criteria. These sites concern different river types.

In order to place the boundaries H/G and G/M inside the acceptable band, new national EQR are defined for BE_W:

Former HG National EQR 0.944 (ICM_mean 1.025; outside acceptable band)

Former GM National EQR 0.722 (ICM_mean 0.798; outside acceptable band)

New HG National EQR 0,833 (ICM_mean 0.912; inside acceptable band)

New GM National EQR 0.611 (ICM_mean 0.684; inside acceptable band)

DE: Germany

- Single regression computed for all national types.
- Strong linear response.
- National boundaries for H/G fall within acceptable band for all types, with the exception of the RC-5 boundary that occurs marginally above the acceptable band.
- National boundaries for G/M fall within acceptable band for types R-C1 – R-C4; the G/M boundary for R-C5 occurs above the acceptable band.

German comment:

All DE boundaries fell inside the acceptable bands, with the exception of the H/G and G/M for R-C5. However, in light of the relatively small dataset used for these calculations, the location of this boundary should be considered to be tentative until it can be checked using a larger dataset.

EE: Estonia

- Distinctly curvilinear response (fits a second order polynomial).

- Not included in calculation of acceptable band due to the non linear response.
- H/G boundary falls within the acceptable band.
- G/M boundary occurs marginally above the acceptable band (upper 95% confidence limit falls within acceptable band).

Comment from Estonia:

The non linear response of the Estonian EQR to the ICM (**Fig. 5.8**) can be explained by examining the characteristics of the Estonian dataset (**Table 3.1**): H-55; G-4; M-2; P-1; B-0. The dataset is biased towards high quality sites and therefore not normally distributed.

ES: Spain

- Strong linear response.
- Both H/G and G/M boundaries are above the acceptable band. The G/M boundary is only marginally above the acceptable band (lower 95% confidence limit falls within the acceptable band) while the H/G boundary is significantly above the acceptable band.

Comment from Spain:

We have assumed that the performance of the indices is the same in all countries within the GIG and for all types. This assumption may not be true, as ecological and environmental data affecting the species tolerances used in the indices construction many not work locally. The process still lacks checking with pressure gradients and better data. The datasets will probably improve substantially with new data collections due to WFD requirements across Europe.

FR: France

- Single regression computed for all national types.
- The H/G boundaries for all national types included in the exercise fall marginally below the acceptable band.
- The G/M boundaries are within the acceptable band for all national types.

Comment from France:

H/G boundary: the French boundary values are only very slightly below the acceptable band and, in fact, the upper 95% confidence limits lie within the acceptable band. Our analyses suggest that the difference between the lower limit of the acceptable band and the FR boundaries depends on the statistical software used. We conclude that the present FR boundaries lie within the statistical error of the exercise and that there is not a problem with either the FR reference concept or ecological status definition.

IE: Ireland

- Not included in calculation of acceptable band (low slope).

- The H/G boundary falls within the acceptable band.
- The G/M boundary occurs significantly above the acceptable band.

Comment from Ireland:

The national metric for IE is identical to that for the UK. The difference in boundaries may reflect the smaller dataset used for IE and, in particular, the low number of poor and bad status sites, which may have affected the properties of the regression. See also comments for UK.

LU: Luxembourg

- Strong linear response where EQR < 1.
- The H/G boundary falls below the acceptable band.
- The G/M boundary falls within the acceptable band.

Comment from Luxembourg:

The most important boundary, the G/M boundary, falls inside the acceptable band. As stated by the WFD, this boundary is used to set the remediation limit. The H/G boundary overlaps the acceptable band. The accuracy of this value is therefore less important than the G/M boundary.

National EQR values (IPS_EQR) are derived from one of the two component metrics of the ICM. The national metric is better at distinguishing sites of Bad to Good quality than those ranging from Good to High, thus reflecting the comments made in **Section 5.2** of the report and represented in **Fig. 5.5**. The second component metric of the ICM, the TI_EQR, is better at distinguishing between sites of Good and High quality classes. The ICM therefore aligns very closely with the national metric for poor quality sites (up to the Good quality class), but the relationship is not as strong for High quality sites where the regression is less linear.

NL: Netherlands

- Strong linear response.
- Not included in calculation of acceptable band (no reference sites).
- Both H/G and G/M boundaries fall below the acceptable bands.

Comment from Netherlands:

The Dutch EKR is based on the presence of negative indicator species, which are promoted by increasing levels of nutrients and other stressors (e.g. acidification). Such negative indicators may also be present in reference conditions, albeit in small quantities. Moreover the relative abundance of indicator species in many samples is low. Work is underway to adapt the Dutch system in a way that meets WFD objectives.

PL: Poland

- Strong linear response.
- The H/G boundary falls within the acceptable band.
- The G/M boundary falls marginally below the acceptable band.
- This dataset was not used in the calculation of the acceptable bands as the national assessment is not yet nationally recognised.

Comment from Poland:

The PL dataset is small and the positions of these boundaries need to be validated on a more comprehensive dataset spanning all status classes. As the PL boundaries are not yet nationally agreed, we will use the outcome of this exercise to adjust the G/M boundary so that it falls within the 'acceptable band'.

SE: Sweden

- Strong linear response.
- The H/G and G/M boundaries fall within the acceptable bands.

Comment from Sweden:

Sweden has a relatively high reference value for the national method. This high value might in turn be the outcome of Sweden using a more strict selection of reference sites, e.g. the national threshold for Tot-P for reference sites is 10 µg/l.

Sweden has changed the boundaries during the process after comparison with other countries and after analysing more national data.

Former H/G National EQR: 0.87 (ICM_Mean 0.86; inside acceptable band)

Former G/M National EQR: 0.71 (ICM_Mean 0.67; inside acceptable band)

New H/G National EQR: 0.89 (ICM_Mean 0.88; inside acceptable band)

New G/M National EQR: 0.74 (ICM_Mean 0.70; inside acceptable band)

UK: United Kingdom

- Single regression computed for all national types.
- Polynomial regression gave slightly stronger relationship with ICM than linear regression (r^2 = 0.56 and 0.62 respectively).
- Not included in calculation of acceptable band (non linear regression).
- Both H/G and G/M boundaries are within the acceptable band.

Comment from UK:

There is a clear link between the location of the UK's status class boundaries and the NDs (Table 2.2.4) and both boundaries fall within the acceptable band. The slightly curvilinear response may be a statistical artefact caused by different sensitivities assigned to some common taxa in the UK system compared with the metrics included in the ICM. Paradoxically, these differences may mean that the UK national metric is particularly sensitive across the high-good-moderate portion of the gradient. Bilateral comparisons with national metrics of neighbouring countries, and the results of the Northern GIG phytobenthos intercalibration exercise, may shed further light on the reasons for these high ICM values.

2.2.6 Open issues and need for further work

Conclusions/Recommendations

Reference conditions

Issues and Problems

- A variety of approaches were adopted by MS for screening reference conditions

- The IC typology did not discriminate between reference sites (Fig. 2.2.2). In addition, there was a strong trophic gradient within the reference sites (Fig. 2.2.3). It is not clear whether differences between MS are due to screening procedures (Table 2.2.8) or to genuine ecological differences. Several samples with floras indicative of high nutrients came from those MS which had apparently adopted comprehensive screening procedures. However, the protocol for reference site selection does not ascertain that actual pressures are determined on the same basis in all MSs. Land use categories can represent a wider range of effective nutrient loading, some types of point source pollution may be neglected and some MSs included a final screening involving (different) biological criteria, whereas others did not. This may reduce the overall effectiveness of the screening procedure.

- It was not possible to derive a diatom-specific typology due to the lack of comparability of environmental data. However, we believe that the present approach, with all types pooled is 'fit for purpose'.

Recommendations

- Problems associated with reference site screening are shared by other intercalibration exercises and a means of validating and publishing criteria used for reference site selection is needed in order to ensure that the intercalibration process is open and transparent.

- Testing the validity of the IC typology should be a priority in future phytobenthos intercalibration exercises. Future work should improve the approach used for assessing the comparability of the results in order to confirm them

- Future phytobenthos intercalibration exercises should consider developing a common format for collecting key environmental data in order to facilitate development of a diatom-specific typology.

National approaches

Issues and problems

- The normative definition has been interpreted in a number of different ways. Parts of the normative definitions are vague or ambiguous.

- Most national metrics are based on existing metrics, calibrated against national reference values. All assume a nutrient / organic gradient but there are different views on whether this should be treated as a single stressor or multiple stressors.

- All national metrics consider taxonomic composition and do not address absolute abundance, undesirable disturbances or bacterial tufts (although some of these are included in methods that fall outside the remit of this exercise)

- Some methods are multimetric in design although not all will be compliant with the forthcoming CEN Guidance Standard on multimetric indices.

- Because only the effects of eutrophication and organic pollution were considered as relevant pressures in this exercise, the possible effects of interacting stressors (e.g., acidification, hydromorphology, toxic substances) on the ICM were not considered. This may affect the comparison of methods.

- Not all participants provided rationales for status class boundaries that were linked unambiguously to the normative definitions. This problem is particularly acute for 'good status' (see Table 2.2.4).

- There was a difference of opinion amongst the phytobenthos group about whether or not a status class boundary can be based solely on the value of a metric, without supporting ecological criteria.

Recommendations

- The relationship between the four components of the normative definition (taxonomic composition, abundance, undesirable disturbances, bacterial tufts) needs further examination.

- The phytobenthos IC report has not resolved the issue of status class definition. It has described national approaches and seeks further guidance from ECOSTAT.

Intercalibration metric

Issues and problems

- The ICM was derived from two widely-used metrics – the IPS and TI. National methods which incorporate one or both of these metrics tend to have stronger relationships with the ICM than those that do not. EE was an exception: although it used the IPS as its national metric, it had a curvilinear response to ICM.

- No attempt was made to harmonise taxonomy, so some 'noise' in the relationship between national metric and ICM may reflect different taxonomic treatments between countries.

- The relationship between national metrics and ICM was evaluated using stricter criteria than were used for the CB GIG invertebrate intercalibration: these criteria were: r^2, RMSE, slope and a visual examination of scatterplot to ensure linearity.

- A Model I regression was used to convert national metrics to the ICM but there are also valid arguments in favour of using a Model II regression. However, the deviation between regression lines fitted using these models is small so long as there is a strong relationship between national metric and the ICM.

- Six out of 13 MS involved in the exercise were included in calculation of the acceptable band.

Recommendations

- All MS who fall outside the 'acceptable band' should be encouraged to perform bilateral comparisons with neighbouring MS in order to validate their own boundaries.
- The ICM should be re-evaluated and, if necessary, revised before the start of phytobenthos IC exercises in other GIGs. This is particularly important for North GIG as two countries that are in both CB GIG and North GIG had poor statistical fits to the ICM.
- Ideally, the revised ICM should be independent of all national metrics.
- There should be close co-ordination between those GIGs about to start phytobenthos IC exercises, in order to share 'best practice'.
- A taxonomic harmonisation exercise – either a desk study or workshop – should be included in future exercises.

Integration with other biological elements

Issues and problems

- This exercise considered phytobenthos as a discrete entity whereas the WFD refers to 'macrophytes and phytobenthos' and some MS assess diatoms and non-diatom phytobenthos separately. It is possible that the outcome of an assessment of 'macrophytes and phytobenthos' will differ from one of diatoms alone.

Recommendations

- CB GIG should investigate how MS are integrating 'macrophytes' and 'phytobenthos' within national assessment systems.

2.2.7 Acknowledgements

We are grateful to members of the Central-Baltic GIG steering group: Roger Owen (UK), John Murray-Bligh (UK), Jean-Gabriel Wasson (France), Gisela Ofenböck (Austria), Andrea Buffagni (Italy), Nicolas Mengin (France), Sebastian Birk (Germany), Wouter van de Bund (JRC), Isabel Pardo (Spain).

Thanks, too, to Steve Juggins (University of Newcastle, UK) for statistical guidance.

2.3 Alpine GIG

2.3.1 Intercalibration approach

The intercalibration of the boundary values for the Phytobenthos method in the Alpine River GIG was done following the procedure of the Central Baltic GIG (see Kelly et al. 2006) and the results show that a further modification is not necessary. Thus, maximum consistency with the approach in the other Geographical Intercalibration Groups is guaranteed.

The **technical steps** of the calculation are:

a) Calculation of the ICM–metrics IPS and TI (done by Michel Coste and Juliette Tison)
b) Calculation of EQR values for the ICM–metrics (reference values is the median of the reference sites)
c) Regression between national method and ICM
d) Converting national boundary values into ICM boundary values (including 95% CL) using the regression formula

Similar to the Central Baltic GIG the **IPS** (indice de polluosensibilité (Cemagref 1982)) and the **TI** (Trophic diatom index Rott et al. (1999)) were used.

The calculation was done in the same way as in the Central Baltic GIG:

IPS:

EQR_IPS = Observed value / reference value

TI:

as this is a trophic index it needs to be adjusted so that high values represent high EQR values (4 is the maximum possible value of the TI):

EQR_TI = (4-observed value) / (4-reference value)

Two options (similar to the approach in the Central Baltic GIG) we tested:

A) Arithmetic mean of IPS and TI
B) Minimum of IPS and TI (worst case)

Option A resulted in better correlations between the national methods and the ICMi and was therefore selected for the use in the intercalibration procedure. This option was also used by the Central Baltic GIG. Results for option B are not included in this report.

2.3.2 National methods that were intercalibrated

The national methods that were intercalibrated are described in Table 2.3.1 below.

Table 2.3.1 Overview of nationall diatom methods that were intercalibrated in the Alpine river GIG

MS	*National metric*
Austria	Multimetric method consisting of 3 modules/metrics: • trophic status module (based on TI: Rott et al. 1999) • saprobic status module (based on SI: Rott et al. 1997) • reference species module (portion of defined reference and bioregion-specific species in total abundance and species number) worst-case-approach
France	French WFD classification Indice Biologique Diatomées (IBD) - norm AFNOR NF T 90 354 (2000) and circular MEDD/DE 05 n°14 (July 05)
Germany	Diatom Module: WFD Diatom Index = Average of the sum of abundances of type specific reference species (following Schaumburg et al. 2005) and Trophic Index (Rott et al., 1999) or (in one special case) Saprobic Index (Rott et al., 1997). Additional metrics are available for cases of acidification or salinisation. Non Diatom Module: WFD Reference species Index depends on type specific taxa and abundances (following Schaumburg et al. 2005) Macrophyte Module: WFD Reference species Index depends on type specific taxa and abundances (following Schaumburg et al. 2005). Additional metrics are available for cases of mass growth stands of special taxa. Ecological status is calculated and classified from the average of the three module scores. If a module is absent, status class can be calculated with two modules or, exceptionally, with a single module. For this reason every module is classified separately and can be considered separately for intercalibration purposes. The national classification system needs all modules of the benthic flora occurring in a monitoring section of a water body.
Italy	*work in progress (see the Annex 1)*
Slovenia	Multimetric method consisting of 2 modules/metrics: • Saprobic index (Zelinka & Marvan 1961) • Trophic index (Rott et al. 1997) Setting of boundary value: Median of ref. samples worst case approach
Spain	three indexes: IPS (Coste in Cemagref 1982) IBD (Prygiel & Coste 2000) CEE (Descy & Coste 1990, 1991) (IPS seems to be the most adequate index for the Alpine GIG)

2.3.3 Results of the comparison

Figure 2.3.1 and Table 2.3.2 show the results of the regression:

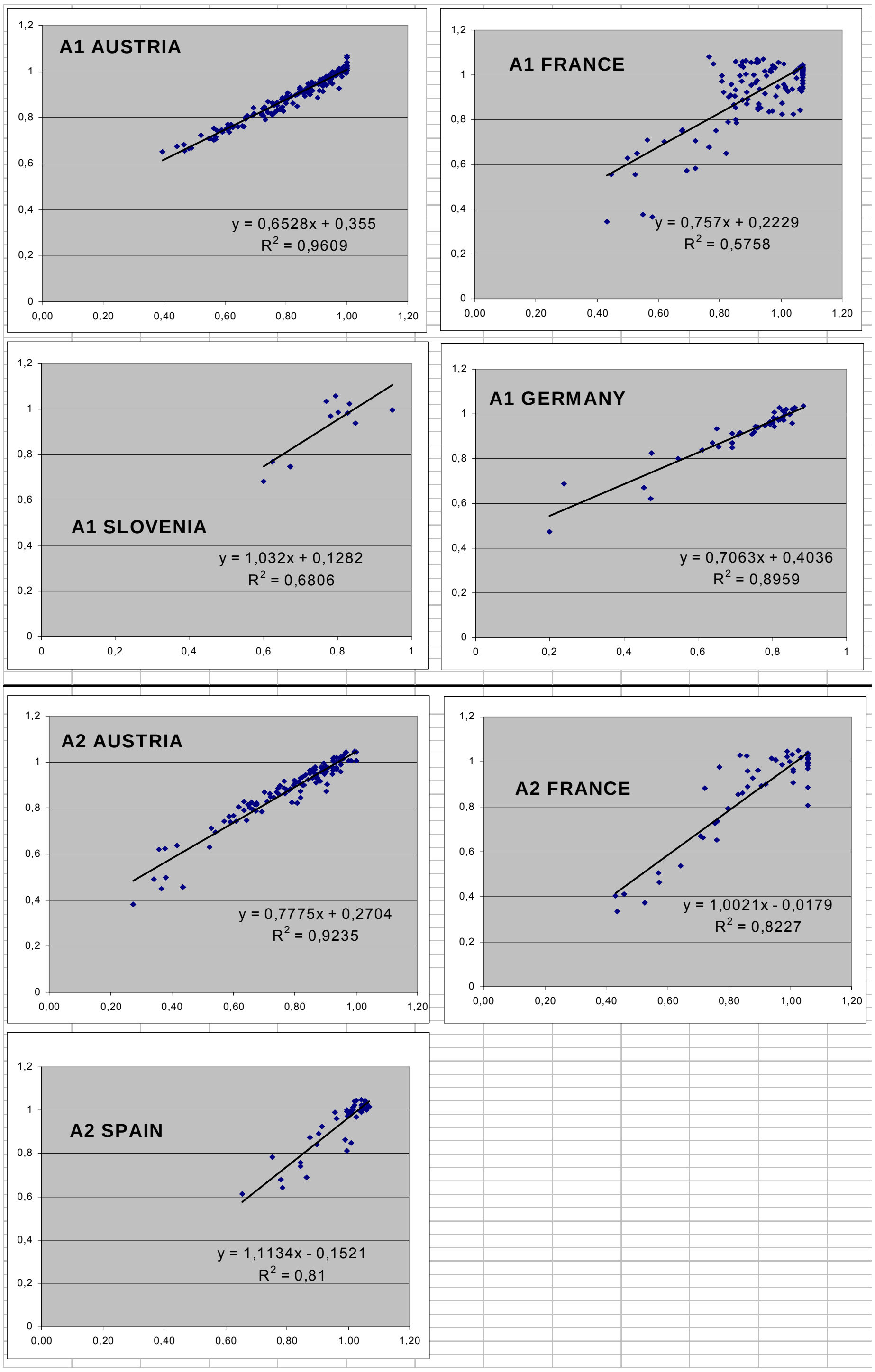


Figure 2.3.1: Regression: x-axis = national method and y-axis = ICM

Table 2.3.2: Data from Regression in Figure 1

	number of sites	number of reference sites	R^2 of national method vs ICM	lowest EQR value in data set
Type R-A1				
Austria	157	18	0.96	0.39
France	117	66	0.56	0.43
Germany	46	9	0.90	0.20
Italy	-	-	-	-
Slovenia	11	4	0.68	0.60
Type R-A2				
Austria	111	17	0.92	0.27
France	52	26	0.82	0.43
Spain	40	6	0.81	0.66
Italy	-	-	-	-

2.3.4 Results of the harmonisation – Boundary EQR values

Table 2.3.3 and Figure 2.3.2 show the results from the conversion of the national boundary values into ICM values. Table 2.3.4 shows the average ICMi boundary values +/- 5 percent confidence interval (upper and lower limit of the band.

Table 2.3.3: National boundary values converted into ICM – values; the 95% CL is taken from the regression

	National boundary		**National boundary ICMi**			
MS	**H/G**	**G/M**	**H/G value**	**+/-95% CL**	**G/M value**	**+/-95% CL**
Type R-A1						
Austria	0.87	0.56	**0.92**	*0.003*	**0.72**	*0.006*
France	0.86	0.71	**0.87**	*0.020*	**0.76**	*0.031*
Germany	0.73	0.54	**0.92**	*0.011*	**0.78**	*0.018*
Italy	-	-	**'-**	*'-*	**'-**	*'-*
Slovenia	0.80	0.60	**0.95**	*0.054*	**0.75**	*0.106*
Type R-A2						
Austria	0.87	0.56	**0.95**	*0.008*	**0.71**	*0.012*
France	0.86	0.71	**0.84***	*0.024*	**0.69**	*0.034*
Spain	0.94	0.74	**0.89**	*0.023*	**0.67**	*0.051*
Italy	'-	'-	**'-**	*'-*	**'-**	*'-*

* The French boundary value is only very slightly below the acceptable band (0.00124 point below) and the upper 95% confidence limits lie within the acceptable band. The analyses suggest that the present FR boundary lies within the statistical error of the exercise and that there is not a problem with either the FR reference concept or ecological status definition. The exact values are:
France High-Good Boundary for RA2: 0.8439
Lower limit of the band (= average - 0.05): 0.84514

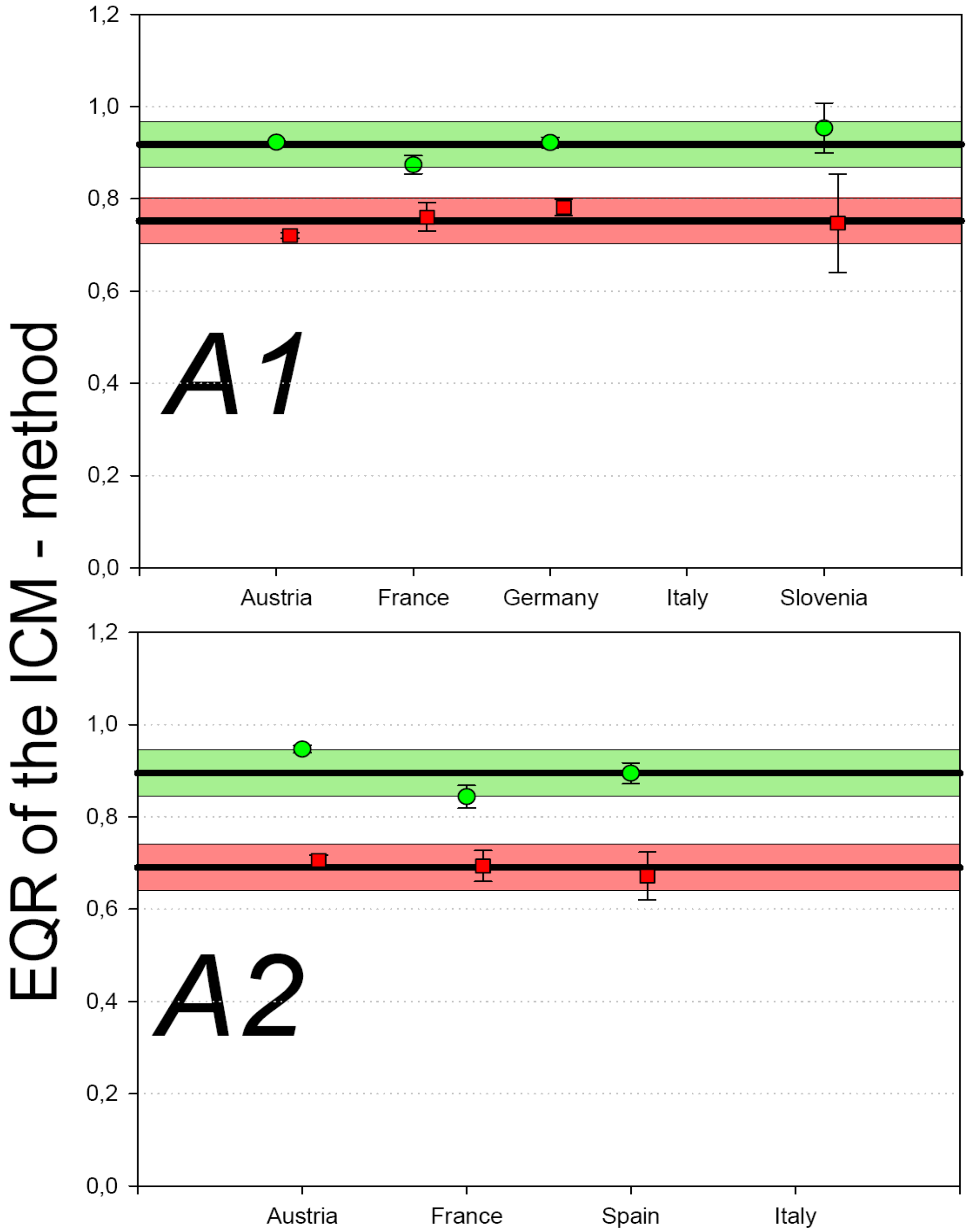


Figure 2.3.2: Comparison of the national boundary values when converted into ICM – values (+/- 95% CL). Green: High/Good boundary, Red: Good/Moderate boundary. The bands show the average boundary values (bold line) +/- 0.05.

Table 2.3.4: Harmonised boundary values: average ICMi boundary values +/- 5 percent "confidence interval "(upper and lower limit of the band) expressed as EQR-value of the ICMi

River Type and boundary	***average boundary*** (ICMi)	***lower limit*** (ICMi)	***upper limit*** (ICMi)
R-A1 *HIGH-GOOD*	0.92	0.87	0.97
R-A1 *GOOD-MODERATE*	0.75	0.70	0.80
R-A2 *HIGH-GOOD*	0.90	0.85	0.95
R-A2 *GOOD-MODERATE*	0.69	0.64	0.74

2.3.5 Open issues and need for further work

No open issues have been identified.

2.4 Mediterranean GIG

2.4.1 Intercalibration approach

Spain, Portugal and France were actively involved in the Intercalibration (IC) process. Cyprus is currently (2007-2008) developing a national assessment method and will join the group for phase 2 of the intercalibration exercise..

The IC types of the Med GIG were initially considered. MS contributed with data sets from the different types and the process was first intended to be done at type level. R-M3 was rejected because only one MS had data and the selection of reference sites or definition of reference conditions is a critical issue (Table 2.4.1).

Table 2.4.1: Description of the intercalibrated types.

Type	*R-M1*	*R-M2*	*R-M4*	*R-M5*
	drainage area 10-100km2 altitude 200-800m mixed geology regime highly seasonal	100-1000km2 <400m mixed geology highly seasonal	10-1000km2 400-1500m non-siliceous geology highly seasonal	10-100km2 <300m mixed geology temporary

2.4.2 National methods that were intercalibrated

Annex V of the WFD treats ‘macrophytes and phytobenthos’ as a single biological element for the purpose of ecological status assessment and identifies four characteristics of this biological element (taxonomic composition, abundance, likelihood of undesirable disturbances and presence of bacterial tufts) that need to be considered when setting status class boundaries. All member states (MS) in Med GIG have chosen to develop separate methods for macrophytes and phytobenthos and, in addition, to use diatoms as proxies for phytobenthos (Table 2.4.2).

Table 2.4.2: National metric for phytobenthos intercalibration.

MS	National metric
Spain	IPS (Lenoir et Coste, 1996)
Portugal	IPS (Lenoir et Coste, 1996) for R-M1 and R-M2 CEE (Descy et Coste, 1990) for R-M5
France	French WFD classification Indice Biologique Diatomées (IBD) - norm AFNOR NF T 90 354 (2000) and circular MEDD/DE 05 n°14 (July 05) (Coste in Cemagref, 1982)

All MS participating in phytobenthos IC were asked to justify their methods in terms of the normative definitions (NDs) and their responses will be considered below. It should be borne in mind that a phytobenthos assessment method does not necessarily need to consider all properties defined in the NDs either because these are considered in a macrophyte method that will be used in parallel with the phytobenthos method or because the MS can demonstrate a relationship between properties defined in the NDs which means that measurement of one property provides an indication of the state of another. In such cases, MS can use a cost-effective method for routine estimation of ecological status whilst, at the same time, demonstrating *de facto* compliance with the NDs.

Table 2.4.3 shows the extent to which the four properties listed in the NDs are incorporated into the national assessment methods. All methods assess taxonomic composition of diatoms alone.

Abundance is problematic. All MS report that abundance is assessed, but the measurement is of relative, rather than absolute, abundance of diatom taxa alone. But several studies (Bernhardt and Likens 2004, Pan *et al.,* 1999, Biggs & Close, 1989; Biggs, 1996) suggest that routine evaluation of absolute abundance may not yield significant extra information on ecological status.

This suggests that the requirement for assessment of abundance as outlined in the NDs might be better served by macrophyte survey methods, particularly where these include macroalgae. Phytobenthos biomass is very spatially and temporally heterogeneous and therefore quantitative assessment is unlikely to yield detailed insights about ecological status at low or moderate pressure levels. However, at higher pressure levels, visually-obvious growths of macroalgae such as *Cladophora* are likely to be conspicuous, often at the expense of macrophyte diversity more generally, and routine assessment of such growths using straightforward survey techniques may well yield more useful information than quantitative assessment of phytobenthos abundance.

Similarly, assessment of 'bacterial tufts' are not included directly in any of the assessment systems evaluated here. Again, a precautionary approach to boundary setting should ensure that the probability of such growths should be minimal when ecological status is good or better.

We agree with the view of the CB GIG phytobenthos expert group that if a precautionary approach to boundary setting is taken using other properties (e.g. taxonomic composition), then the probability of undesirable disturbances and bacterial tufts should be minimal when ecological status is good or better.

Table 2.4.3: Phytobenthos methods: compliance with WFD normative definitions. 1 - included in the national metric; 0 - not included.

MS	Taxonomic composition	Abundance	Undesirable disturbances	Bacterial tufts
FRANCE	1	1	0	0
Comment	Species level of identification (diatoms only).	Only relative abundance	Each taxa included inside the national routine index (IBD) as in IPS gets a quality profile in 7 classes based on the sensitiveness-tolerance to undesirable disturbance (mostly organic, -trophic, salinity)	- Not included inside the diatom index. - Some diatoms (ex : Nitzschia umbonata) are as informative on the worst organic / trophic pollution levels as bacterial or fungal tufts - Anyway, included inside French macrophyte assessment tool (IBMR)
SPAIN	1	1	0	0
Comment	Species level of identification			
PORTUGAL	1	1	0	0

MS	Taxonomic composition	Abundance	Undesirable disturbances	Bacterial tufts
Comment	Species/variety level of identification		But IPS Index is based on the sensitiveness of diatom taxa to pollution; taxa are divided in sensitivity groups according to their tolerance CEE follows a similar approach These two indexes were selected after testing the response of a high number of diatom metrics to general degradation for each national river type.	But IPS responds to high pollution pressure: a group of taxa relates to highly polluted environments. CEE follows a similar approach

2.4.3 Reference conditions and class boundary setting

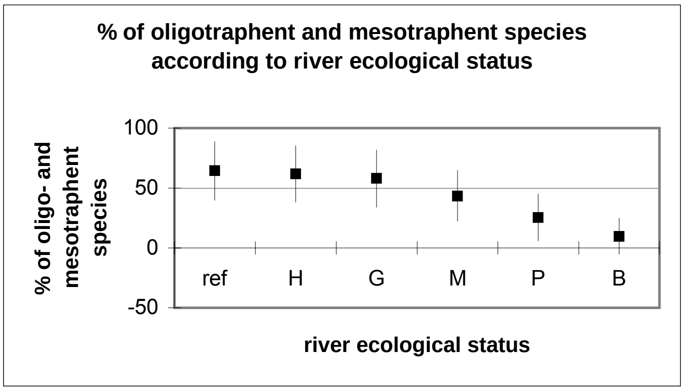
The metrics used by MS convert the response to a pressure gradient into a continuous variable which then has to be converted into an EQR, computed from Observed and Expected in reference situations values. MS adopted a variety of approaches to split this EQR scale into separate status classes. Table 2.4.4 summarises these approaches, and Table 2.4.5 presents the ecological status boundaries expressed as national metric (EQR) for each participating country.

The Normative Definitions define high, good and moderate status in terms of their deviation from the biota expected at the reference state and, therefore, a national method, if it is to be compliant with the NDs, has to be able to express each status class in terms of change from the reference state.

Table 2.4.4: Rationales for defining Member State status class boundaries

	High / Good Boundary	Good / Moderate Boundary
FRANCE	25th percentile of reference values for IBD or IPS (for every diatom natural biotypes covering all the national river types)	H/G boundary – [(H/G – minimum note)/4] + 1(for every diatom natural biotypes covering all the national river types)
Comment	The **good/moderate boundary** was calculated using a two steps procedure (this procedure based on diatom natural biotypes was used to define the provisional threshold values of the good ecological status of french river (ministerial circular DE/MAGE/BEMA 05 n°14 of the 28th July 2005)): 1: For each type, the remaining range below the H/G boundary and the IBD minimum value was split into 4 equal classes to derive a preliminary G/M boundary, following a procedure proposed in the REFCOND guidance. 2: This preliminary boundary was then increased by 1 point on the IBD scale for all national types. This procedure of boundaries calculation was chosen to be congruent with the French macroinvertebrates approach. The IBD values obtained were then checked to verify their compliance with normative definitions: the graph below shows the percentage of sensitive species (‘oligotraphent’ + ‘mesotraphent’ species: van Dam *et al*., 1994) in reference conditions and along the ecological status gradient.	

	High / Good Boundary	Good / Moderate Boundary



This graph shows:

-no significant difference in sensitive species % between reference conditions and high status;

-a very slight but significant decrease of sensitive species between high and good status;

-a drop in the percentage of sensitive species between good and moderate status.

	High / Good Boundary	Good / Moderate Boundary
SPAIN	25th percentile of reference sites	High/Good Boundary x 0.75
Comment	There is large variation between values in different seasons, related with the Mediterranean character of the systems.	
PORTUGAL	25th percentile of reference sites	High/Good Boundary x 0.75
Comment	No evident discontinuity was detected on the indexes response to the pressure gradient. The approach followed for Phytobenthos-Diatoms was the same as for the other BQEs: the so-called REFCOND method. With this method, the H/G boundary is set as the 25th percentile and then the range below is divided into 4 equal width classes. So, the G/M boundary is set as 0.75 x H/G boundary. With this approach, the values of the boundaries depend on the selection of reference sites and on the variability of the reference sites. Boundaries lower than expected are the result of variability within the reference samples pool.	

Table 2.4.5: Boundary values expressed as national metric (EQR) for a) Spain, b) Portugal and c) France.

a)

SPAIN	EQR-IPS		
type	reference	High/good	Good/ moderate
RM1	1	0.90	0.67
RM2	1	0.93	0.70
RM4	1	0.91	0.68
RM5	1	0.95	0.71

b)

PORTUGAL	EQR-IPS or EQR-CEE		
type	reference	High/good	Good/ moderate
R-M1 (IPS)	1	0.77	.0,58
R-M2 (IPS)	1	0.90	0.68
R-M5 (CEE)	1	0.85	0.64

c)

FRANCE	EQR-IBD		
type	reference	High/good	Good/ moderate
R-M1	1	0.93	0.80
R-M2	1	0.93	0.80
R-M4	1	0.86	0.71

Intercalibration common metric

The MedGIG has followed the hybrid Option 2 described in the ECOSTAT Boundary Setting Protocol (IC Process Guidance, Annex III). In this approach, the different national boundaries are compared after being translated into a common index. This common index is called Intercalibration Common Metric (ICM).

The ICM used for Phytobenthos IC process by CBGIG (cf. Report of CBGIG - Phytobenthos IC) and AlpineGIG was used as a translation index in the Med GIG, as well. This ICM was considered as an appropriate translation index (see regressions in 2.4.3 of this report) and this option makes possible the inter-GIG comparison.
The ICM is calculated as:

ICM = (EQR-IPS + EQR-TI)/2
EQR_IPS = Observed value / reference value
EQR_TI = (4-observed value) / (4-reference value)

2.4.4 Results of the comparison

Based on the regression of the National Index to ICM, the national boundaries are translated and compared. The same approach as for CB GIG Phytobenthos IC process was used to calculate boundary values and acceptable range of boundary values.
The MS had a range of boundary values, depending on the IC type; the plotted value is the mean (the median could not be calculated for 2 countries) of all the boundaries, along with the highest of the upper 95% confidence intervals of the predictions and the lowest of the lower 95% confidence intervals.

The acceptable range of boundary values was calculated as the median boundary value ± 0.05 EQR units for all MS that fulfilled the following list of criteria:
-National agreed assessment system and boundary values;
-At least six reference samples (representing at least four sites) screened according to ECOSTAT and Med GIG guidelines;
-A statistically significant linear relationships with the ICM. More particularly:
-Root mean square error (RMSE) ≤ 0.15;
Coefficient of determination (r^2) ≥ 0.5;
Slope ≥ 0.5 and ≤ 1.5

Boundaries within (or higher than) this band are accepted.

NATIONAL DATASETS

All data required for the IC exercise was stored in a central relational database, managed by the Cemagref (France). The database comprises three main components: raw diatom data, supporting chemical data and sample information. A summary of the number of sites available for each river type (Table 2.4.6), and in each quality class (including reference sites) from each MS (Table 2.4.7) are presented below.

Table 2.4.6: Summary of the number of sites available in each river type from each MS.

Member State	M1	M2	M4	M5
SPAIN	64	87	35	12
PORTUGAL	32	20		15
FRANCE	9	42	16	

Table 2.4.7: Summary of the number of sites available in each quality class (including reference sites) from each MS.

Member State	**High**	Good	**Moderate**	**Poor**	**Bad**	Total
FRANCE	41	11	13	4	0	**69**
SPAIN	51	53	39	38	17	**198**
PORTUGAL	47	7	11	2	0	**67**
Total						

STANDARDIZATION OF REFERENCE CONDITIONS

The concept of 'type-specific reference conditions' is central to the WFD as ecological status is defined in terms of deviation from the biota expected under such conditions. Different interpretations of 'reference conditions' may lead to different values being used as the denominator in EQR calculations causing different ecological status assessments.

Member States participating in the Phytobenthos IC were asked to supply the raw biological data for all reference samples in their datasets, along with information on how candidate reference sites were screened in relation to criteria established by REFCOND (Working Group 2.3 - REFCOND Guidance Document No 10.) and CB GIG (Tables 2.4.8, 2.4.9 adapted from CB GIG criteria, 2.4.10).

Table 2.4.8: **Most relevant reference screening criteria and the way each MS evaluated each parameter. Key: 1 - not used; 2 - measured; 3 - estimated; 4 - field inspection; 5 - expert judgement.**

	Landuse data (e.g. CORINE)	**BOD_5**	**O_2**	**N-NH_4**	**P**	**N-NO_3**
FRANCE	2	2	2	2	2	2
SPAIN	4	1	2	2	2	2
PORTUGAL	2	2	2	2	2	2

Table 2.4.9: Chemical reference thresholds for reference screening.

	R-M1	**R-M2**	**R-M3**	**R-M4**	**R-M5**

BOD (mg/L)					
Mean	2.4	2.4	2.4	2	2.4
90th percentile	3.6	3.6	3.6	2.75	3.6
Dissolved Oxygen (% saturation)					
Mean	95-105	95-105	95-105	95-105	95-105
10-90th percentile	85-115	90-110	90-110	90-110	85-115
N-NH4 (mg/L)					
Mean	0.1	0.1	0.1	0.05	0.1
90th-percentile	0.25	0.25	0.25	0.12	0.25
P-PO$_4$ or SRP (µg/ L)					
Mean	40	40	40	20	40
N-NO$_3$ (mg / L)					
Mean (invertebrates)	6	6	6	2	6
Mean (phytobenthos)	4	4	4	2	4

Spain indicated the chemical characteristics of the reference samples, which in a small number of cases present some differences, though in general are similar or more stringent.. (Table 2.4.10).

Table 2.4.10: Chemical characteristics of the reference samples indicated by Spain.

	R-M1	R-M2	R-M3	R-M4	R-M5
BOD (mg/L)					
Mean					
90th percentile					
Dissolved Oxygen (% saturation)					
Mean	110	103	110	100	
90th percentile	138.8	143.6	140.0	114.4	
N-NH4 (mg/L)					
Mean	0.13	0.08	0.05	0.06	0.06
90th percentile	0.25	0.13	0.07	0.12	0.06
P-PO4 (ug/L)					
Mean	16	52	29	45	47
90th percentile	43.9	101.4	34.5	127.8	72.0
N-NO$_3$ (mg/L)					
Mean	0.45	1.95	1.90	1.04	0.64
90th percentile	0.96	4.38	2.39	2.32	1.14

The number of reference samples from each MS and type is indicated in Table 2..4.11.

Table 2.4.11: **Number of reference samples by MS and river type.**

	M1	M2	M3	M4	M5
SPAIN	33	13	6	26	6
PORTUGAL	20	12			10

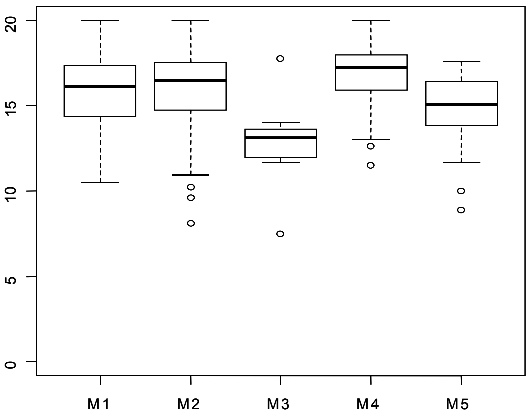
FRANCE	7	21	1	16	

INTERCALIBRATION PROCESS: ANALYSIS AND RESULTS

Evaluation of IC typology

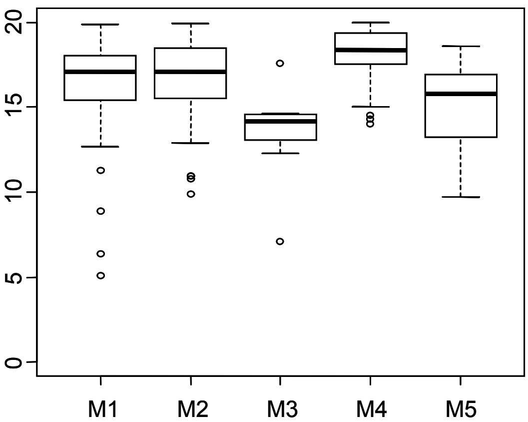
Taxonomical harmonisation was performed before data treatment.
Using only reference samples, IBD, IPS and TI values were plotted according to IC type (Figure 2.4.1).

IBD



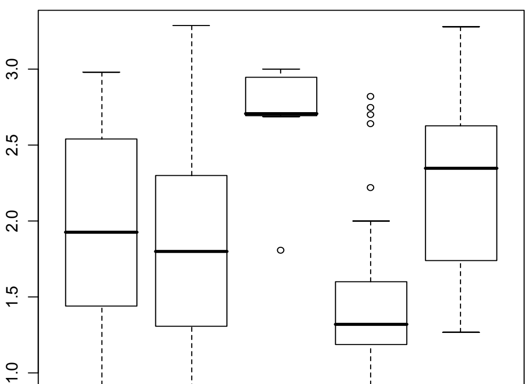
Kruskal-Wallis chi-square = 19.5808,
df = 4, p-value = 0.0006041
alternative hypothesis: two.sided

IPS



Kruskal-Wallis chi-square = 31.156,
df = 4, p-value = 2.845e-06
alternative hypothesis: two.sided

TI



Kruskal-Wallis chi-square = 35.0061,
df = 4, p-value = 4.632e-07
alternative hypothesis: two.sided

Figure 2.4.1: Evaluation of IC typology

The difference between box-plots was performed with a Kruskall-Wallis test and then the box-plots were compared 2 by 2 by a complementary non-parametric test (Conover, W.J., 1980 - Practical Nonparametric Statistics. 2nd edition. p. 231).The results ($p<0.05$) show that according to TI, IPS and IBD, M1 and M2 are identical, and M3 and M5 also. With TI and IBD, M1, M3 and M5 are all identical.

A DCA (Figure 2.4.2) was then performed with diatom data from reference stations, and confirmed that according to the flora composition the IC typology seems to be not really relevant.

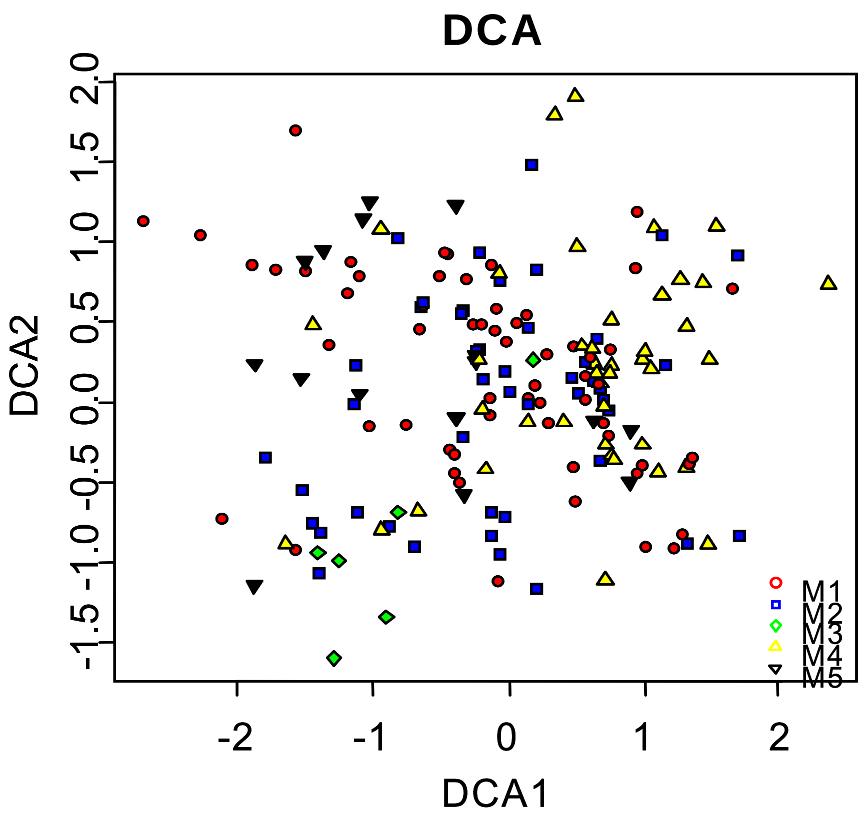


Figure 2.4.2: Detrended Correspondence Analysis (DCA), plotted by intercalibration type.

2.4.5 Results of the harmonisation – Boundary EQR values

Based on the regression of the National Index to ICM (Figure 2.4.3), the national boundaries are translated and compared. The results of the regressions show that the translation of the national indexes into ICM values is feasible.

Status class boundaries for each MS, expressed as the ICM ± 95% confidence limits of the prediction are presented in Table 2.4.12.

The acceptable range of boundary values was calculated as the median boundary value ± 0.05 (mean value for M5) EQR units for all MS that fulfilled the GIG list of criteria All MS. boundaries were i

Figure 2.4.3: Relatio

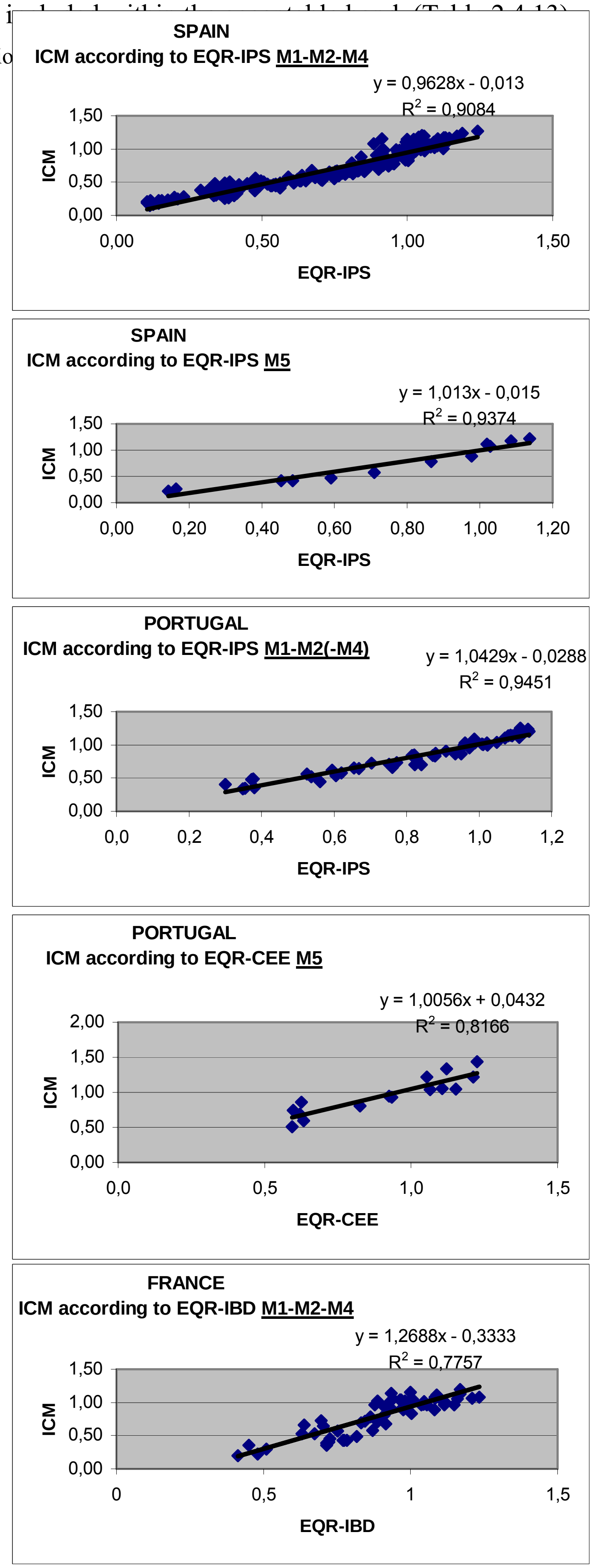


Table 2.4.12: High/Good and Good/Moderate boundaries and 95% confidence limits (lower and upper) expressed as ICM for all national datasets

	Properties of regression			H/G boundary				
	R-squared	RMSE	slope	lower	H/G	upper	Comment	lower
SPAIN								
M1	0.91	0.04	0.96	0.84	0.85	0.87		0.62
M2	0.91	0.04	0.96	0.87	0.88	0.90		0.65
M4	0.91	0.04	0.96	0.85	0.86	0.88		0.63
results M1-M2-M4				0.84	0.86	0.90	**ok**	0.62
results M5	0.94		1.01	0.88	0.95	1.01	**ok**	0.65
PORTUGAL								
M1	0.94	0.06	1.04	0.75	0.78	0.76		0.54
M2	0.94	0.06	1.04	0.89	0.92	0.93		0.65
results M1-M2				0.75	0.85	0.97	**ok**	0.54
results M5	0.82		1.00	0.85	0.90	0.97	**ok**	0.62
FRANCE								
M1	0.77	0.12	1.27	0.82	0.85	0.88		0.65
M2	0.77	0.12	1.27	0.82	0.85	0.88		0.65
results M1-M2				0.82	0.85	0.88	**ok**	0.65

Table 2.4.13: Minimum acceptable boundary values in EQR-ICM.

	M1-M2-M4	M5
H/G boundary	0.80	0.88
G/M boundary	0.61	0.65

The Med GIG decided not to intercalibrate each IC type separately, as mentioned before, considering that:

- Only 3 countries participate to the intercalibration process;
- IC types are not statistically different;
- R-M3 is not considered in this exercise, as reference sites for large river types are open to criticism and only Spain presents samples from this type;
- M5 is a very specific river type, as temporary rivers present a considerably higher natural variability

The Med GIG decided to intercalibrate M1, M2 and M4 types together. Even if R-M4 sites are mainly from mountain areas, EQR are expected to be more are less equivalent. Moreover, only Spain present enough data for M4, as France only gathers reference samples of this type. Only Spain and Portugal present data for M5 and so deriving boundaries and bands with only two countries may rise some criticism. However this report includes results for this type.

Figure 2.4.4 illustrates the results of intercalibration: the boundary values from all participating countries for M1+M2+M4 and for M5 are in agreement with the acceptability band.
Table 2.4.14 presents the ICM and National Classification values for IC types and MS satisfying the minimum acceptable values.

As may be observed, M5, in spite of its higher natural variability, present higher boundaries (Tables 2.4.12 and 2.4.14), possibly because no different hydrological years were included in each national data sets. As mentioned before, the boundaries from this type should be considered as provisional as only two MS provided data.

It must be emphasised that a possible way to deal with the natural temporal variability of M5 (and possibly other types) is to derive different reference conditions for different sets of hydrological years (e.g. dry years) and use them according to the characteristics of the data subsets. With this approach, EQRs are calculated with the appropriate reference values (e.g., *dry* reference value for *dry* year samples) and, because of the standardisation properties of EQR, the problems related to inter-annual variability are solved or at least reduced.

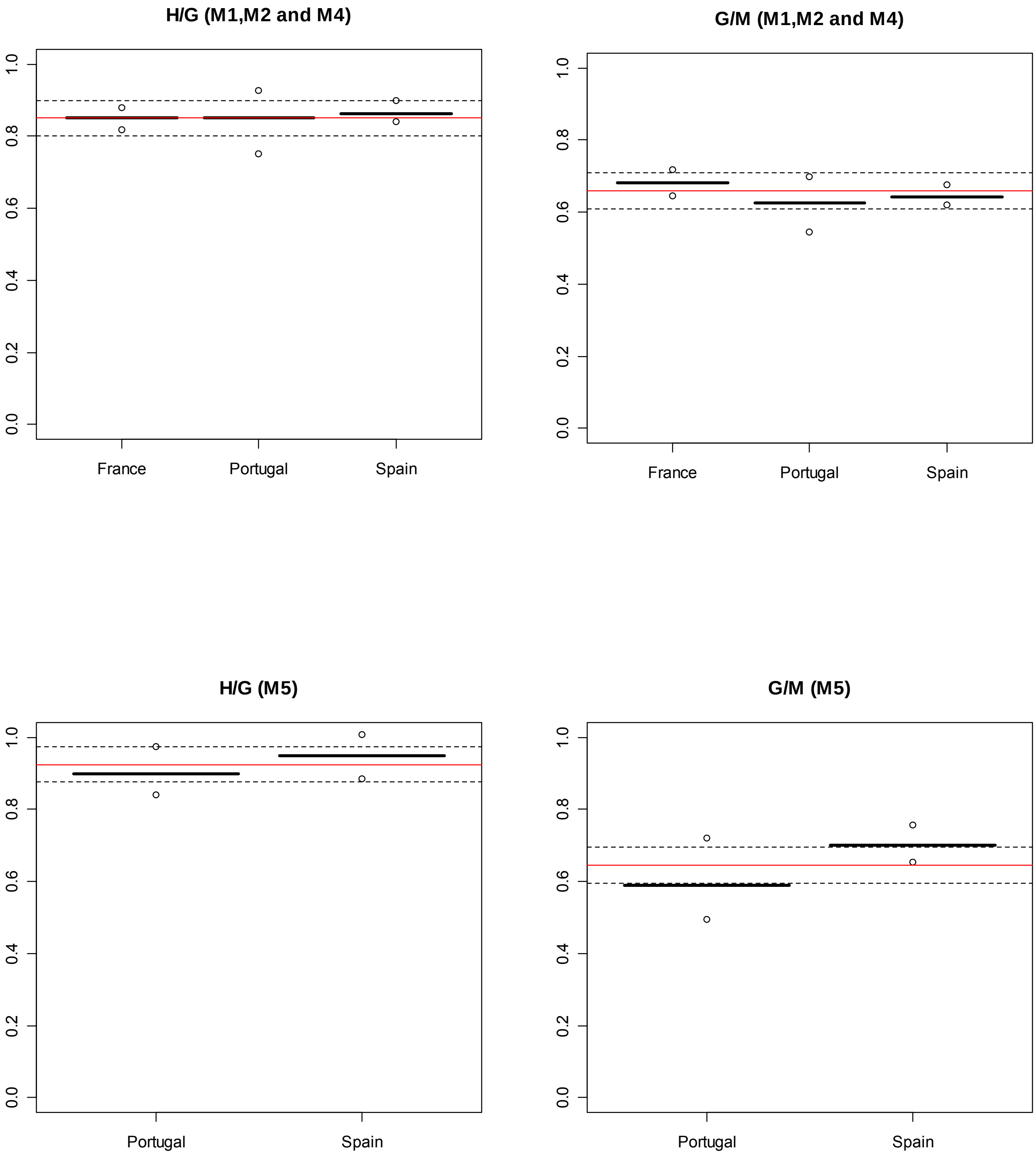


Figure 2.4.4: High/Good and Good /Moderate boundaries proposed by participants in the Med GIG phytobenthos intercalibration exercise, and acceptability bands (dashed line).

Table 2.4.14: Results of the Intercalibration - H/G and G/M Boundaries for each type and country as EQR National Classification values and EQR ICM values derived through regression.

Type and country	***Ecological Quality Ratios for the classification systems***			
	High-Good boundary		***Good-Moderate boundary***	
	National Classification System	Corresponding ICM	National Classification System	Corresponding ICM
R-M1				
France	0.93	0.85	0.80	0.68
Portugal	0.84	0.85	0.62	0.62
Spain	0.90	0.85	0.67	0.63
R-M2				
France	0.93	0.85	0.80	0.68
Portugal	0.84	0.85	0.62	0.62
Spain	0.93	0.88	0.70	0.66
R-M4				
Spain	0.91	0.86	0.68	0.64
Minimum acceptable ICM-EQRs for M1-M2-M4		**0.80**		**0.61**
R-M5				
Portugal	0.83	0.88	0.64	0.69
Spain	0.95	0.95	0.71	0.70
Minimum acceptable ICM-EQRs for M5		**0.88**		**0.65**

2.4.6 Open issues and need for further work

As a general comment, results from this first stage of Phytobenthos IC show that intercalibration is feasible, at least with the national methods from the participating MS. However, only three MS (France, Portugal, Spain) participated to the Phytobenthos intercalibration. Only data from Spain and Portugal were used for M5 and so results for this type included in this report should only be interpreted as very preliminary.

With 1 or 2 more years of work, very probably more data will be available and more MS will be involved (cf. Annex from Italy mentioning the possibility to join soon the Phytobenthos IC). Under these circumstances, a review of the intercalibration may be carried out in a near future providing more reliable results. For this reason, the MedGIG suggests the continuation of the IC process in order to work with larger data sets and more MS.

3 Discussion

3.1 Comparability between GIGs

The intercalibration approaches for the four GIGs that have completed the work for phytobenthos is generally comparable, all focusing on diatoms applying a common metric approach.

The CB GIG exercise involved 12 Member States; whilst the exercise for the other GIGs is much smaller, with just a few participants. An important implication is that the exercise has lower statistical power and it is not always clear if those MS that fall outside the 'acceptable band' do so because there are issues that those MS need to address or because the 'acceptable band' is itself based on a small (and potentially atypical sample). On the other hand, however, the 'acceptable band' should not be equated with 'best practice'. MS that comply with the minimum requirements of the exercise are included in the acceptable band and the position of this band, therefore, reflects the consensus of those.

This must affect how results from smaller intercalibration exercises are judged. In particular, a 'Type 1 error' (i.e. erroneous rejection of the [null] hypothesis that boundaries are the same) may lead to the conclusion that a MS needs to adjust boundaries when, in fact, the median value of the ICM (which anchors the acceptable band) is unlikely to be stable with such a small sample size.

The approach adopted in the Northern GIG here was, therefore, to perform a suite of tests using different permutations of the statistical criteria and to make final judgements about the need (or otherwise) to adjust boundaries based on the weight of evidence. Whilst the CB GIG exercise evaluated two versions of the ICM (one based on the mean of component metrics, the other based on the minimum), the Northern GIG exercise used both versions. TISI-min favoured Ireland and UK, both of whose national metric was the TDI, which correlates more strongly with the nutrient-sensitive TI, whilst TISI-mean favoured Finland and Sweden whose national metric was the IPS, which correlated more strongly with the SI. Whilst TISI-mean is not biased by a low value of one or other metric, TISI-min better embodies the 'one out, all out' principle used when comparing biological elements as part of status assessments.

Three of the four MS taking part in the NGIG exercise were also involved in the CB GIG exercise. Boundaries calculated in this exercise are broadly consistent between the two exercises. For H/G, Ireland, Sweden and UK were all inside the acceptable band for the CB GIG exercise whilst, for Northern GIG, UK were inside whilst Sweden was above the acceptable band for TISI-min but inside for TISI-mean and Ireland was marginally below for TISI-mean. For G/M, UK and Sweden were inside the acceptable band whilst Ireland was above. For the N GIG exercise, Ireland and UK were inside the acceptable band on all occasions whilst Sweden was again above the acceptable band when TISI-min was used. In the case of Ireland, the relatively small size of the dataset plus the low number of poor quality sites may be responsible for the differences in regression equations.

Whilst Sweden were above the acceptable band on two out of four occasions for each of H/G and G/M comparisons, it is only those MS that fall below the acceptable band that need to consider harmonisation. In this exercise, both Ireland and Finland fell below the acceptable band on one out of four occasions, both were only marginally below the acceptable band on these occasions and we believe that there is no case for either MS to adjust their boundaries.

For the Mediterranean GIG, only three MS (France, Portugal, Spain) participated to the Phytobenthos intercalibration. Only data from Spain and Portugal were used for M5 and so results for this type included in this report should only be interpreted as very preliminary. With 1 or 2 more years of work, very probably more data will be available and more MS will be involved (cf. Annex from Italy mentioning the possibility to join soon the Phytobenthos IC). Under these circumstances, a review of the intercalibration may be carried out in a near future providing more reliable results. For this reason, the Mediterranan GIG suggests the continuation of the intercalibration process in order to work with larger data sets and more countries.

3.2 Open issues and need for further work

Typology and reference conditions

It was not possible to derive a diatom-specific typology due to the lack of comparability of environmental data. However, the diatom intercalibration expert groups believe that the present approach, with all types pooled is ‘fit for purpose’A variety of approaches were adopted by MS for screening reference conditions. The intercalibration typology did not discriminate between reference sites (Fig. 2.2.2). In addition, there was a strong trophic gradient within the reference sites (Fig. 2.2.3). It is not clear whether differences between MS are due to screening procedures (Table 2.2.8) or to genuine ecological differences. Several samples with floras indicative of high nutrients came from those MS which had apparently adopted comprehensive screening procedures. However, the protocol for reference site selection does not ascertain that actual pressures are determined on the same basis in all MSs. Land use categories can represent a wider range of effective nutrient loading, some types of point source pollution may be neglected and some MSs included a final screening involving (different) biological criteria, whereas others did not. This may reduce the overall effectiveness of the screening procedure.

The following recommendations were agreed upon to deal with these issues:

- Problems associated with reference site screening are shared by other intercalibration exercises and a means of validating and publishing criteria used for reference site selection is needed in order to ensure that the intercalibration process is open and transparent.
- Testing the validity of the IC typology should be a priority in future phytobenthos intercalibration exercises. Future work should improve the approach used for assessing the comparability of the results in order to confirm them
- Future phytobenthos intercalibration exercises should consider developing a common format for collecting key environmental data in order to facilitate development of a diatom-specific typology.

Differences in national approaches to classification and boundary setting

Although the results of the intercalibration exercise ensure a satisfactory level of comparability between Member State’s class boundaries, the normative definitions have been interpreted in a number of different ways; one of the reasons for this is that parts of the normative definitions are perceived as somewhat vague or ambiguous.

Most national metrics are based on existing metrics, calibrated against national reference values. All assume a nutrient / organic gradient but there are different views on whether this should be treated as a single stressor or multiple stressors. All national metrics consider taxonomic composition and do not address absolute abundance, undesirable disturbances or bacterial tufts (although some of these are included in methods that fall outside the remit of this exercise). Some methods are multimetric in design although not all will be compliant with the forthcoming CEN Guidance Standard on multimetric indices. Because only the effects of eutrophication and organic pollution were considered as relevant pressures in this exercise, the possible effects of interacting stressors (e.g., acidification, hydromorphology, toxic substances) on the ICM were not considered. This may affect the comparison of methods.

Not all participants provided rationales for status class boundaries that were linked unambiguously to the normative definitions. This problem is particularly acute for 'good status' (see Table 2.2.4). There was a difference of opinion amongst the phytobenthos group about whether or not a status class boundary can be based solely on the value of a metric, without supporting ecological criteria.

The following recommendations were argreed upon to deal with these issues:

- The relationship between the four components of the normative definition (taxonomic composition, abundance, undesirable disturbances, bacterial tufts) needs further examination.
- The phytobenthos IC report has not resolved the issue of status class definition. It has described national approaches and seeks further guidance from ECOSTAT.

The intercalibration metric

The ICM was derived from two widely-used metrics – the IPS and TI. National methods which incorporate one or both of these metrics tend to have stronger relationships with the ICM than those that do not. Estonia was an exception: although it used the IPS as its national metric, it had a curvilinear response to ICM. No attempt was made to harmonise taxonomy, so some 'noise' in the relationship between national metric and ICM may reflect different taxonomic treatments between countries.

The relationship between national metrics and ICM was evaluated using stricter criteria than were used for the Central-Baltuic GIG invertebrate intercalibration: these criteria were: r^2, RMSE, slope and a visual examination of scatterplot to ensure linearity. A Model I regression was used to convert national metrics to the ICM but there are also valid arguments in favour of using a Model II regression. However, the deviation between regression lines fitted using these models is small so long as there is a strong relationship between national metric and the ICM.

The following recommendations were argreed upon to deal with these issues:

- All MS who fall outside the 'acceptable band' should be encouraged to perform bilateral comparisons with neighbouring MS in order to validate their own boundaries.
- The ICM should be re-evaluated and, if necessary, revised before the start of phytobenthos IC exercises in other GIGs. This is particularly important for North GIG as two countries that are in both CB GIG and North GIG had poor statistical fits to the ICM.
- Ideally, the revised ICM should be independent of all national metrics.
- There should be close co-ordination between those GIGs about to start phytobenthos IC exercises, in order to share 'best practice'.
- A taxonomic harmonisation exercise – either a desk study or workshop – should be included in future exercises.

Integration with other biological elements

This exercise considered phytobenthos as a discrete entity whereas the WFD refers to 'macrophytes and phytobenthos' and some MS assess diatoms and non-diatom phytobenthos separately. It is possible that the outcome of an assessment of 'macrophytes and phytobenthos' will differ from one of diatoms alone.It is recommended to investigate how MS are integrating 'macrophytes' and 'phytobenthos' within national assessment systems.

4 References

AFNOR (2000). Détermination de l'Indice Biologique Diatomées (IBD). Norme NF T 90-354.

Bernhardt, E.S. and Likens, G.E. (2004). Controls on periphyton biomass in heterotrophic streams. Freshwater Biology, 49, 14-27.

Biggs, B.J.F. (1996): Patterns in Benthic Algae of Streams in Stevenson, R.J., Bothwell, M.L. and Lowe, R.L. (eds). Algal Ecology: Freshwater Benthic Ecosystems, Academic Press, San Diego.

Biggs, B.J.F. & Close, M.E. (1989). Periphyton biomass dynamics in gravel bed rivers – the relative effects of flows and nutrients. Freshwater Biology, 22, 209-231.

Buffagni A., Erba S., Birk S., Cazzola M., Feld C., Ofenböck T., Murray-Bligh J., Furse M. T., Clarke R., Hering D., Soszka H. & W. van de Bund (2005). Towards European inter-calibration for the Water Framework Directive: procedures and examples for different river types from the E.C. Project STAR. 468pp. Istituto di ricerca sulle acque, Rome.

CEN (2003). Water quality - Guidance standard for the routine sampling and pretreatment of benthic diatoms from rivers. EN 13946: 2003. Comité European de Normalisation, Geneva.

CEN (2004). Water quality - Guidance standard for the identification, enumeration and interpretation of benthic diatom samples from running waters. EN 14407:2004. Comité European de Normalisation, Geneva.

Common Implementation Strategy (CIS) for the Water Framework Directive (2000/60/EC) Guidance Document No. 14, Guidance on the Intercalibration Process 2004-2006.

Coste M and Ayphassorho (1991). Étude de la qualité des eaux du Bassin Artois-Picardie à l'aide des communautés de diatomées benthiques (application des indices diatomiques). Rapport CEMAGREF. Bordeaux – Agence de l'Eau Artois-Picardie, Douai.

Coste in CEMAGREF, 1982. Etude des méthodes biologiques d'appréciation quantitative de la qualité des eaux. Rapport Q.E. Lyon A.F. Bassin Rhône-Méditérannée-Corse, 218p.

Denys L. (2006). Validation and revision of the reference concept for river phytobenthos in Belgium – Flanders proposed for the European Water Framework Directive, based on diatom assemblages from reference sites in the Central-Baltic GIG region. Advies Instituut voor Natuur- en Bosonderzoek INBO.A.2006.163, Brussel, 8p.

ECOSTAT [Working Group A] (2005). Activity on eutrophication. Eutrophication assessment in the context of European water policies. Version 10, 25 October 2005. Draft Guidance.

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

Hill, M.O. (1979). DECORANA - A FORTRAN program for Detrended Correspondence Analysis and Reciprocal Averaging. Ecology and Systematics, Cornell University, Ithaca, New York.

Hendrickx, A. and Denys, L. (2005). Toepassing van verschillende biologische beoordelingssystemen op vlaamse potentële interkalibratielocaties overeenkomstig de europese kaderrichtlijn water : partim 'fytobenthos'. Rapport van het Instituut voor Natuurbehoud, IN.R.2005.06. Instituut voor Natuurbehoud: Brussel : Belgium. 107 pp.

Kelly, M.G., Cazaubon, A., Coring, E., Dell'Uomo, A., Ector, L., Goldsmith, B., Guasch, H., Hürlimann, J., Jarlman, A., Kawecka, B., Kwandrans, J., Laugaste, R., Lindstrøm, E.-A., Leitao, M., Marvan, P., Padisák, J., Pipp, E., Prygiel, J., Rott, E., Sabater, S., van Dam, H., Vizinet, J., 1998. Recommendations for the routine sampling of diatoms for water quality assessments in Europe. *Journal of Applied Phycology,* 10, 215-224.

Kelly, M.G. (2006). A comparison of diatoms with other phytobenthos as indicators of ecological status in streams in northern England. Pp. 139-151. In: Proceedings of the 18th International Diatom Symposium (edited by A. Witkowski). Biopress, Bristol.

Kelly MG and Whitton BA (1995). The trophic diatom index: a new index for monitoring eutrophication in rivers. J. Appl. Phycol. 7 433-444.

Kelly, M.G., Rippey, B., King, L., Ní Chatháin, B, McQuillan, C. & Poole, M. (2006a). *Use of phytobenthos for evaluating ecological status in Ireland.* Report to North-South Shared Aquatic Resource (NSShARe) project.

Kelly, M.G., Juggins, S., Bennion, H., Burgess, A., Yallop, M., Hirst, H., King, L., Jamieson, J., Guthrie, R., Rippey, B. (2006b*). Use of diatoms for evaluating ecological status in UK freshwaters*. 160pp: Draft final report to Environment Agency.

Leclerq L and Maquet B (1987). Deux nouveaux indices chimique et diatomique de qualité d'eau courante. Application au Samson et à ses affluents (bassin de la Meuse belge). Comparaison avec d'autres indices chimiques, biocénotiques et diatomiques. Institut Royal des Sciences Naturelles de Belgique, document de travail 28.

Lenoir A and Coste M (1996). Development of a practical diatom index of overall water quality applicable to the French National Water Board network. In: WHITTON BA and ROTT E (eds.) Use of Algae for Monitoring Rivers II. Institut für Botanik. Universität Innsbruck. pp 29-43.

Pan, Y., Stevenson, R.J., Hill, B.H., Kaufmann, P.R. and Herlihy, A.T. (1999). Spatial patterns and ecolocical determinants of benthic algal assemblages in mid-atlantic streams, USA. Journal of Phycology, 35, 460-468.

Pardo, I. A. M. Olsen, C. Delgado, L. García, A. Nebra & M. Domínguez. (2005). Implantación da Directiva Marco da Auga 2000/60/CE no Ámbito territorial Galicia-Costa. Technical report.

R Project Core Development Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. **http://www.R-project.org**

Rott, E.; Hofmann, G.; Pall, K.; Pfister, P. & Pipp, E. (1997). Indikationslisten für Aufwuchsalgen. Teil 1: Saprobielle Indikation. Publ. Wasserwirtschaftskataster, BMfLF, 1-73.

Rott, E.; Van Dam, H.; Pfister, P.; Pipp, E.; Pall, K.; Binder, N. & Ortler, K. (1999). Indikationslisten für Aufwuchsalgen. Teil 2: Trophieindikation, geochemische Reaktion, toxikologische und taxonomische Anmerkungen. Publ. Wasserwirtschaftskataster, BMfLF, 1-248.

Schaumburg, J., U. Schmedtje, B. Köpf, C. Schranz, S. Schneider, P. Meilinger, D. Stelzer, G. Hofmann, A. Gutowski & J. Foerster (2005). Makrophyten und Phytobenthos in Flüssen und Seen. Leitbildbezogenes Bewertungsverfahren zur Umsetzung der EG-Wasserrahmenrichtlinie. Informationsbericht Heft 1/05. Bayerisches Landesamt für Wasserwirtschaft, München

Schiefele S and Schreiner C (1991). The use of diatoms for monitoring nutrient enrichment, acidification and impact of salt in rivers in Germany and Austria. In: WITTON BA, ROTT E and G. FRIEDRICH (eds.) Use of Algae for Monitoring Rivers. Institut fur Botanik, Univ. Innsbruck: 103-110.

Slàdeček V (1986). Diatoms as indicators of organic pollution. Acta Hydrochim. Hydrobiol. 14(5) 555-566.

Van der Molen, D.T. (Ed.) (2004). Referenties en concept-maatlatten voor rivieren voor de Kaderrichtlijn Water. STOWA-rapport 2004/43. STOWA, Utrecht. 365p.

Wallach, D. & Goffinet, B. (1989). Mean squared error of prediction as a criterion for evaluating and comparing system models. Ecological Modelling, 44, 299-306.

Wallin, M., Wiederholm, T. & Johnson, R. (2003). Guidance on establishing reference conditions and ecological status class boundaries for inland surface waters. Produced by CIS Working Group 2.3 – REFCOND.

Warton, D. (2005). SMATR: (Standardised) Major Axis Estimation and Testing Routines. R package version 2.0, translated to R by John Ormerod.

Wasson (2006). Proposals for Reference Thresholds of selected chemical parameters for the Central-Baltic GIG intercalibration. Working paper for the CB GIG. April 4th, 2006.

Working Group 2.3 - REFCOND Guidance Document No 10. Rivers and Lakes – Typology, Reference Conditions and Classification Systems. Common Implementation Strategy For The Water Framework Directive (2000/60/EC).

Zelinka, M. & Marvan, P (1961). Zur Präzisierung der biologischen Klassifikation des Reinheit fliessender Gewässer. Archiv für Hydrobiologie 57: 389-407.

5 Glossary

Term	**Explanation**
Biological metric	A calculated value representing some aspect of the biological population's structure, function or other measurable characteristic that changes in a predictable way with increased human influence.
BQE	Biological quality element.
CEN	Comité European de Normalisation.
CIS	Common Implementation Strategy.
Class boundary	The EQR value representing the threshold between two quality classes.

Ecological status	One of two components of surface water status, the other being chemical status. There are five classes of ecological status of surface waters (high, good, moderate, poor and bad).
ECOSTAT CIS	Common Implementation Strategy (CIS) Working Group A Ecological Status.
EQR	Ecological Quality Ratio.
GIG	Geographic Intercalibration Group i.e. a geographical area assumed to have comparable ecological boundaries conditions.
Good ecological status	Status of a body of surface water, classified in accordance with WFD standards (cf. annex V of the WFD).
Harmonisation	The process by which class boundaries should be adjusted to be consistent (with a common European defined GIG boundary). It must be performed for HG and GM boundaries.
ICM	Intercalibration Common Metric.
Intercalibration	Benchmarking exercise to ensure that good ecological status represents the same level of ecological quality everywhere in Europe.
MS	Member State (of the European Union)
Pressures	Physical expression of human activities that changes the status of the environment (discharge, abstraction, environmental changes, etc...).
REFCOND	Development of a protocol for identification of reference conditions, and boundaries between high, good and moderate status in lakes and watercourses. EU Water Framework Directive project funded by the European Commission Environment Directorate-General.
Reference conditions	The benchmark against which the effects on surface water ecosystems of human activities can be measured and reported in the relevant classification scheme.
Water body	Distinct and significant volume of water. For example, for surface water: a lake, a reservoir, a river or part of a river, a stream or part of a stream.
WFD	Water Framework Directive.

European Commission

EUR XXXXX LL – Joint Research Centre – Institute for Environment and Sustainability
Title:
Author(s):
Luxembourg: Office for Official Publications of the European Communities
20YY – nnnn pp. – x cm
EUR – Scientific and Technical Research series – ISSN 1018-5593
ISBN X-XXXX-XXXX-X
DOI XXXXX

Abstract
Text...............................

How to obtain EU publications

Our priced publications are available from EU Bookshop (http://bookshop.europa.eu), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

AA- BB- XXXXX- LL- C

EUROPEAN COMMISSION

Publications Office
Publications.eu.int

Bar code