Technical Report - 2011 - 045

COMMON IMPLEMENTATION STRATEGY FOR THE WATER FRAMEWORK DIRECTIVE (2000/60/EC)

Guidance document No. 14

GUIDANCE DOCUMENT ON THE INTERCALIBRATION PROCESS 2008-2011

EUROPEAN COMMISSION

environment

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

ISBN : 978-92-79-18997-5
DOI : 10.2779/99432

Disclaimer:
This technical document has been developed through a collaborative programme involving the European Commission, all the Member States, the Accession Countries, Norway and other stakeholders and Non-Governmental Organisations. The document should be regarded as presenting an informal consensus position on best practice agreed by all partners. However, the document does not necessarily represent the official, formal position of any of the partners. Hence, the views expressed in the document do not necessarily represent the views of the European Commission.

Foreword

The Water Directors of the European Union (EU), Acceding Countries, Candidate Countries and EFTA Countries have jointly developed a common strategy for supporting the implementation of the Directive 2000/60/EC, "establishing a framework for Community action in the field of water policy" (the Water Framework Directive). The main aim of this strategy is to allow a coherent and harmonious implementation of the Directive. Focus is on methodological questions related to a common understanding of the technical and scientific implications of the Water Framework Directive. In particular, one of the objectives of the strategy is the development of non-legally binding and practical Guidance Documents on various technical issues of the Directive. These Guidance Documents are targeted to those experts who are directly or indirectly implementing the Water Framework Directive in river basins. The structure, presentation and terminology are therefore adapted to the needs of these experts and formal, legalistic language is avoided wherever possible.

In the context of the above-mentioned strategy, several guidance documents directly relevant for intercalibration have been developed and endorsed by the Water Directors. They provide Member States with guidance e.g. on typology and setting reference conditions for inland waters (CIS Guidance No. 10) as well as for coastal and transitional waters (CIS Guidance No. 5), on the classification of water bodies (CIS Guidance No. 13) and on monitoring (CIS Guidance No. 7). In 2002 the Water Directors endorsed the document 'Towards a guidance on establishment of the intercalibration network and on the process of the intercalibration exercise' (CIS Guidance Document No. 6), which served as a basis for the establishment of a register of intercalibration sites. The first phase of the intercalibration has been carried out following CIS Guidance Document No. 14 "Guidance on the Intercalibration Process 2004-2006", which was published in 2005. The results of the first phase showed a number of gaps and uncertainty as to the comparability of results. For the second phase of intercalibration (2008-2011) there was therefore a need to update the guidance.

A Drafting Group was created under the auspices of Working Group A on Ecological Status (ECOSTAT). The Drafting Group was coordinated by the Joint Research Centre and involved a number of experts from different Member States. This document is the revised Guidance No. 14 taking into account the experiences and the results of the first round of intercalibration, ongoing discussions in WG ECOSTAT, and the recommendations of the expert networks on lakes, rivers, and coastal and transitional waters.

The Water Directors have examined and endorsed the main text and Annexes I, II, and IV of this guidance during their informal meeting under the Swedish Presidency in Malmö (30 November - 1 December 2009) and asked the Geographical Intercalibration Groups to continue the work on this basis. The Water Directors endorsed the Annexes III and VI of this guidance during their informal meeting under the Spanish Presidency in Segovia (27-28 May 2010). Annex V of this guidance was endorsed by the Water Directors during their informal meeting under the Belgian Presidency in Spa (2-3 December 2010), thereby completing the endorsement of this guidance.

We would like to thank the Drafting Group for preparing this high quality document. We are convinced that the amended guidance for phase 2 intercalibration will help streamline intercalibration across the different geographic regions of Europe and improve the comparability of the intercalibration results. We strongly believe that this and other Guidance Documents developed under the Common Implementation Strategy will play a key role in the process of implementing the Water Framework Directive.

December 2010

Table of Contents

BACKGROUND AND PURPOSE OF THIS DOCUMENT 7

1. KEY PRINCIPLES OF THE INTERCALIBRATION PROCESS 9

2. STEPS OF THE INTERCALIBRATION PROCESS 15

3. CONTENTS OF THE TECHNICAL INTERCALIBRATION REPORT 33

4. ORGANISATION OF THE WORK AND TIMETABLE 35

ANNEX I: LIST OF GEOGRAPHICAL INTERCALIBRATION GROUPS (GIGS) 39

ANNEX II: RECOMMENDATIONS ON THE ESTABLISHMENT OF A COMMON DATASET FOR INTERCALIBRATION 43

ANNEX III: GUIDANCE FOR DERIVING REFERENCE CONDITIONS AND DEFINING ALTERNATIVE BENCHMARKS FOR INTERCALIBRATION 47

ANNEX IV: THE DEVELOPMENT OF A BOUNDARY SETTING PROTOCOL FOR THE PURPOSES OF THE INTERCALIBRATION EXERCISE 59

ANNEX V: DEFINITION OF COMPARABILITY CRITERIA FOR SETTING CLASS BOUNDARIES 67

ANNEX VI: REPORTING TEMPLATE FOR THE MILESTONE REPORTS 89

GLOSSARY 99

Background and purpose of this document

1. The first phase of the intercalibration (IC) has been carried out following CIS Guidance Document No. 14 “Guidance on the Intercalibration Process 2004-2006”, published in 2005. It contained key principles of the intercalibration exercise, a framework for deriving class boundaries consistent with the Water Framework Directive's (WFD) normative definitions, process options for intercalibration, the contents of the final intercalibration report, the organisation of the work and its timetables, and the composition of the GIGs.

2. During and after completion of the first round of intercalibration, several additional documents were added addressing specific aspects and/or problems that were encountered:
   - Class boundary setting protocol was agreed outlining the general principles of boundary setting in compliance with the WFD normative definitions (latest version: 1.2 of 6 June 2005). This document was used as the basis for the reporting templates for the GIG ‘milestone reports’;
   - Discussion document on the comparability of the intercalibration results – presenting an analysis of the results and summarising the comparability of the GIG results. It contains recommendations for improving the level of comparability for future IC exercises.[1]

3. The results of the first round of intercalibration are laid down in the Commission Decision of 30 October 2008[2] that was accompanied by the following documents:
   - Intercalibration technical reports;
   - The ‘intercalibration guidelines’ to translate the IC results into national methods and to derive reference conditions;
   - Work plan for future intercalibration, aiming for complete cover of all quality elements by 2011 in time for the second round of River Basin Management Plans (RBMP).

4. The results of the first intercalibration showed a number of gaps:
   - Transitional waters were not intercalibrated at all and for other water categories, some quality elements were missing (e.g. fish and macrophytes for rivers, and macroinvertebrates and phytobenthos for lakes);
   - Some of the results did not cover the full biological quality element (BQE) but only parts of them (e.g. phytoplankton in lakes and coastal waters; macroalgae and angiosperms in some coastal GIGs);
   - The results for some of the GIGs did not include all the participating Member States;
   - In some cases a close look at the results also cast doubt on the degree of comparability achieved;

---

[1] Bund van de W., Poikane S., Rodriguez Romero J. Comparability of the results of the intercalibration exercise - summary of responses and way forward. Discussion document. January 2008.

[2] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:332:0020:0044:EN:PDF

- There were gaps in the coverage of water body types and pressures;
- There was a lack of comparability in the application of criteria for setting reference conditions and class boundaries.

The aim of the second phase of intercalibration is to close these gaps and improve the comparability of the results in time for the second river basin management plans due in 2015.

5. For the second phase an update of the CIS Guidance Document No. 14 “Guidance on the Intercalibration Process 2004-2006” is needed, taking into account the experiences of the first phase. The purpose of the present document is to provide further detailed guidance for the intercalibration process continuation, which has already started in 2008 and will continue up to the end of 2011. The guidance is based on the previous Intercalibration Guidance, taking into account the experiences and the results of the first round of intercalibration, ongoing discussions in WG ECOSTAT, and the recommendations of the expert networks on lakes, rivers, and coastal and transitional waters.

1. Key Principles of the intercalibration process

Aims of intercalibration

1. The intercalibration process is aimed at ensuring comparability of the classification results of the WFD assessment methods developed by the Member States for the biological quality elements[3]. 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[4]. In the frame of the intercalibration exercise compliance of Member States assessment methods with the provisions of the Directive are checked.

2. The essence of intercalibration is to ensure that the high-good and the good-moderate boundaries in all Member States' assessment methods for biological quality elements correspond to comparable levels of ecosystem alteration. In this way, the intercalibration process described in this guidance is aimed at identifying and resolving:
   - Any 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 Water Framework Directive;
   - Any significant incomparability between the values established for the good status class boundaries by different Member States.

3. In the first phase of the intercalibration process an intercalibration register[5] was established for a limited number of water body types consisting of sites representing boundaries between the quality classes high-good and good-moderate. These were based on the WFD normative definitions. The intention was to compare the class boundaries of the Member States at those sites. The first intercalibration exercise (2004-2007) showed that, generally, the data was not sufficient and that a larger data set is needed that should ideally cover the whole gradient of the pressure. In the second phase of intercalibration (2008-2011) Member States may continue to use the data from the sites of the intercalibration register, but there will be no specific role for the register in this phase.

---

[3] The WFD describes intercalibration in Annex V, 1.4.1. using the term 'to ensure the comparability of monitoring systems'. The term 'monitoring system' in the way it is commonly used includes the whole process of sampling, measurement and assessment including all quality elements (biological and others). The term 'monitoring system' in WFD Annex V, 1.4.1. should be interpreted to mean only the biological assessment, applied as a classification tool, the results of which can be expressed as EQR. To be clear, this guidance uses the term 'WFD assessment method' instead of the term 'monitoring system'.

[4] WFD Annex V, 1.4.1 (ii), (iii), (iv), (vi)

[5] Commission Decision of 17 August 2005 (2005/646/EC): http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2005:243:0001:0048:EN:PDF

4. In Phase 2 the gaps assessed in the first phase of intercalibration should be closed. Any biological quality elements that have not been intercalibrated or not fully intercalibrated in the first phase (for example, phytoplankton) should be fully intercalibrated in Phase 2. Furthermore, all Member States in the GIG need to participate in the intercalibration. In order to improve the comparability of the results, the intercalibration procedure has been refined, now defining more clearly the individual intercalibration steps and introducing a number of checking criteria (details described in section 2).

5. Although priority should be given to the quality elements for which intercalibration has not been completed in the first phase, it will be necessary to check if the results for BQEs that have been fully intercalibrated in Phase 1 are in agreement with the criteria defined in this guidance and to review the intercalibration results following the procedure described in this guidance where the criteria are not fulfilled. The results of the checking and, where necessary, the review should be discussed and endorsed by the ECOSTAT, the SCG and the WFD Committee. The results need to be reported by the GIGs in the format requested (Annex VI).

WFD-compliant assessment methods

6. In principle, only results from WFD-compliant assessment methods can be intercalibrated (a list of checking criteria is given in section 2.1). Where methods are only partially developed, Member States may use parameter level methods for a partial intercalibration. The results of both the full and the partial intercalibration will be documented in the Technical Report. However, the COM Decision will in principle only include the results of the full intercalibration at the BQE level (compare point 22).

7. Should it turn out that – based on existing scientific knowledge – it is not possible to develop a WFD-compliant method for a BQE (i.e. only parameters can be developed and intercalibrated) then this needs to be discussed at the BQE level and a conclusion should be drawn firstly within the GIG (comparison of methods across MS). The conclusion should be supported by scientific arguments explaining why this is not possible. This needs to be checked and verified by the BQE lead. The issue should then be discussed at the cross-GIG level and at WG ECOSTAT. The IC Steering Group would need to support the conclusion. It could then be discussed whether this conclusion should be included in the COM Decision (compare point 22). The decision on which BQEs can actually be intercalibrated in Phase 2 will need to be taken very early in the process, preferably by the end of 2009, of which a last revision may be possible by mid-2010.

8. In two cases a BQE consists of two components: 1) "macrophytes and phytobenthos" for rivers and lakes, and 2) "macroalgae and angiosperms" for transitional and coastal waters. Macrophytes and phytobenthos react at different time and spatial scales, e.g. macrophytes generally react within years to changes

in pollution whereas phytobenthos can react within days or even hours. Furthermore, macrophytes react on larger spatial scales than phytobenthos. Depending on the type and magnitude of the existing pressure(s) it may be sufficient to use only one of the two components. In other cases it may be necessary to use both to get a fuller and clearer picture of the impacts or the responses to a given measure. It is up to the Member State to decide how it develops its methods. If only one component is used then it must be demonstrated that the impacts of the existing pressures are being sufficiently detected by that component. The same applies to macroalgae and angiosperms.

9. If the assessment methods developed by a Member State differ so much that the data cannot be compared and therefore the assessment method cannot be intercalibrated by one of the options provided in this guidance, then the Member State (in collaboration with the GIG) will need to find an alternative intercalibration approach. The alternative approach will need to be approved by WG ECOSTAT. If no alternative method can be found, the Member State will need to carry out an on-site comparison (comparative field exercise on a selected number of sites). The results of the IC based on this field exercise must be approved by the GIG.

Practical implementation

10. The intercalibration exercise is undertaken within Geographical Intercalibration Groups (GIGs) rather than the ecoregions defined in Annex XI of the Water Framework Directive. This is to enable intercalibration between a greater number of Member States. GIGs consist of Member States sharing common intercalibration types. It is also possible to undertake the exercise in one EU-wide GIG with the establishment of one central database and development of common intercalibration metrics. A full list of the GIGs is provided in Annex I.

11. Within each GIG 'common intercalibration types' have to be selected for intercalibration based on factors described in the WFD (Annex II, 1.2). These common intercalibration types should cover the main surface water types occurring in the GIG. Member States need to identify which national types correspond to the common intercalibration types. The common intercalibration types should be shared by at least two countries in the GIG and should be sufficiently common to allow for a meaningful comparison. The common intercalibration types defined in the first intercalibration phase should be reviewed and adapted as necessary. If there are main surface water types missing, it may be necessary to define new common intercalibration types. This is to be decided within the GIG. For those surface water types that are not intercalibrated in the intercalibration exercise, the IC boundaries of high-good and good-moderate status classes need to be translated accordingly. If a significant number of national types do not match the common intercalibration types, then this has to be reported to WG ECOSTAT.

12. The intercalibration exercise is focused on combinations of common intercalibration types, biological quality elements and specific pressures or specific combinations of pressures. The selection of these combinations should cover the major pressures occurring in the GIG. Major pressures that have not been covered in the first intercalibration need to be included in the second phase.

13. It is important to ensure that the reference conditions of the surface water types being intercalibrated are comparable. The definition of the reference conditions must correspond to the criteria given in the REFCOND Guidance. If natural or near-natural reference conditions are not available or cannot be reliably derived for a certain type (for example, for large rivers) intercalibration needs to be carried out against an alternative reference / alternative benchmark (e.g. good ecological status for that surface water type). Annex III contains guidelines for deriving reference conditions and alternative benchmarks.

14. The first intercalibration showed that the definition of common intercalibration types and the pressures acting upon them, the definition of reference conditions, and the criteria for assessing the comparability of boundaries need to be improved. As these are essential elements of the intercalibration process it is of utmost importance that they are based on sound definitions shared by all Member States in the GIG. They should be agreed and validated at the cross GIG/BQE level before the implementation begins.

15. As in Phase 1, intercalibration in Phase 2 will focus on the intercalibration of good ecological status. Good ecological potential (GEP) will not be intercalibrated in Phase 2 due to the complexity of defining GEP and the fact that the procedure how to intercalibrate GEP is not yet clear.

16. In certain cases data from HMWBs/AWBs can nonetheless be used for the intercalibration of good ecological status:
    - Where the BQE to be intercalibrated is not impacted by the hydromorphological conditions leading to the designation of the HMWB or AWB, the BQE can be intercalibrated (e.g. phytoplankton in reservoirs). This means that the maximum ecological potential of the BQE is comparable to the reference conditions of the corresponding natural type. These HMWBs or AWBs should be intercalibrated separately from natural surface water types, i.e. they should be treated as separate common intercalibration types (e.g. a certain type of reservoir).
    - Where the BQE to be intercalibrated is impacted by the hydromorphological conditions leading to the designation of a HMWB or AWB (e.g. benthic invertebrates in diverted streams), data may be used from all water bodies (including HMWBs/AWBs) in order to cover the whole gradient of hydromorphological alterations. It is important to note that in this case ecological status is intercalibrated, not ecological potential.

17. This guidance describes three different options that can be used for intercalibration of WFD-compliant methods. Where Member States have not yet defined the high-good and good-moderate boundaries, intercalibration may be used to define these.

18. The choice of the appropriate intercalibration option depends on how comparable the approaches of the national methods are:
    - Option 1: same data acquisition and same numerical evaluation means that Member States are using a common assessment method and intercalibration then concentrates on the harmonisation of reference conditions and class boundary comparison/setting;
    - Option 2: different data acquisition and numerical evaluation requires the development of common metrics for intercalibration;
    - Option 3: similar data acquisition but different numerical evaluation necessitates direct comparisons (Option 3) in which the pairwise differences of national assessment results are investigated. Common metrics are highly recommended as a supporting approach to evaluate the influences of biogeographical differences, the definition of reference conditions and the actual boundary setting.

19. The results of the intercalibration exercise are expressed as Ecological Quality Ratios (EQRs), which link class boundaries to type-specific reference conditions. The calculation of EQRs varies depending on how a particular parameter responds to changes in water quality (detailed explanations are given in Chapter 2). Because of these differences in calculation methods among others, it is not possible to compare the values of the EQRs across methods and biological quality elements. Therefore, intercalibration is not about agreeing common EQR values for the good status class boundaries but on demonstrating that those boundaries represent a comparable level of anthropogenic alteration to the biological quality element.

Organisation and time-table

20. The time-table laid down in Chapter 4 needs to be followed closely to ensure the timely completion of the intercalibration exercise.

21. The intercalibration is steered through a bottom-up process with the main work being carried out in the GIGs. In addition, BQE leads have been established to address cross-GIG issues related to BQE-specific assessment methods. The BQE groups should also steer the process of the review of intercalibration results of Phase 1. The water category leads address issues across GIGs and BQE groups. A description of the groups' tasks and responsibilities is given in Chapter 4. The Intercalibration Steering Group consists of the water category leads as well as other experts, e.g. GIG leads and/or BQE leads, and is chaired by the Joint Research Centre. The Steering Group will be used as a review panel to check on the implementation of the intercalibration and to resolve issues that

cannot be solved at the GIG or BQE level. Any issue that cannot be resolved must ultimately be brought forward to WG ECOSTAT. Should any issues arise that cannot be resolved by WG ECOSTAT, then these will be forwarded to the Strategic Co-ordination Group and/or WFD Committee, as appropriate.

22. The GIGs are obligated to report on the results of the intercalibration including the review of intercalibration results of Phase 1 for all BQEs. The results of the intercalibration exercise will be discussed and agreed at WG ECOSTAT and then forwarded to the Strategic Co-ordination Group and the WFD Committee for approval. Once approved the Commission will decide on the adoption of the results and publish them in a Commission Decision thereafter. The "Technical Report on the WFD Intercalibration Exercise" will be prepared by JRC based on the reports of the GIGs and will describe in detail how the intercalibration exercise has been carried out in each GIG.

23. After completion of the intercalibration exercise it is the obligation of the Member States to translate the results of the intercalibration exercise into their national classification systems in order to set the boundaries between high and good status and between good and moderate status for their national types. For some types that are either very specific (e.g. volcanic lakes) or very rare (e.g. some large lake type that occurs only once within the Member States) or even unique in Europe, it may not be possible to translate the intercalibration results to that type. In such cases an explanation should be given for each type why this is not possible.

2. Steps of the intercalibration process

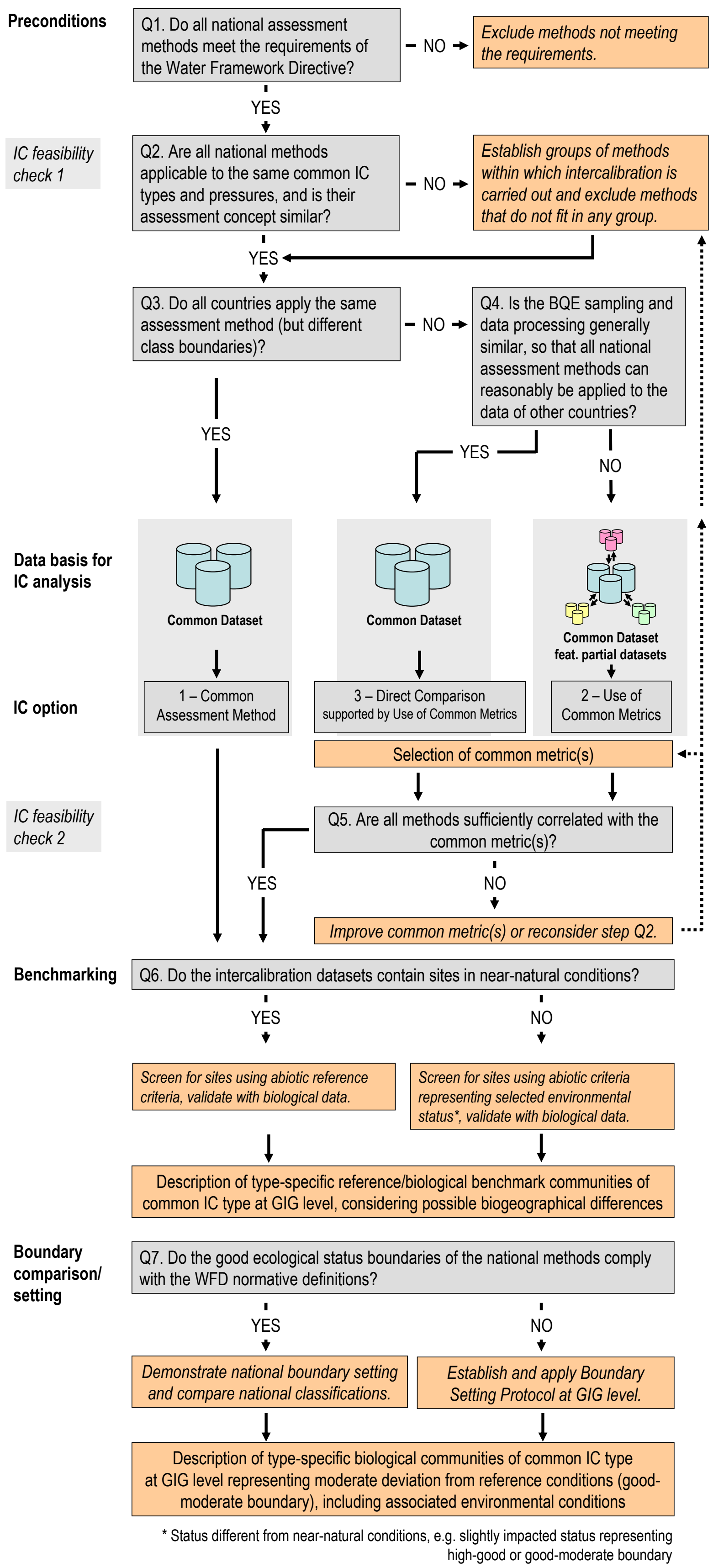


Figure 1: Flow chart of the main steps of the intercalibration process

This chapter describes the general approach of the technical intercalibration process for the second round of intercalibration. The main steps of this process are presented in the flow chart depicted in Figure 1. The questions that are asked in the flow chart serve the purpose of performing four basic checks for the identified necessary steps of the intercalibration exercise:

- **Preconditions check**: Check the compliance of national assessment methods with the WFD requirements with the help of **WFD compliance criteria**;
- **Intercalibration feasibility check**: Screening of Member States' assessment methods for acceptance in the current intercalibration exercise with the help of **method acceptance criteria**;
- **Data set check**: Evaluation of Member States' datasets for inclusion in common dataset / boundary calculations with the help of **data acceptance criteria;**
- **Comparison of boundaries**: Assess level of agreement of boundaries with the help of **comparability criteria**.

These checks are related to the main tasks of the intercalibration process that comprise:

- Documentation of national assessment methods including response to pressures and class boundary setting (Q1, Q2);
- Evaluation of general method comparability for intercalibration ("IC feasibility checks") (Q2, Q3, Q4, Q5);
- Collation of common intercalibration dataset (Chapter 2.3);
- Definition of intercalibration reference conditions/benchmark including description of the respective biological community (Q6),
- Common boundary setting / analysis of boundary comparability (Q7);
- Description of biological communities at conditions representing the harmonised good-moderate boundary ("borderline conditions") (Q7).

2.1 - Preconditions for intercalibration: WFD compliance criteria

Q1. Do all national assessment methods meet the requirements of the Water Framework Directive?

In principle, only methods meeting the requirements of the WFD can be intercalibrated (compare key principle no. 6). The first step in the intercalibration process requires the checking of national methods considering the following WFD compliance criteria. If the criteria are not met, the methods will be excluded from the next step.

Status classification:

- Ecological status is classified by one of five classes (high, good, moderate, poor and bad);

- High, good and moderate ecological status are set in line with the WFD's normative definitions (Boundary setting procedure).

Numerical evaluation:

- All relevant parameters indicative of the biological quality element are covered (see Table 1). A combination rule to combine parameter assessment into BQE assessment has to be defined. If parameters are missing, Member States need to demonstrate that the method is sufficiently indicative of the status of the QE as a whole.
- Assessment is adapted to intercalibration common types that are defined in line with the typological requirements of the Annex II WFD and approved by WG ECOSTAT.
- The water body is assessed against type-specific near-natural reference conditions.
- Assessment results are expressed as EQRs.

Data acquisition (i.e. sampling and data processing):

- Sampling procedure allows for representative information about water body quality/ecological status in space and time;
- All data relevant for assessing the biological parameters specified in the WFD's normative definitions are covered by the sampling procedure;
- Selected taxonomic level achieves adequate confidence and precision in classification.

Table 1: Indicative parameters to be included in biological assessment methods for the surface water categories and BQEs ([a] or depth distribution/cover for macroalgae and angiosperms, [b] only lakes, [c] only macroalgae, [d] bioaccumulation-bioassays). The table gives an overview of the normative definitions in the WFD and of the parameters mentioned in the CIS Guidance No 7 - Monitoring (WG 2.7) (optional issues are put between brackets).

Surface Water Category	Biological Quality Element	Taxonomic composition	Abundance [a]	Disturbance sensitive taxa	Diversity	Age structure	Frequency and intensity of algal blooms	Biomass	Absence of major taxonomic groups	Taxa indicative of pollution
Rivers and Lakes	Phytoplankton	x	x				x	x[b]		
	Macrophytes and Phytobenthos	x	x							
	Benthic invertebrate fauna	x	x	x	x				x	
	Fish fauna	x	x	x		x				
Transitional Waters	Phytoplankton	x	x				x	x		
	Macroalgae	x	x							
	Angiosperms	x	x							
	Benthic invertebrate fauna	x	x	x	x					x
	Fish fauna	x	x	x						(x[d])
Coastal Waters	Phytoplankton	x	x		(x)		x	x		
	Macroalgae and Angiosperms		x	x	(x)					
	Benthic invertebrate fauna	x	x	x	x			(x)		x

Task 1 (Member State):

- Description of national assessment method.

The necessary information will be compiled in a joint questionnaire of the WISER project and the intercalibration exercise. The questionnaire should be filled in by each Member State per BQE and water category. The information should be compiled by the GIG in an overview table and will be part of the intercalibration technical report.

Task 2 (GIG):

- Collation and evaluation of national descriptions concerning WFD requirements.

2.2 - Methods' intercalibration feasibility check: method acceptance criteria

Q2. Do all national methods address the same common type(s) and pressure(s), and follow a similar assessment concept?

The intercalibration process ideally covers all national assessment methods within a GIG. However, the comparison of dissimilar methods ("apples and pears") has clearly to be avoided. Intercalibration exercise is focused on specific type / biological quality element / pressure combinations. The second step of the process introduces an "IC feasibility check" to restrict the actual intercalibration analysis to methods that address the same common type(s) and anthropogenic pressure(s), and follow a similar assessment concept.

Typology criteria as restricting factor

- At first, the existing intercalibration typology should be reviewed. Are the common type descriptions suited for the specific BQE intercalibration exercise? And are all major types in the GIG covered?
- The individual type allocations of each country (national type to common intercalibration type) need to be checked by the MS.

Pressure criteria as restricting factor

Moreover, the exercise has to consider that pressure specific assessment approaches feature distinct characteristics. Organic pollution and hydromorphological degradation, for instance, show different effects on aquatic communities. While for the latter the alteration or loss of habitats is decisive, the impact of oxygen depletion is most relevant in organically polluted water bodies. These effects are often more dominant, superimposing the influence of habitat alteration. Methods designed to assess individual pressures often follow different assessment concepts.

In the intercalibration process the consequences of these differences need to be evaluated. The process flowchart (Figure 1) links this step to the second "IC feasibility check" that investigates the comparability based on the results of data analyses (see Chapter 2.4).

Assessment concept criteria as restricting factor

- Different community characteristics - structural, functional or physiological - can be used in assessment methods which can render their comparison problematic. For example, biodiversity indices may give a different view on structural characteristics of the community compared to species composition indices.
- In several cases, the concept of the method requires more specific typology issues to be taken into account to ensure comparability of results, e.g., in lakes it may be necessary to define the water body zone from which the samples were taken. Lake macroinvertebrate assessment systems may focus on different lake zones - profundal, littoral or sublittoral - and subsequently may not be comparable.
- Additional important issues may be the assessed habitat type (soft-bottom sediments versus rocky sediments for benthic fauna assessment methods) or life forms (emergent macrophytes versus submersed macrophytes for lake aquatic flora assessment methods).

Task 3 (Member State):

- Demonstration of applicability of national method to common IC type, coverage of pressure-impact relationship and of similarity of assessment concept of national method with those of other Member States in the GIG (to be included in general method description, see Task 1).

Task 4 (GIG):

- Compilation of groups with similar assessment methods, and evaluation of "outlying" methods.

Assessment method criteria: numerical evaluation as restricting factor

Q3. Do all countries apply the same assessment method (but different status classifications)?

The "same assessment method" means the application of identical protocols of sampling and data processing, and the use of the same numerical evaluation for classification. The latter means that:

- The same BQE parameters are used in the assessment and their results are combined in a similar way up to the BQE level;
- The assessment has to be comparable at the spatial scale (combination of assessments at sample level / assessment at water body level) or the temporal scale (combination of temporal series).

In this case the intercalibration efforts only need to concentrate on the harmonisation of status classifications, i.e. the definition of reference conditions and the setting of the boundaries between high and good, and good and moderate ecological status, respectively.

Task 5 (GIG)

- Evaluation of national method descriptions with regard to data acquisition and numerical evaluation.

Assessment method criteria: data acquisition as restricting factor

Q4. Is the BQE sampling and data processing generally similar, so that all national assessment methods can reasonably be applied to the data of other countries?

Most assessment methods are adapted to regional conditions and often follow national traditions. In case the answer on Q3 is negative, this means that different techniques of numerical evaluation are applied (e.g. focus on different aspects of the biological community being reflected in the selection and combination of biological metrics, different choice and scoring of indicator taxa).

The process step Q4 interrogates the scope of the base data sampled according to the national protocols:

- Do the raw taxa lists contain all biological information required by the individual methods? Are the required levels of taxonomic precision similar?
- Are spatial and temporal requirements (e.g. sampling season or minimum sampling area) met?

If this is the case, national methods can reasonably be applied to the base data of other countries. However, the effect of biogeographical differences has to be determined, and data may need to be harmonised/adjusted (e.g. taxonomic adjustment, abundance scale conversions) (see Annex II).

Task 6 (GIG)

- Evaluation of national method descriptions with regard to data acquisition.

2.3 - Data base for intercalibration analysis

Generally, BQE groups within a GIG shall collate a common dataset for intercalibration. Central data collection and analysis facilitate a reproducible and transparent intercalibration process considerably. The quest for intercalibration solutions is often a laborious task including trial and error. Central data processing allows in-depth examination and testing of various approaches. The common dataset is also the basis for the description of biological communities of the intercalibration types (see Chapters 2.5 and 2.6) and should therefore be comprehensive and representative for the common IC types. The common dataset should furthermore:

- sufficiently cover the geographical area in which the common type occurs within the GIG,
- encompass sampling sites covering the entire gradient of the pressure to be intercalibrated, and hence the complete ecological quality gradient ranging from high to poor ecological status, and
- contain non-biological (environmental) and biological data to conduct pressure-impact analyses.

The collection of the dataset involves several difficulties and limitations, e.g., there may be considerable differences among datasets from different countries regarding sampling, analytical methods and taxonomic precision; this may reduce the comparability of the data and increase the uncertainty of the results. So it is important to agree on criteria for minimum data requirements and data quality criteria in order to obtain comparable datasets.

Data acceptance criteria have to include the following aspects:

- Data requirements (obligatory and optional), e.g., providing physico-geographical parameters for checking type allocations (e.g. altitude, alkalinity, mean depth for lakes);
- The sampling and analytical methodology;
- Level of taxonomic precision required and taxalists with codes;
- The minimum number of sites / samples per intercalibration type
- Sufficient covering of all relevant quality classes per type;
- Other aspects where applicable.

The datasets should be screened by the GIG and a report describing the full list of acceptance criteria with the evaluations of whether the Member State met each of the required criteria should be compiled. Data that only partly fulfill the required criteria should be clearly identified, and the differences between the quality of the various datasets should be clearly mentioned and kept in mind all along the process of intercalibration (from the analysis of the data to the presentation of the results).

If the data acquisition is significantly different between countries, GIGs shall nevertheless aim at establishing a common dataset. In this case it can be composed of partial datasets, i.e. national subsets that fulfil the data requirements of particular methods. The common database, for instance, may contain monitoring data at the taxonomical level of family for most of the countries but include data at species level for some countries. Methods requiring species level information cannot be applied to family level data. However, their assessment can be related to common metrics at family level. Common metric development can be done using the complete dataset, benchmarking (Chapter 2.5) and boundary comparison/setting (Chapter 2.6) has to be done using the partial dataset. The biological communities of the common intercalibration types need to be described based on the “least common denominator” (in this case: family level data).
For further details on the common dataset see Annex II.

Task 7 (Member States):

- Providing required data for the intercalibration dataset.

Task 8 (GIG):

- Establishment of a common taxonomical checklist (taxa names and codes) for the needs of the Intercalibration;

- Collation of the common intercalibration dataset including biological and pressure data. For some issues (e.g. description of types, reference criteria and conditions, pressures etc.) the collation of common datasets useful for various GIGs shall be privileged as much as possible;
- Data access and storage etc.

2.4 - Intercalibration options

The choice of the appropriate intercalibration option depends on how comparable the national methods are the following:

(a) Same data acquisition, same numerical evaluation → IC Option 1
The national techniques of data sampling, processing and evaluation are the same and all countries in the GIG are using the same assessment method. The intercalibration exercise can concentrate on the harmonisation of reference conditions and class boundary comparison/setting. IC Option 1 represents the most straightforward option since the difficulties and uncertainties involved in comparing the results of different assessment methods are avoided.

(b) Different data acquisition and numerical evaluation → IC Option 2
If data sampling and evaluation procedures are significantly different between countries the use of common metrics for intercalibration is necessary. Common metrics can be selected from the national assessment methods and other existing biological indices, or they can be generated for the intercalibration exercise (see Birk & Willby, 2009[6]). Ideal common metrics have to:
- cover all relevant parameters indicative of the BQE;
- respond to the pressures being intercalibrated;
- be ecologically meaningful (i.e. clearly be related to common ecological principles);
- show no (or only minor) bias due to biogeographical differences or differences in national sampling protocols.

(c) Similar data acquisition, but different numerical evaluation → IC Options 3 supported by the use of common metric(s)
In the direct comparison (IC Option 3) the pair-wise differences of national assessment results are investigated at sampling site level or water body level. This method allows for a comprehensive analysis of Member States' classifications of sites, including reference sites, if available. The influence of biogeographical differences can be investigated by comparing biological data and assessment results of reference sites between Member States if these have been defined by common criteria. A proper national assessment system is likely to perform better on its own reference sites than on reference sites of other Member States. Biogeographical

[6] Birk, S. & N. Willby, 2009. Towards harmonization of ecological quality classification: Establishing common grounds in European macrophyte assessment for rivers. Submitted to *Ecological Indicators*.

differences, and the different way of Member States to account for such differences, can thus be an important source of incomparability. This may hamper the possibilities for successful intercalibration, as well as the fact that the number of reference sites is often too low to investigate the role of biogeographical differences.

The use of common metrics is a supporting approach if the biogeographical differences are estimated as large. Regression analyses of national EQRs against the common metric reveal the positions of the national reference and class boundaries on the common metric scale. This provides insight into the reasons for possible incomparabilities. Common metrics can be used as “international currencies” to which common boundary setting (including harmonised reference definition) and the GIG-wide descriptions of reference and “borderline” conditions can be related (see Chapters 2.5 and 2.6). The ecological relevance further enhances the transparency of the intercalibration process. From these supporting analyses individual intercalibration exercises may result between subgroups of Member States that are biogeographically more comparable.

Task 9 (GIG)

- Selection of most appropriate intercalibration option.

When IC Options 2 or 3 are used, the following intercalibration feasibility check is needed.

Q5. Are all methods reasonably related to the common metric(s)?

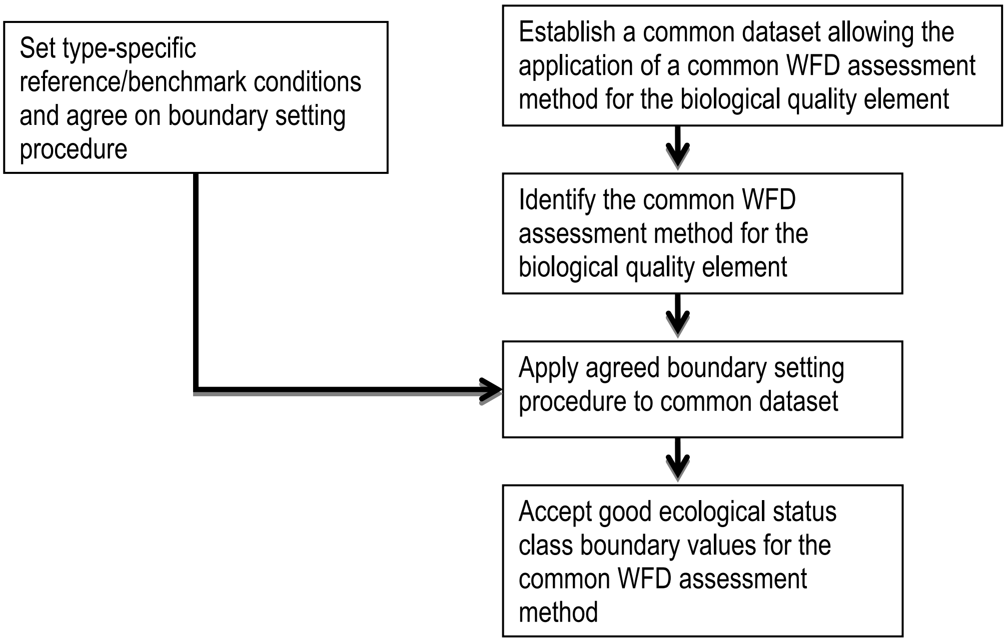
The relationship between common metric(s) and national assessment methods has to be sufficiently strong. The feasibility check can also be carried out against the mean of national EQRs when applying IC Option 3. As a general rule the process shall strive for highest correlation coefficients. For the criteria for inclusion of intercalibration results see Annex V. In such cases when relationship between common metric(s) and national assessment methods is not sufficiently strong, the common metric selection needs to be reconsidered, or the respective methods have to be excluded from the particular exercise based on well-founded explanations (“IC feasibility check 2”).

Task 10 (GIG)

- Check of “IC feasibility” and evaluation of “outlying” methods.

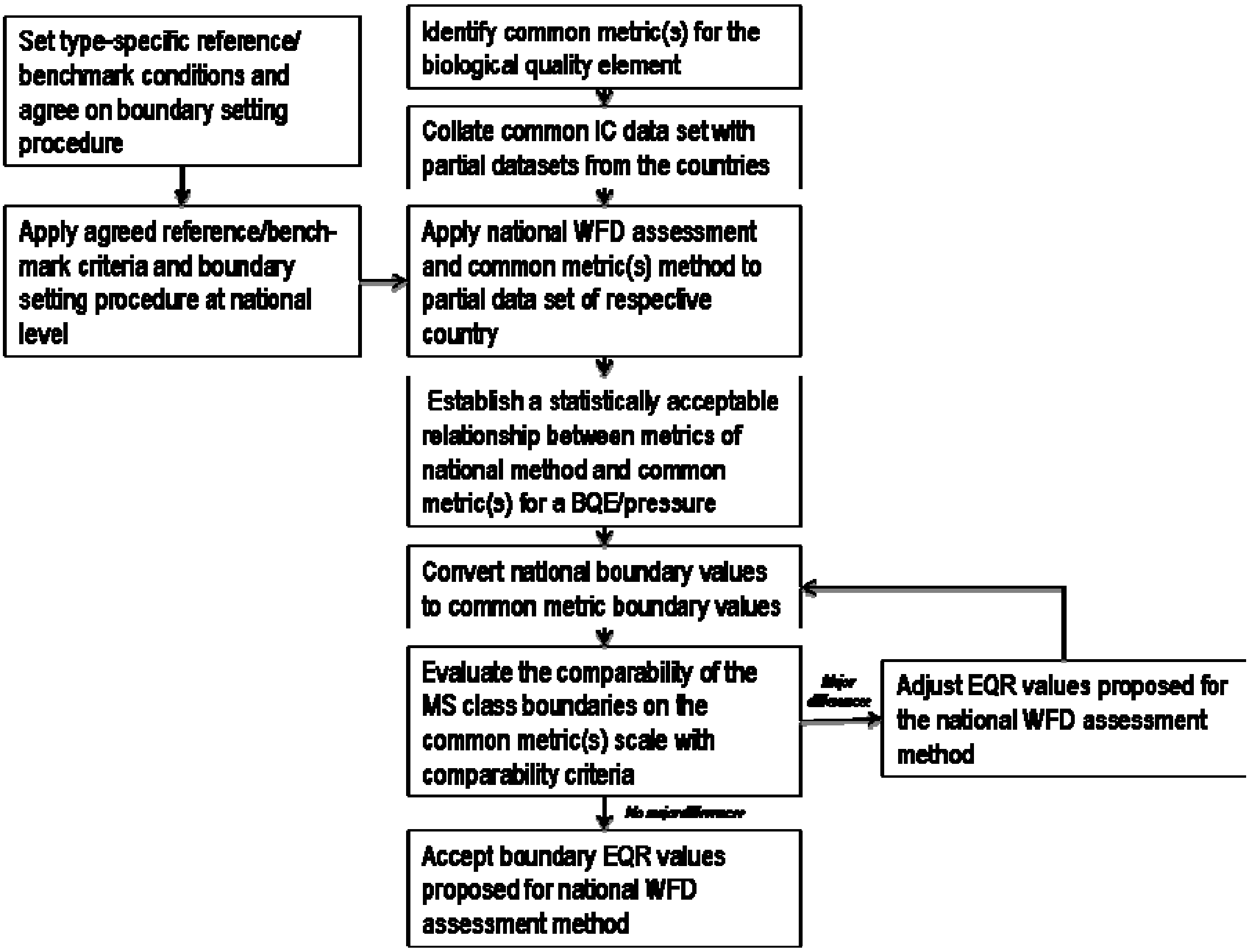
The next section outlines the different options for the process of intercalibration in more detail.

"Option 1": Common WFD assessment method



Features	All Member States in the GIG use the same WFD assessment method, and agree on high-good and good-moderate class boundaries of the EQR scale for this common method by applying the class boundary setting procedure for the common intercalibration types. No further harmonisation is required.
Application	Where Member States can agree to use the same WFD assessment method.
Data requirements	Data to demonstrate how the boundaries are set.
Advantages	The most straightforward option since the difficulties and uncertainties involved in comparing the results of different assessment methods are avoided. Comparability between Member States is assured.

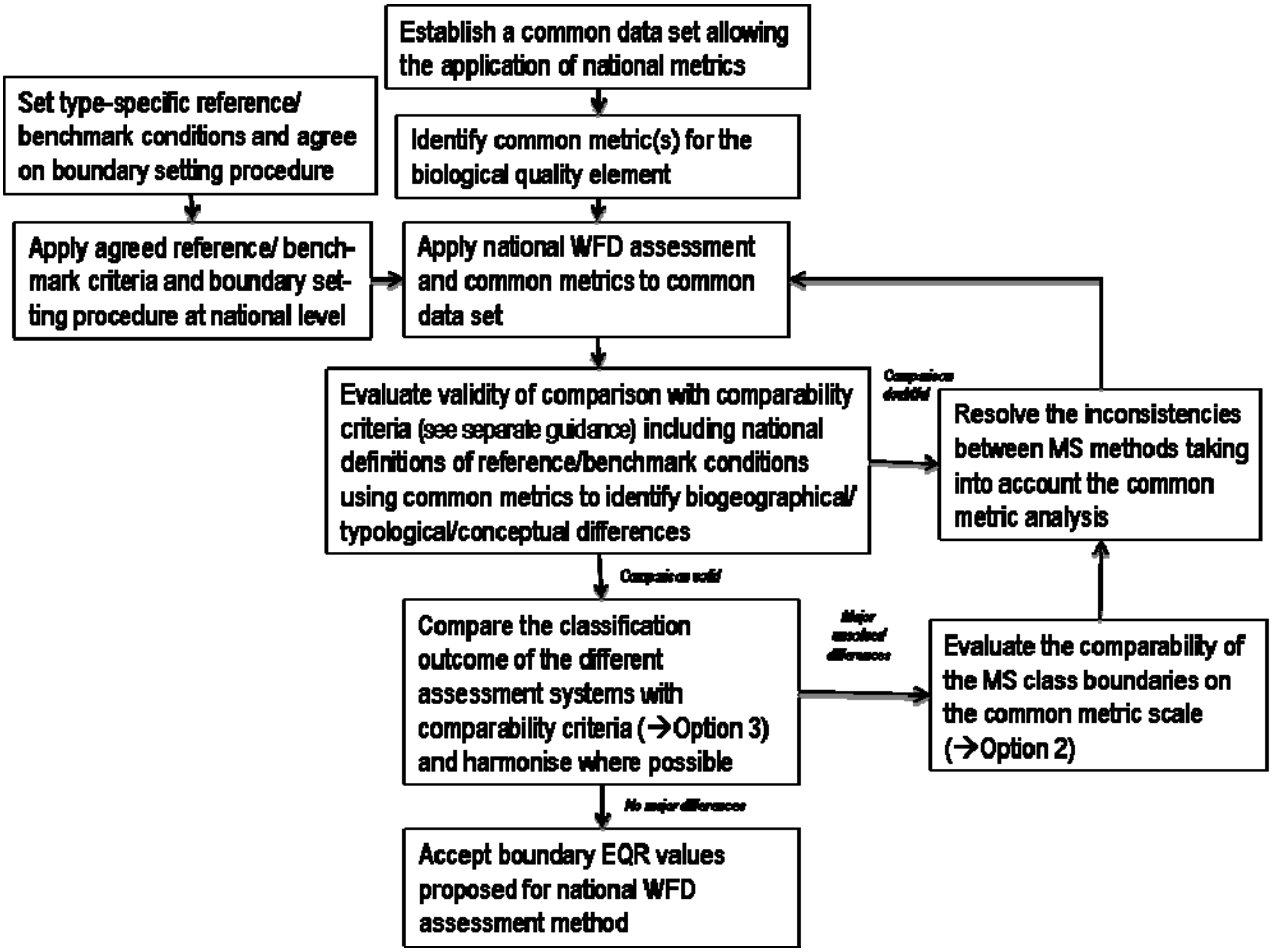
“Option 2”: Comparability of class boundaries of Member States’ methods is assessed indirectly using a common intercalibration metric(s)



Information on Option 2	
Conditions for use	All Member States in a GIG have sufficiently developed their national WFD assessment methods. Member States can agree on common metric(s) that is indicative of the relevant biological quality element, sensitive to the pressure that is assessed, and is reasonably related with the Member States’ methods. Suitable Member States’ datasets are collated in an IC database from which these common metric(s) can be calculated to enable reliable comparison between the Member States’ assessment methods, containing data from a sufficient number of reference (or benchmarking) sites allowing a reliable and comparable boundary setting.
Application	Where it is not possible to apply Option 1 (because Member States

	do not use the same assessment method) or Option 3 (due to different data acquisition and numerical evaluation). Where Option 3 is used, the use of common metric(s) is recommended as a complementary analysis to increase transparency and to provide insights in reasons for possible incomparabilities (see Chapter 2.4).
Features	In this approach boundaries are initially set by the Member State compared on a common metric(s) EQR scale, and harmonised where necessary. Common metrics enable a GIG-wide comparison of classification results.
Data requirements	The GIG establishes a common IC database with data from each Member State that allow for calculating both the national WFD assessment method and the common metric(s).
Advantages	Common metrics allow for the comparison of national good status boundaries if the data acquisition techniques are different. Common metrics provide “international currencies” to which common boundary setting and GIG-wide descriptions of reference and “borderline” conditions can be related.
Disadvantages	Because comparisons are made indirectly on an EQR scale, Option 2 can only give valid results if reference/benchmark conditions are comparable throughout the GIGs. Possible differences in classifications of different Member States’ assessment methods when applied to individual water bodies are not made transparent.

“Option 3”: (supported by the use of common metrics): Comparability of class boundaries of Member States’ methods is assessed by direct comparison of classification outcomes using a common dataset



Information on Option 3 (supported by common metrics)	
Conditions for use	All Member States in a GIG have sufficiently developed their national WFD assessment methods. Availability of suitable datasets on which Member States’ assessment method can be calculated to enable reliable application of the agreed boundary setting procedure. Availability of a means of estimating and taking into account differences in the bias of the methods when applied to the dataset referred to above. Member States agree on a common metric that is indicative of the relevant biological quality element, sensitive to the pressure that is assessed, and is reasonably related with the Member States’ methods.
Application	Except where Option 1 is available

Features	Member States apply the boundary setting procedure using their own datasets and identify the high-good and good-moderate class boundaries. Comparability is tested by checking whether there are major differences in the results given by different Member States' assessment methods when applied to the same dataset. A common metrics analysis (following Option 2) is used to help resolve inconsistencies between Member States' methods.
Data requirements	A common dataset allowing the application of all Member States' national methods, as well as the common metric(s).
Advantages	Comprehensive and robust comparison due to the combination of the direct comparison of the methods and the use of common metrics.
Disadvantages	Application of Member States' national methods outside the geographical range for which they are tested may be questionable. Data requirements are difficult to meet without making compromises. Using a common metric (Option 2) intercalibration in parallel helps to overcome these disadvantages.

2.5 - Reference/alternative benchmark conditions

In the intercalibration exercise reference or alternative benchmark conditions have to be established for the common IC types in order to be able to compare the national class boundary settings. It is important to ensure that the reference conditions of the surface water types being intercalibrated are comparable. The definition of the reference conditions must correspond to the criteria given in the REFCOND Guidance. If natural or near-natural reference conditions are not available or cannot be derived for a certain type (for example, for large rivers) intercalibration needs to be carried out against an alternative benchmark (e.g. good ecological status for that surface water type). To enhance the transparency of the intercalibration process defining reference or benchmark conditions shall be done using the common dataset. This requires finding references or benchmarks based on actual data sampled at existing sites. The availability of a comprehensive database that especially covers sites in reference or alternative benchmark conditions (pristine or impacted by similar levels of impairment) is essential. As a guideline and where possible, a minimum of 15 sites meeting these conditions per common intercalibration type should be used to make a statistically reliable estimate.

Q6. Do the intercalibration datasets contain sites in near-natural conditions?

The intercalibration benchmark shall preferably be derived from sites in near-natural reference conditions (see Figure 2). Based on a harmonised set of reference criteria abiotic data in the intercalibration dataset have to be screened for near-natural sites[7]. The biological conditions of these sites need to be reviewed to avoid the influence of impacts caused by pressures not regarded in the screening process.

For several surface water types near-natural conditions no longer exist. These types require a different benchmarking approach based on the definition of "Least Disturbed Conditions" (LDC)[8] that refer to the best available physical, chemical and biological habitat conditions given today's modified landscape (Figure 3). LDC sites have to be identified from the common intercalibration dataset. This can be done by screening for sites meeting abiotic criteria that represent a similar low level of impairment (see Birk & Hering, 2009[9]). This approach also requires the review of the biological conditions. It is important to identify the position of the benchmark on the gradient of impact, i.e. to document the deviation of the selected benchmark from reference conditions. This allows for integrating the approach into the Cross-GIG harmonisation efforts for benchmarking (see Cross-GIG activity on reference condition refinement). When appropriate, modelling approaches can be used to support the setting of alternative benchmarks.

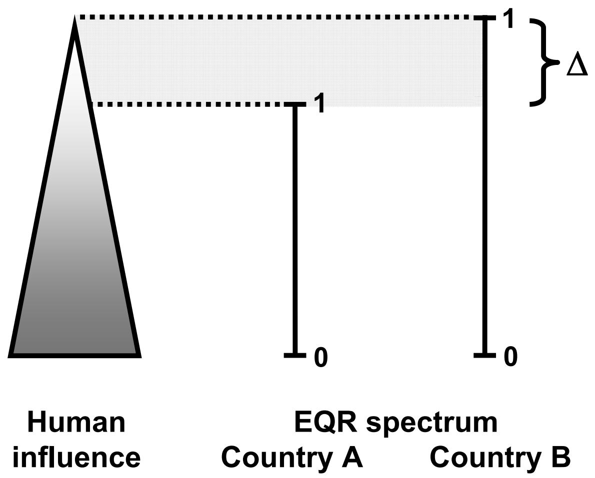


Figure 2: The importance of a common definition of near-natural reference conditions in intercalibration. If the national assessment methods of two countries refer to different levels of human influence (Δ), the same EQRs represent different levels of impairment (Figure taken from Birk & Böhmer 2007[10])

---

[7] Reference conditions to be used in the intercalibration exercise are currently reviewed by the Cross-GIG activity on reference condition refinement.

[8] Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson & R. H. Norris, 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. Freshwater Bioassessment 16: 1267-1276.

[9] Birk, S. & D. Hering, 2009. A new procedure for comparing class boundaries of biological assessment methods: A case study from the Danube Basin. Ecological Indicators 9: 528-539

[10] Birk, S. & J. Böhmer, 2007. Die Interkalibrierung nach EG-Wasserrahmenrichtlinie - Grundlagen und Verfahren. Wasserwirtschaft 9: 10-14.

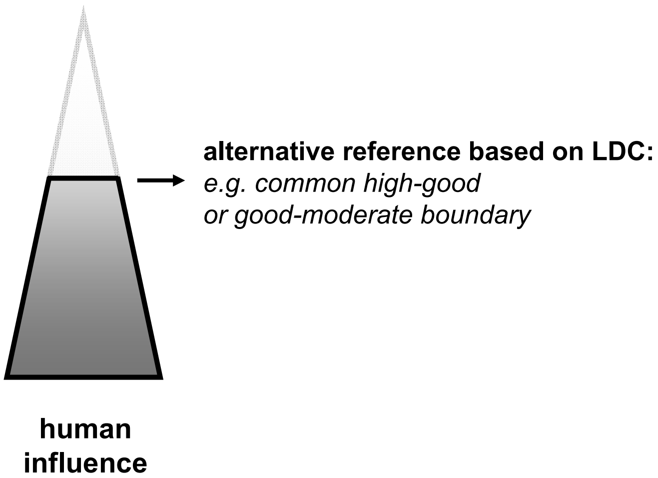


Figure 3: Definition of an alternative reference in intercalibration by using sites impacted by a similar level of impairment (Least Disturbed Conditions - LDC) instead of near-natural reference sites (Figure taken from Birk & van Kouwen 2009[11])

The biological communities at reference/benchmark conditions have to be described, considering potential biogeographical differences. The description shall be based on the analysis of sites in the common dataset, possibly confined to the "least common denominator" level of data resolution (see Annex II). It is recommended to relate these descriptions to the characteristic value ranges of the common assessment method (IC Option 1) or the common metric(s) (IC Option 2 and 3), respectively.

Task 11 (GIG):

- Definition and application of reference conditions/benchmark criteria;
- Description of intercalibration type specific reference/benchmark communities.

2.6 - Boundary setting/comparison

Q7. Do the good ecological status boundaries of the national methods comply with the WFD normative definitions?

In the final step of the process a distinction is made between ecological status classifications of national methods established either individually by the Member States prior to the intercalibration process, or jointly by the GIG based on a common approach for boundary setting (IC Option 1 and 2). The former requires a detailed demonstration of national boundary setting according to a Boundary Setting Protocol (see Annex IV). The position of national class boundaries should be reviewed also with regard to the pressure-impact relationship (see Task 3).

---

[11] Birk, S. & L. van Kouwen, 2009. Supportive analysis of the second Joint Danube Survey data (typology, intercalibration) and Technical support of the Eastern Continental Geographical Intercalibration Group. Final report. April 2009. Hamm (Sieg).

National boundary setting: comparison of national boundaries is determined using the standardised analytical procedure and harmonised comparability criteria (see Annex V). National methods not complying with these criteria have to adjust their national method, and re-enter the comparison process. The adjustments do not necessarily need to be confined to elevating the status class boundaries, but may include more profound changes on the level of data acquisition or numerical evaluation.

Joint boundary setting requires the design of a common Boundary Setting Protocol (see Annex IV). Basic element of this protocol is to establish a relation between abiotic pressure parameters and the common WFD assessment method (IC Option 1) or the common metric(s) (IC Option 2), respectively. Depending on the type of relationship the GIG should agree on the most suitable boundary setting option that needs to be applied to the national classifications.

Similar to the benchmarking step the biological communities representing the “borderline” conditions between good and moderate ecological status have to be described. This shall be done using sites of the common dataset that fall into a selected boundary range (e.g. harmonisation band of national good-moderate boundaries expressed in common metric scale).

Task 12 (Member States):

- If boundaries were set individually by the Member State: Demonstration of national boundary setting according to a Boundary Setting Protocol and boundary adjustment in case of deviation (indicated by comparability analysis);
- If boundaries are jointly set: Transfer of common boundaries into national status classification.

Task 13 (GIG):

- If boundaries were set individually: Approval of national boundary setting and performance of comparability analysis;
- If boundaries are jointly set: Elaboration and execution of Boundary Setting Protocol;
- Description of intercalibration type specific biological communities at “borderline” conditions.

3. Contents of the Technical Intercalibration Report

According to the timetable set out in the **Intercalibration work programme 2008-2011**, the final report of the intercalibration exercise should be finalised in December 2011. The final Intercalibration report will consist of the final reports of the intercalibration groups (Annex VI) and necessary considerations at GIG, BQE and cross-GIG level.
This chapter gives an outline of the expected key elements of this report following the major steps of the IC flowchart (Fig.1)

1. National assessment method:
    1.1. Description of Member States' assessment methods (See table 2 and 3);
    1.2. Results of WFD compliance check (meeting the requirements of normative definitions);
    1.3. Results of intercalibration feasibility check (compliance with method acceptance criteria for the IC feasibility: type / pressure / method concept / metrics);

2. Common intercalibration types:
    2.1. Characterisation of common IC types;
    2.2. Correspondence of national typology to common IC typology;

3. Data basis:
    3.1. Description of dataset;
    3.2. Sampling strategy and analyses methods;

4. Intercalibration option used:
    4.1. Selection of the IC option;
    4.2. Development of the IC common metrics (if applicable) or common metric used;
    4.3. Application of IC procedure to the dataset(s);

5. Reference conditions/Benchmarking:
    5.1. Description of reference/benchmark setting process;
    5.2. Description of IC type-specific reference or benchmark communities, considering possible biogeographical differences;

6. Boundary comparison/setting:
    6.1. Description of boundary setting procedure;
    6.2. Description of IC type-specific biological communities representing the "borderline" conditions between good and moderate ecological status, considering possible biogeographical differences;
    6.3. Boundary comparison and harmonisation.

7. Boundary EQR values established for the type/quality element/pressure combination for the common metric (where applicable) and each national WFD assessment method:

Member State	Classification Method	EQR High-Good boundary	EQR Good-Moderate boundary
	Common metric	0.85	0.65
MS1	Method 1	0.85	0.60
MS2	Method 2	0.85	0.75
MS3	Method 3	0.70	0.60
MS4	Method 4	0.90	0.75
MS5	Method 5	0.85	0.60

8. Open issues:
   8.1. Gaps - what is not achieved in the current intercalibration exercise;
   8.2. Possible way forward.

4. Organisation of the work and timetable

4.1 The intercalibration process will be carried out under the umbrella of WFD Common Implementation Strategy WG ECOSTAT. An overview of the intercalibration organisational structure is given in Table 2. The Member States participating in the GIGs are given in Annex I.

Table 2. Overview of the organisational structure for the intercalibration process. The Lakes, Rivers, and Coastal/Transitional Waters expert groups are subdivided into GIGs, and horizontally subdivided into BQE sub-groups that work across GIGs.

WFD Common Implementation Strategy Working Group ECOSTAT
Intercalibration Steering group (JRC water category coordinators plus additional members from GIG/BQE leads)

RIVER Intercalibration coordinator

BQE / GIG	Alpine GIG lead	Central Baltic GIG lead	Eastern Continental GIG lead	Mediterranean GIG lead	Northern GIG lead
Benthic fauna BQE lead	R-Alp-Bf lead	R-CB-Bf lead	R-EC –Bf lead	R-Med –Bf lead	R-N-Bf lead
Phytobenthos BQE lead	R-Alp-Phb lead	R-CB-Phb lead	R-EC-Phb lead	R-Med-Phb lead	R-N-Phb lead
Macrophytes BQE lead	R-Alp-Mp lead	R-CB-Mp lead	R-EC-Mp lead	R-Med-Mp lead	R-N-Mp lead
Fish fauna BQE lead	R-Alp-F lead	R-CB-F lead	R-EC-F lead	R-Med-F lead	R-N-F lead

LAKE Intercalibration coordinator

BQE / GIG	Alpine GIG lead	Central Baltic GIG lead	Eastern Continental GIG lead	Mediterranean GIG lead	Northern GIG lead
Phytoplankton BQE lead	L-Alp-Ph lead	L-CB-Ph lead	L-EC-Ph lead	L-Med-Ph lead	L-N-Ph lead
Macrophytes BQE lead	L-Alp-Mp lead	L-CB-Mp lead	L-EC-Mp lead	L-Med-Mp lead	L-N-Mp lead
Benthic fauna BQE lead	L-Alp-Bf lead	L-CB-Bf lead	L-EC-Bf lead	L-Med-Bf lead	L-N-Bf lead
Fish fauna BQE lead	L-Alp-F lead	L-CB-F lead	L-EC-F lead	L-Med-F lead	L-N-F lead

COASTAL/TRANSITIONAL WATERS Intercalibration coordinator

BQE / GIG	North East Atlantic GIG lead	Baltic Sea GIG lead	Black Sea GIG lead	Mediterranean GIG lead
Phytoplankton BQE lead	C-NEA-Ph lead	C-BS-Ph lead	C-BC-Ph lead	C-Med-Ph lead
Macroalgae/Angiosperms BQE lead	C-NEA-Mp lead	C-BS-Mp lead	C-BC-Mp lead	C-Med-Mp lead
Benthic fauna BQE lead	C-NEA- Bf lead	C-BS-Bf lead	C-BC-Bf lead	C-Med-Bf lead
Fish fauna BQE lead	C-NEA-F lead	C-BS-F lead	C-BC-F lead	C-Med-F lead

4.2 The intercalibration organisational structure includes:

- IC Groups exist for each combination of water category, GIG and biological quality element, e.g. Lake Eastern Continental Phytoplankton group (L-EC-Ph) or River Central-Baltic macrophyte group (R-CB-Mp);
- GIGs comprise regional intercalibration groups for each relevant water category and biological quality element, e.g., Lake Central-Baltic GIG (L-CB GIG) includes Lake Central-Baltic Phytoplankton, Macrophytes, Benthic fauna and Fish fauna groups;
- Biological Quality element groups (BQE groups) ensure cross-GIG cooperation within a water category e.g. Lake Phytoplankton IC group includes Alpine, Central-Baltic, Eastern Continental, Mediterranean and Northern phytoplankton groups.

One of the Member States in each IC group will act as a group coordinator responsible for the practical work. Every GIG and every BQE group will appoint a coordinator responsible for cooperation and organization of cross-group work.

In addition, there are two cross-GIG groups:

- Reference Conditions Working Group (REFCON): The task of this group is to analyse comparability of Member States' definitions of reference conditions for rivers, lakes, coastal and transitional waters and to make recommendations to Member States and GIGs how to improve comparability.
- Large Rivers Intercalibration Group: This group is developing a harmonised approach for intercalibration of very large rivers (catchments < 10.000 km²) across GIGs and BQEs.

4.3 The practical work will be carried out in the **intercalibration groups**, following the timetables set out in this guidance document. Basically IC group leads are responsible for organization of the IC process in their group:

- Collection of common dataset (recommended deadline: October 2009);
- Datasets established and common metrics developed (June 2010);
- Reference conditions/Benchmarking and boundary setting (October 2010);
- Boundary comparison and harmonisation (March 2011).

4.4 Cooperation between the IC groups at the BQE level is ensured through the **BQE leads**. Tasks of BQE leads include streamlining IC at BQE level, addressing BQE-specific problems, and ensuring the comparability of approaches taken by the IC groups:

- Validation of WFD compliance checking;
- Validation of how groups set reference conditions (together with GIG leads);
- Validation of BQE and pressure specific dataset requirements;
- Validation of common metric method elaboration by different GIGs
- Validation of IC results
- Discussions of “cross-GIG” issues: e.g. how to deal with hydromorphological pressures for Macrophyte BQE ?

4.5 Tasks of the **GIG leads** include overall coordination of the IC process:

- Organizational aspects (resources, experts);
- Coordination of all IC groups/ all BQEs in a GIG;
- Definition of the common intercalibration types and their description;
- Validation of how groups set reference conditions (together with BQE leads) to ensure the comparability of the reference conditions between the IC groups in a regional context;
- Support for collection of datasets;
- Overview of pressures addressed by different BQEs to ensure that all relevant pressures are addressed in the GIG.
- In the course of the intercalibration process, the GIGs/Intercalibration groups should regularly report the progress to WG ECOSTAT to check whether approaches followed in different GIGs are sufficiently comparable.

4.6 The Intercalibration process needs to be transparent and the results need to be coherent and consistent between regions, biological quality elements and between water categories. WG ECOSTAT and the **Intercalibration Steering group** is responsible for evaluating the results of the intercalibration exercise and making recommendations to the Strategic Coordination Group or WFD Committee, as and when appropriate. WG ECOSTAT is responsible for the consistency and harmonisation of the process between GIGs and between water categories (lakes, rivers, and coastal and transitional waters), but the Cross-GIG groups should carry out the work necessary to ensure the consistency and harmonisation of the intercalibration process.

4.7 In addition, an **IC Review Panel** should be set up consisting of the water category leads as well as some other experts, e.g. from GIGs or possibly external experts. The review panel will have such tasks as checking WFD compliance of the methods and approving the results of the intercalibration.

4.8 The intercalibration process is facilitated by the EC Joint Research Centre (JRC). JRC has established a reporting structure where IC groups report and update the results of the different steps of the IC process, and will compile the draft final technical report of the intercalibration exercise.

4.9 The Member States in the GIGs have the collective responsibility to bring together the data enabling comparison of the classification results of different countries within the GIG. Additional sampling during the IC exercise may be considered in the GIGs. The GIGs are free to specify the aggregation level and format for this data. To ensure transparency of the intercalibration process the original data source(s) should be specified, and the data should be made publicly available in such a form that the Intercalibration procedure can be verified.

4.10 JRC is responsible to regularly report the progress of the intercalibration process to the CIS Strategic Co-ordination Group, the Water Directors, and the WFD Committee.

4.11 The general timetable of the intercalibration exercise (Table 3) is constrained by the legal deadline to finalise the intercalibration report by December 2012. This requires that WG ECOSTAT agrees on the report in June 2011. WG ECOSTAT will meet twice every year and regularly provide progress reports and recommendations to the Strategic Co-ordination Group and the WFD Committee.

Table 3. Timetable of the Intercalibration - Phase 2.

Steps of the Intercalibration	**2009**	**2010 Jan-Jun**	**2010 Jul-Dec**	**2011 Jan-Jun**	**2011 Jul-Dec**	**2012 Jan-Jun**
Test preconditions						
- Apply criteria for WFD compliance						
- Apply criteria for IC feasibility						
Report to WG ECOSTAT Oct 2009	Milestone 1					
Collect IC dataset						
Design IC working procedure						
Select IC option, develop common metric						
Report to WG ECOSTAT Apr 2010		Milestone 2				
Define benchmarks (Q6)						
Compare/propose class boundaries						
Report to WG ECOSTAT Oct 2010			Milestone 3			
Boundary harmonisation						
Report to WG ECOSTAT Apr 2011				Milestone 4		
Submit final IC reports						
Final report to WG ECOSTAT Jun 2011					Milestone 5	
Formal adoption of IC results, Final report						IC Report

Reporting milestones are related to the major steps of the IC flowchart (Figure 1); further details on the contents of the milestone reports are given in Annex VI.

ANNEX I: List of Geographical Intercalibration Groups (GIGs)

GIG co-ordinator(s) are indicated in **bold.**

Geographical Intercalibration Groups

1) Rivers

Name of the Geographical Intercalibration Group	Member States being part of this Geographical Intercalibration Groups
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	Austria Bulgaria **Czech Republic** Greece Hungary Romania Slovakia Slovenia

Mediterranean	Cyprus France Greece Italy Malta **Portugal** Slovenia Spain

2) Lakes

Name of the Geographical Intercalibration Group	**Member States being part of this Geographical Intercalibration Group**
Northern	**Finland** Ireland Norway Sweden United Kingdom
Central/Baltic	Belgium Czech Republic Denmark Estonia France Germany Latvia Lithuania **Netherlands** Poland Slovakia United Kingdom
Alpine	Austria France Germany Italy Slovenia
Eastern Continental	Bulgaria Hungary **Romania**
Mediterranean	Cyprus France Greece Italy Malta Portugal Romania **Spain**

3) Transitional and coastal waters

Name of the Geographical Intercalibration Group	**Member States being part of this Geographical Intercalibration Groups**
Baltic	**Denmark** Estonia Finland Germany Latvia Lithuania Poland Sweden
North-East Atlantic	Belgium Denmark France Germany Ireland Netherlands Norway Portugal Spain Sweden **United Kingdom**
Mediterranean	Cyprus France Greece **Italy** Malta Slovenia Spain
Black Sea	Bulgaria **Romania**

ANNEX II: Recommendations on the establishment of a common dataset for intercalibration

II.1 Content

The common dataset should contain

- biological data (e.g. taxonomical composition and abundance of BQE, …),
- localisation data (country code, name and coordinates of sampling site, …),
- typological data (e.g. altitude, geology, dominant substrate, …), and
- pressure data (e.g. catchment land use, physico-chemical measurements, …).

II.2 Features

The common dataset should allow central data processing and testing of various intercalibration options. It fosters the application of the Boundary Setting Protocol (pressure-response analyses, description of biological communities at various quality states). However, the collation of a representative common dataset for intercalibration requires a laborious process (data collection, data quality control, data harmonisation).

II.3 Harmonisation of biological data in common dataset

Data sampling and processing often differs between countries (e.g. record of abundance, level of taxonomical identification). Thus, deviating data features need to be harmonised. It may be necessary to slightly adjust a national assessment method, so it can be applied to the common dataset. In such a case, the results of the adjusted method have to be related to the outputs of the original method, based on a data subset including all required data parameters. The relations have to be sufficiently strong.

The harmonization of the common dataset may include the following aspects:

- Taxonomical adjustment and coding: The taxonomical data need to be standardized (according to recent reference literature) and coded (using international species codes).
- Level of taxonomic identification: Generally, data need to show the lowest taxonomical level that is required for applying all national assessment methods

appropriately. Higher levels can be accepted if the effect on national methods is minimal (refer to relation between adjusted and original method).

- Record of abundance: The selection of the most precise scheme is desirable (e.g. Individuals per square metre). However, it is often necessary to agree on a “least common denominator” that can be provided by all Member States. A transformation scheme to convert national abundance data might be suitable.
- Record of different aspects of the BQE: If national assessment methods record different/additional aspects of the BQE, but focus on similar pressures (e.g. macroalgae within macrophyte assessment), the common dataset may not provide the complete data relevant to all methods. In these cases, a “least common denominator” solution has to be found, e.g. applying the common dataset to an adjusted national method and referring to relation between adjusted and original method.

II.4 Collation of pressure data

Alongside the biological information pressure data of the sampling site need to be collated. Ideally, these data have to be standardized (comparable) measurements of anthropogenic influence to which the BQE is responding. Suitable parameters are, for instance, anthropogenic land use in the surrounding of the sampling site, derived from Corine Land Cover data. Chemical data sampled according to similar protocols can also be used, but should include the specification of data aggregation (e.g. monthly mean, spot) and spatial/temporal match with biological data. Data on habitat quality evaluation is useful, although the Member States often use individual systems. These data thus require harmonization efforts before including it in the common database.

II.5 National data quality criteria

Most national assessment methods employ quality criteria to evaluate the acceptability of a sample for ecological status classification. For example, the sample has to contain a minimum number of indicator taxa, or the total abundance has to exceed a specific value. When collating the common dataset, these national data quality criteria need to be considered. If compliance is difficult due to specific national requirements (e.g. short regionally specific indicator lists), modified criteria can be used.

II.6 Biogeographical differences

If the common dataset covers very large geographical gradients, the data may be prone to biogeographical differences. For example, climatic factors cause distinct community compositions within the same common intercalibration type. Or by natural means, a certain species is rare in country A (limit of distribution) and ubiquitous in country B (centre of distribution). The effect of biogeographical differences on the national classifications needs to be evaluated. Significant impacts on the intercalibration results have to be taken into account, e.g. by including sub-types (individual reference conditions) between which national classifications are compared.

ANNEX III: Guidance for deriving reference conditions and defining alternative benchmarks for intercalibration

1. Common benchmarking for intercalibration

1.1. The setting of a common benchmark is a crucial step of intercalibration, as it establishes the harmonized basis for comparing the national class boundaries. The same class boundary values may reflect different levels of human impact if the reference states are defined differently. In the intercalibration exercise the common intercalibration benchmark is set based either on reference sites or on alternative approaches; a tiered approach is recommended:

- Tier 1 - “true” reference sites – sites with no or minimal anthropogenic pressure that fulfill all criteria proposed in REFCOND Guidance for all pressures (so for all the BQEs);
- Tier 2 - “partial” reference sites – subject to greater anthropogenic disturbance but certain biological quality element parameters do not differ from true reference biological conditions (e.g. “phytoplankton reference sites” with no or minimal eutrophication pressure but significant morphological pressure not affecting the phytoplankton community in a significant manner);
- Tier 3 - “alternative benchmark” sites – sites impacted by similar level of disturbance and exerting similar level of impairment to biology (to be used for setting biological benchmark for intercalibration exercise). This approach allows for intercalibration even if reference sites are absent.

1.2. Because the intercalibration results will influence water management decision across Europe, the process must be transparent and verifiable. Harmonization based on reference sites is difficult to verify if these reference sites are identified by the member states themselves. Therefore the benchmarking process must use harmonised criteria independent of national classifications (i.e. countries cannot simply nominate the sites they classify as high status as being their benchmark sites without further checking).

1.3. Harmonised criteria to define these reference conditions or the alternative benchmark for the intercalibration exercise have to be established. These criteria are intended to allow for screening of reference sites or alternative benchmark sites which can be done in two ways:

- In case of separate datasets each country must nominate a set of national reference sites or alternative benchmark sites belonging to the relevant IC type that have been screened against agreed abiotic criteria. If a country employed a geographical analogue approach in establishing reference sites or alternative benchmark sites, and therefore used unimpacted sites (or with a specific impact in the case of alternative benchmark sites) from a different country as the basis for its method, these sites should be submitted for benchmarking.

- In case of a common dataset (preferable approach), harmonised set of reference criteria are applied to abiotic data in the intercalibration dataset to select reference sites or alternative benchmark sites.

1.4. To come to a common understanding for reference conditions or an alternative benchmark in the same type, similar methodologies should be adopted for the characterization of very low pressure levels of reference conditions for all water categories or similar pressure levels in the case of an alternative benchmark. Table 1 illustrates a common approach to allow consistency on pressures identification across water categories. It shows a list of the most important pressures for each water category, together with examples of potential pressure indicators of relevance to the biota that should be analyzed. GIGs should make an appropriate survey and assessment of the driving forces and pressures, at the relevant spatial scales (watershed, water body, site), and agree which pressures are relevant for the BQEs.

Rivers		Lakes		Transitional		Coastal	
Pressure type	Pressure indicators	Pressure type	Pressure indicators	Pressure type	Pressure indicators	Pressure type	Pressure indicators
1. Point source Pollution	Population density, oxygen, phosphate, nitrogen	1. Point source Pollution	Population density, total phosphorus	1. Point source pollution	Population density, oxygen, phosphate, nitrogen	1. Point source pollution (from rivers+coastline)	Population density, oxygen, phosphate, nitrogen
2. Diffuse source Pollution	Agriculture land use, phosphate, nitrogen	2. Diffuse source Pollution	Agriculture land use, total phosphorus	2. Diffuse source Pollution	Agriculture land use, phosphate, nitrogen	2. Diffuse source Pollution	Agriculture land use, phosphate, nitrogen
3. Riparian zone vegetation	Riparian use, riparian composition, riparian longitudinal ad lateral connectivity	3. Riparian zone vegetation	Riparian use, riparian composition, riparian longitudinal ad lateral connectivity	3. Riparian zone vegetation	Riparian use, riparian composition, riparian longitudinal ad lateral connectivity	3. Shoreline modifications/harbours in supralittoral/ terrestrial	Shoreline occupation, continuity between coastal perimeter and natural settings
4. Morphological alterations	Sediment transport, river continuity, channelisation, siltation, river profile, presence of weirs and dams	4. Morphological alterations	Quantity and dynamics od flow, water level, residence time, groundwater connection, depth variation, substrate and structure of shore zone	4. Hydromorpho-logical alterations	Quantity and dynamics od flow, water level, residence time, groundwater connection, depth variation, substrate and structure of shore zone	4. Hydromorpho-logical alterations in littoral and sublittoral/	Changes in deposition/erosional areas, groyns
5. Water abstraction	Abstraction below a threshold	5. Water abstraction	Abstraction below a threshold				
6. River flow regulation	Presence of dams influencing natural flow regime, storage and seasonal patterns						
7. Biological pressures	presence of invasive species, biomanipulation, intensive fishery/aquaculture	7. Biological pressures	presence of invasive species, biomanipulation, intensive fishery/aquaculture	7. Biological pressures	presence of invasive species, biomanipulation, intensive fishery/aquaculture	7. Biological pressures	presence of invasive species, biomanipulation, intensive fishery/aquaculture
8. Other pressures	intensity recreational use	8. Other pressures	intensity recreational use	8. Other pressures	intensity recreational use	8. Other pressures	intensity recreational use

Table 1. List of important REFCOND pressures and potential pressure indicators for each type of pressure per water category.

2. Guidance for selecting 'true' or 'partial' reference sites

2.1. Guidance for selection of reference sites is given in previous CIS guidance documents[12] and in the Guidelines[13]. The approaches described in these documents should be followed as much as possible when selecting reference sites in the intercalibration exercise.

2.2. The Directive provides a number of options for establishing type-specific reference conditions[14]. Reference conditions may be either spatially based or based on modelling, or may be derived using a combination of these methods. Where it is not possible to use these methods, expert judgement may be used to establish such conditions. [Guidelines, section 3.3]

2.3. The use of spatial networks of reference sites is expected to provide the most reliable estimates of biological reference conditions and is therefore the preferred option, where practicable [Guidelines section 3.5]

2.4. Reference conditions do not equate necessarily to totally undisturbed, pristine conditions. They may include very minor disturbance which means that human pressure is allowed as long as there are no or only very minor ecological effects. Therefore, sites subject to a greater anthropogenic disturbance can be used as reference sites provided the relevant biological quality element parameters do not differ from true reference biological conditions. For that reason, the criteria can be modified concerning their relevance for the specific BQE. For example, the criterion of morphological changes could be omitted for setting reference condition for phytoplankton because that pressure does not significantly affect the BQE. Such sites can be used establishing reference conditions for a specific BQE even if they are not 'true' reference sites.

2.5. The level of "very low pressure" corresponding to "very minor modifications" of the biological quality element should be defined, when sufficient data are available, on the basis of statistical relationships demonstrating that the level of pressure accepted to select a reference site is unlikely to have a significant impact on the biological quality element (or parameter) [Guidelines, section 3.7].

2.6. In order to avoid any circular reasoning, biological data should not be taken into account in a first stage. Sites with statistically outlier biological values should be carefully checked for pressures, and dubious sites eliminated. On the other hand, the outlier values that can be explained by natural disturbances (e.g. variability of meteorological and hydrological conditions) that affect temporarily the biological communities can be considered as part of the natural variability of the site and should not be eliminated [Guidelines, section 3.9].

[12] CIS Guidance Document No 10 (2003): Rivers and Lakes – Typology, Reference Conditions and Classification Systems; Guidance Document No 5 (2003): Transitional and Coastal Waters – Typology, Reference Conditions and Classification Systems

[13] Guidelines to translate the intercalibration results into national classification systems and to derive reference conditions

[14] Paragraph 1.3, sub-paragraph (iii), Annex II

2.7. Expert judgement should be part of the benchmarking process. It should be used to :
- consolidate the definition of harmonized criteria for reference conditions and the statistically derived threshold on pressures ;
- drive the implementation of these criteria and thresholds (e.g. selection of reference sites, alternative benchmarking...).

2.8. Setting reference conditions for reservoirs or water body types likely to be designated as heavily modified (HMWB) can be done through the identification of another similar water body, within the same type, which is subject to insignificant human pressures except for those hydro-morphological modifications accepted its designation as HMWB [Guidelines, section 3.10].

2.9. The steps to derive reference conditions (and status class boundaries) when a spatial network of reference sites is available are illustrated in figure 1. Initially reference criteria and thresholds on pressures that have been agreed should be applied to select the spatial network of reference sites (1). Secondly, the biological data from reference sites should be analysed to derive the biological reference benchmark (2). In posterior steps, the other class boundaries should be identified in agreement with the normative definitions and the deviation from reference conditions (points 3, 4, 5 and 6).

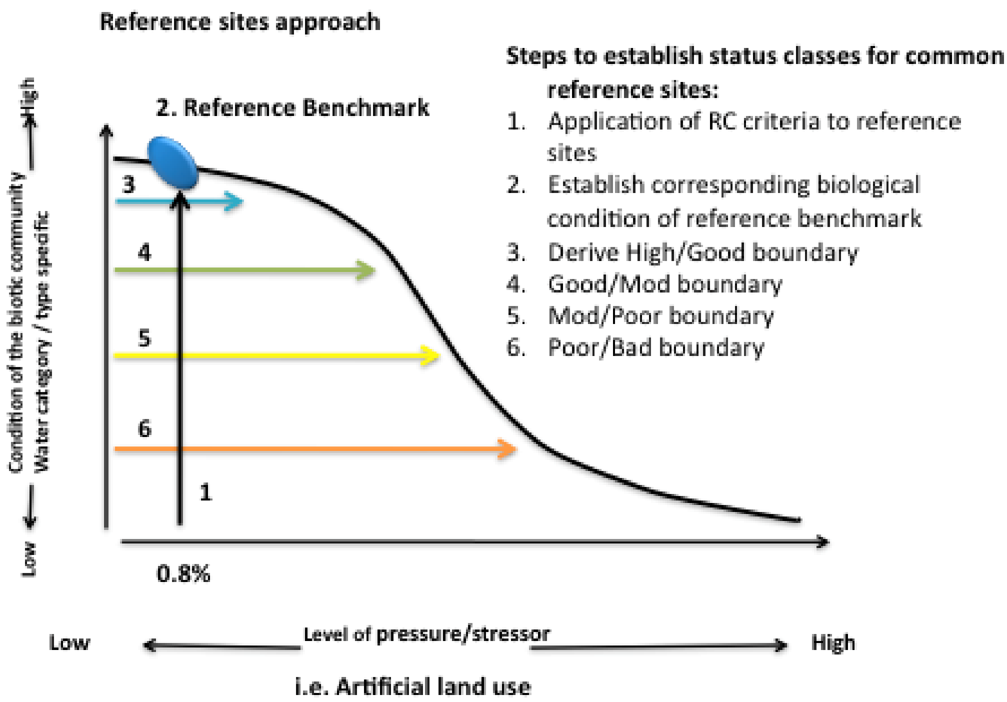


Figure 1. Derivation of the reference benchmark and status class boundaries when a spatial network of reference sites is available.

2.10. It should be ensured that, when collecting any new information on pressures and biological conditions, this is done using standardised methods, or methods currently in use according to the scientific literature, or any new method properly described and tested providing reliable information, and quality assurance procedures where applicable [Guidelines, section 3.11].

2.11. Where existing data is used to derive reference conditions, it should be ensured that data is sufficiently comparable. Where necessary, appropriate conversion factors may be applied to improve the comparability of data [Guidelines, section 3.12].

2.12. The spatial network must consist of sufficient sites to enable to:

i. Confidently estimate the reference value (i.e. statistic) that will serve as the reference biological value for the classification system for the biological quality element; and

ii. Determine whether or not the natural variation in the biological element is too great to establish reliable type-specific reference conditions [Guidelines, section 3.13].

2.13. Where a Member State has insufficient reference sites within its territory to enable it to derive a reliable estimate of biological reference conditions, it should explore the potential for utilising information from suitable sites in the territory of other Member States. In doing this, Member States should:

- consider the comparability of the conditions at those sites with those at relevant sites in its territory (e.g. climatic, geomorphologic, physiographic conditions);

- ensure the effects of differences in these conditions can be estimated and appropriately taken into account when making use of information from the sites to derive biological reference values [Guidelines, section 3.14].

2.14. It may not be possible to establish reliable type-specific reference conditions if the natural spatial variation in the biological element across the type is too large. Where this may be the case, it should be assessed whether reliable reference conditions could be established by using additional factors to identify types representing a narrower range of spatial variation in the biological element concerned. Before doing this, it should be considered whether there would be sufficient numbers of relevant reference sites from which to derive reliable reference conditions for the new types [Guidelines, section 3.15].

2.15. It may not be possible to establish reliable type-specific reference conditions if the natural temporal variation in the biological quality element (or metric) is too large. Where this may be the case, it should be assessed whether reliable reference conditions could be established by using reference data obtained from particular seasons [Guidelines, section 3.16].

Modelling approaches

2.16. Member States may elect to use data from reference sites in combination with modelling approaches to predict the most appropriate biological reference value for individual water bodies or groups of water bodies in order to reduce the effect of natural spatial variation on the reliability of reference conditions [Guidelines, section 3.18]

2.17. Modelling approaches may be used on their own or to improve confidence in the estimates of reference conditions based on a spatial network of reference sites. [Guidelines, section 3.19].

2.18. Models should be designed to estimate the biological reference values expected under the conditions affected by no or very low human pressure [Guidelines, section 3.20].

2.19. When using modelling approaches, it should be ensured that the models provide a sufficient level of confidence about the values for the reference conditions and that the conditions so derived are consistent and valid for each surface water body type. To ensure a sufficient level of confidence, Member States should compare the model predictions with data from known reference sites, historical data or palaeological data; and/or undertake appropriate sensitivity analyses [Guidelines, section 3.21].

2.20. The steps to derive status class boundaries when a model is used are illustrated in figure 2. First, starting from a relationship between biological data and abiotic variables/pressure indicators, derived from available data at a specific point or for a trend along the pressure gradient, a pristine abiotic environment is modelled. These predicted abiotic reference conditions should comply with the set of reference criteria agreed at national level for reference sites (2). Secondly, the predicted biological data are used to describe the biological reference benchmark (3). Subsequently, the class boundaries should be derived in agreement with normative definitions and deviation from reference condition, as for figure 1 (points 4, 5, 6 and 7). All biological predictions resulting from models should be referred to a deviation from the very low pressure levels of reference conditions (5 & 7).

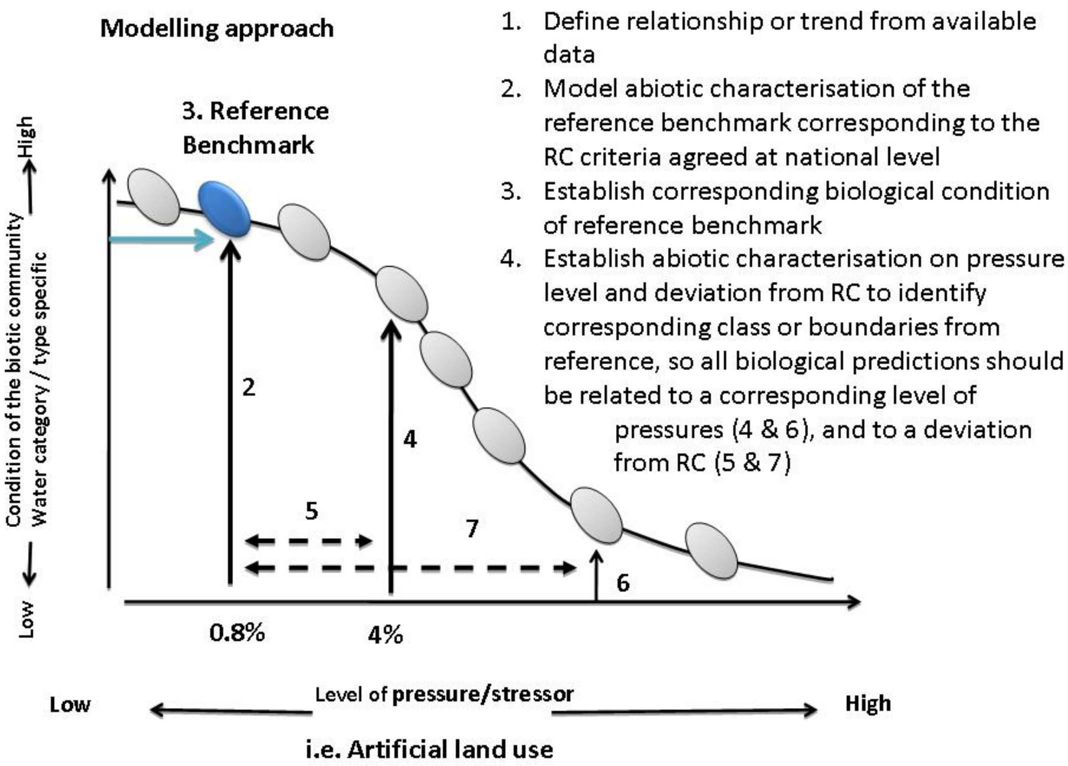


Figure 2. Modelling and reference conditions

Expert judgement

2.21. Member States could base reference conditions on expert judgement where it is not possible to derive reference conditions based on a spatial network of reference sites or from modelling [Guidelines, section 3.22].

2.22. Expert judgement could also be a part of the process of selecting reference sites, when background data or scientific knowledge are not available, to assess the level of pressure corresponding to "very minor modifications of physico-chemistry, hydro-morphology and biology" [Guidelines, Section 3.23].

2.23. In making expert judgements, Member States should use as many sources of information as possible, including monitoring data and relevant information (e.g. historical or palaeological data, background levels identified by the international conventions), to improve confidence in their understanding of how the biological quality element responds to increased pressure and hence the values for that element under conditions of no or only very minor human disturbance [Guidelines, Section 3.24].

3. Alternative benchmarking

3.1. If no or only very few sites meet the reference criteria, alternative benchmarking approach (Figure 3) must be used based on sites impacted by similar levels of disturbance (alternative benchmark sites).

3.2. When using alternative benchmarking sites, the following preconditions need to be fulfilled:

a. the pressure-impact relationship must be the same across the data set used
b. there is a need to account for all relevant pressures
c. if there are multiple pressures, they must be combined in a meaningful way

3.3. Alternative benchmark sites have to be identified from the common intercalibration dataset. This can be done by screening for sites meeting abiotic criteria agreed at GIG level (step 1 in Fig 3) which (1) represent a similar level of human pressure; (2) represent best available (or least disturbed) physical chemical and biological conditions given today's state of landscape. These criteria will vary from region to region, and are developed iteratively with the goal of establishing the least amount of ambient human disturbance in the region under study. An example of the criteria approach is given in Birk & Hering (2009) for the Danube basin countries. To locate the least disturbed sites for diatom and invertebrates intercalibration, a series of criteria were developed describing land use, hydromorphological parameters and chemical criteria (nutrients, biological oxygen demand an conductivity).

3.4. The biological conditions of these sites need to be reviewed to avoid the influence of impacts caused by pressures not regarded in the screening process.

3.5. The biological parameters of these sites are used to establish biological benchmarks (Step 2 in Figure 3) for intercalibration, i.e. the condition of the biological community that represents the trans-national reference point for harmonization (Birk & Hering 2009). Expert discussions should confirm a common notion of type-specific communities at benchmark status.

3.6. A reasonable definition of biological benchmarks requires distinct pressure-impact relationships and high quality pressure data (Figure 3). There are possibly few countries in Europe with the perfect understanding of pressure-biota relationships and sufficiently comprehensive pressure databases for all biology sites. Therefore, the process depends on the accumulation of experience and integration from all lines of evidence rather than statistically rigorous procedures:

- it may be plausible for countries to screen for sites where supporting environmental data are existing, then to check biology of these sites and then add in other sites with matching biology but where less extensive environmental data is available;

- The biology generated by screening has to be considered in the light of the normative definitions and a common understanding of ecological changes;

- A common opinion has to be reached on type-specific communities at benchmark status and their deviation level from "true" reference conditions;

- The final aim has to be the definition of biological conditions ("benchmark communities") that represent common level of biological deviation from natural reference communities.

3.7. It is important to identify the position of the benchmark on the gradient of impact, i.e. to document the deviation of the selected benchmark from reference conditions. Therefore at first “virtual reference” (not existing in reality) has to be derived (Step 3 and 4 in Figure 3). It can be done by several approaches:

- Using the very few “true” reference sites still existing, literature data and expert judgement;

- Defining “virtual” reference sites not existing in reality but conceived as the potential biological components that should be present (Borja *et al*. 2004);

- Deriving biological reference conditions from extrapolating the dose-response relationship (e.g. combined pressure gradient versus common metric);

- Predicting the biological reference values of the common metric in a multiple regression analysis using the individual pressures as independent variables.

- The actual distance of the alternative benchmark sites from the virtual references (Step 5 in Figure 3) allows evaluating the quality status of the available sampling stations in terms of their level of pressure.

3.8. Modelling approaches (as specified in 2.16-2.20) and expert judgment (as specified in 2.21-2.23) can also be used to support the identification of alternative benchmarks

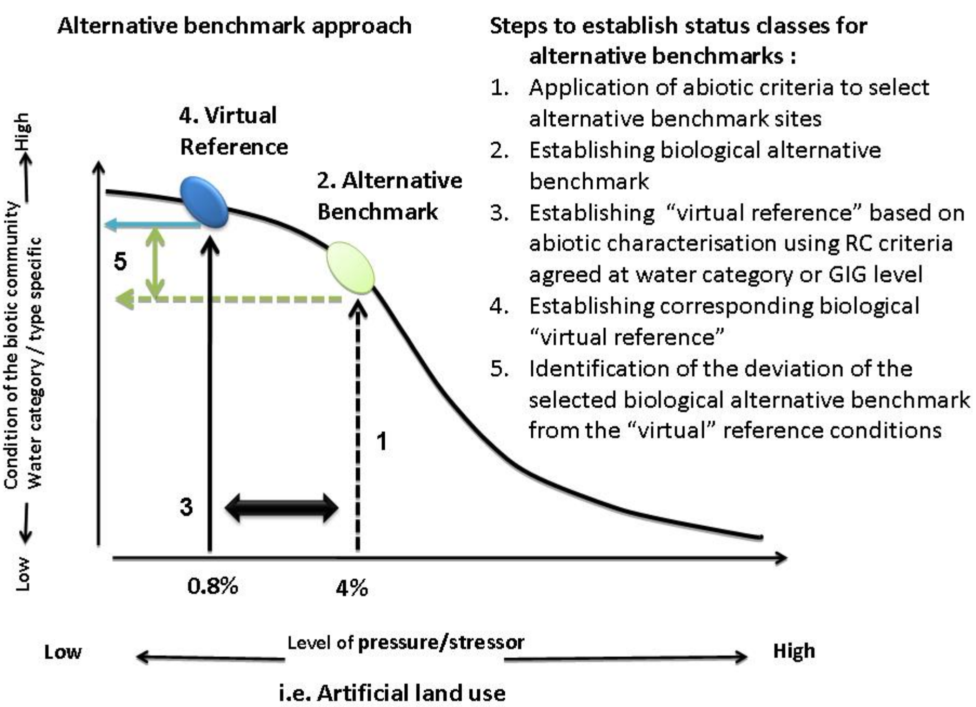


Figure 3. Alternative benchmark and reference conditions.

References

Birk S, Hering D (2009) A new procedure for comparing class boundaries of biological assessment methods: A case study from the Danube Basin. Ecological Indicators 9: 538-539

Borja A, France J, Valencia V, Bald J, Muxica I, Belzunce MJ, Solaun O (2004) Implementation of the European water framework directive from the Basque country (northern Spain): a methodological approach. Marine Pollution Bulletin 48:209-218

ANNEX IV: The development of a boundary setting protocol for the purposes of the intercalibration exercise

Background

This technical paper represents a template boundary setting protocol for the purposes of the intercalibration exercise required by section 1.4.1 of Annex V to Directive 2000/60/EC. The need for this template was identified in the Intercalibration Process Guidance, adopted by Water Directors in December 2004.

The protocol deals with the setting of specific class boundaries for those metrics of the biological quality elements for which suitable assessment methods and data are available for the intercalibration exercise. It does not deal with the overall classification of the ecological status of water bodies.

The template boundary setting protocol should be completed by the GIG (or alternative intercalibration group) and for each biological quality element being intercalibrated. The protocol should be applied in accordance with the agreed approaches to intercalibration. Where the Member States are comparing their own monitoring systems between themselves (Option 2 or Option 3), the boundary setting protocol will be applied by the individual Member States and the GIG will check that the BSP has been applied consistently and then oversee the comparison of boundaries and any harmonisation of boundaries.

The protocol presumes that the GIGs have identified types or sub-types the biology of which is expected to show a broadly similar ecological response to anthropogenic disturbances. For example, the ecology of naturally oligotrophic lakes may show a significantly different characteristic response to nutrient enrichment compared to naturally eutrophic lakes.

The WG ECOSTAT Discussion Paper “Draft Principles of Ecological Status Classification in Relation to Eutrophication”, sets out a proposed common understanding of the Water Framework Directive’s normative definitions in the context of nutrient enrichment, focusing on those key principles of the normative definitions that are relevant across the water categories. This can be used as a framework to apply the class boundary setting protocol regarding eutrophication, and in particular developing the conceptual description of the effects on the biological quality element of increasing impact on the supporting elements (see Step 2).

Step 1: Identify qualifying criteria for type-specific reference conditions (more detailed description given in Annex III)

- Describe the criteria used to identify reference sites for the biological quality element:
    - Identify the specific values for the relevant pressure criteria, hydromorphological and physico-chemical conditions considered to correspond to no, or only very minor, anthropogenic alteration

- State whether it was possible to identify reference values for the biological quality element using data from reference sites:
    - Were sufficient reference sites available for the type?
    - Were there sufficient biological data available from reference sites?

- If it was possible to use reference sites:
    - Specify which summary statistic (e.g. median value or arithmetic mean) of the values for the biological quality elements at reference conditions were used to quantify reference conditions for the purpose of calculating EQRs
    - Specify which summary statistic (e.g. 95 percentile) of the values for the biological quality elements at reference were used to identify the high-good boundary

- If it was not possible to use reference sites:
    - Specify the relevant criteria used to define reference values and the high-good boundary (e.g. when using modelling methods; paleolimnological methods; expert judgement; etc.)

Step 2:
(a) Describe how the biological quality element is expected to change as the impact of the pressure or pressures on supporting elements increases[15]; and
(b) Relate this description to the normative definitions.

- Specify the relevant pressure or combination of pressures and the associated impacts on the supporting elements that are being considered.

[15] The direct effects of most pressures are on the supporting elements (i.e. physico-chemical conditions and hydromorphological conditions). The changes in these supporting elements lead to impacts on biological quality elements. Relatively few pressures act directly on the biological quality elements (e.g. fishing). If relevant, the effects of such pressures should be taken into account when using the protocol

- Specify the quality element(s) being considered.

- In the form of a conceptual model, describe how the biological quality element(s) is expected to respond as the impact (or impacts) on the supporting elements increases. The conceptual model should be designed to highlight key changes to ecosystem structure and function as anthropogenic disturbance increases.

- Based on the normative definitions and the conceptual model, provide an ecological description of the condition of the biological quality element at high, good and moderate status.

Step 3: Select suitable metric(s) of the quality element; assess whether the metric(s) responds to the gradient of impact contained in the data set; and quantify the reference conditions for the metric.

This purpose of this step is to organise the data in the biological data set so that they describe the way in which the biological quality element responds to increasing impacts (i.e. they describe the degradation curve for the biological quality element).

- Select a metric (or metrics) of the quality element that is representative of the effects on the quality element predicted in the Step 2 analysis of the normative definitions (for example, relative biovolume of Cyanobacteria describes effect of eutrophication on Phytoplankton BQE).

- Identify a descriptor, or composite descriptor, of the degree of the relevant pressure or combination of pressures (for example, total phosphorus or chlorophyll-a concentration describes eutrophication pressure).

- Identify whether the biological metric being considered responds over the whole potential gradient of impact on the supporting element(s). If not, try to find a combination of metrics for the quality element that will together cover the whole spectrum[16].

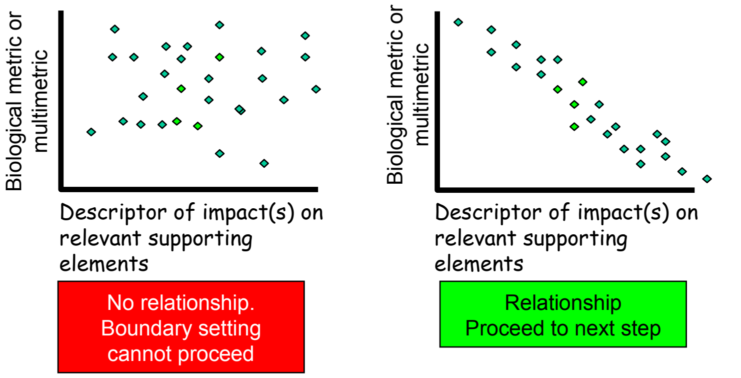
- Collate comparable data on the selected biological metric or metrics from a range of sites subject to varying degrees of anthropogenic impact, including reference sites if possible.

- If the metric shows relationships with the impact gradient:
  (i) Quantify the reference conditions and the high-good boundary following the procedure outlined in step 1;

[16] If it is not possible to calculate metrics responding over the whole spectrum of the impact gradient, ensure a metric is selected that shows a response likely to span at least high, good and moderate status

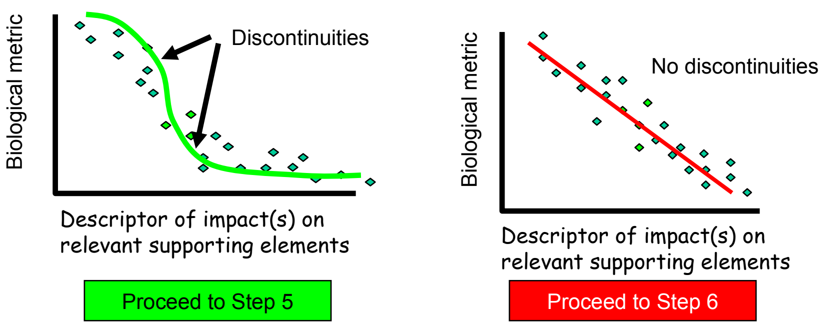
(ii) Continue with step 4.

- If the metric shows no relationship with the impact gradient represented in the dataset, the boundary setting process for this metric cannot proceed. In such cases:
  (i) The use of another metric of the quality element should be considered;
  (ii) The collection of better data on the original metric of the quality element should be considered; and
  (iii) The appropriateness of the way in which the impact gradient has been defined should be considered (e.g. Are other pressures acting? Is the definition of the impact gradient sufficiently type-specific?)



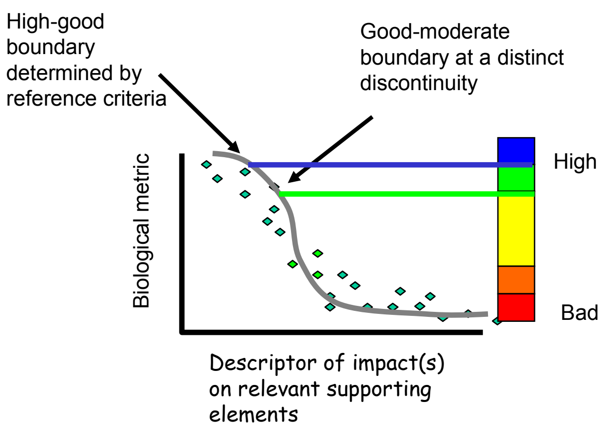
Step 4: Identify if there are any discontinuities in the relationship between the metric and the gradient of impact represented by the data set.

- If there are distinct discontinuities in the relationship between the biological metric and the gradient of impact represented in the data set, specify how the values for the discontinuity are derived from the data and proceed to Step 5. If not, proceed to Step 6



Step 5: Determine if the discontinuity relates to a class boundary or a class centre

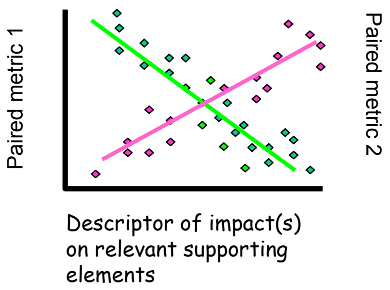
- Compare the data at the discontinuities with the Step 2 analysis of the normative definitions.
- Decide if the discontinuities correspond to class centres or class boundaries and identify to which classes they relate (for example, steep decrease of macrophyte abundance corresponds to the Good/Moderate class boundary).
- Set out the reasons for the decision and set class boundaries accordingly.
- Specify how errors in the estimate of the class boundaries or class centres are taken into account in setting class boundaries.



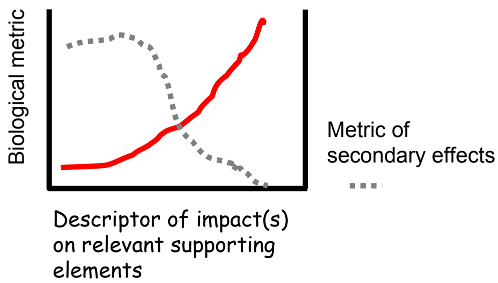
Step 6: Taking account of the results of Step 2, assess whether class centres or class boundaries can be located using paired metrics.

- Select appropriate paired metrics based on the Step 2 analysis of the normative definitions.

> **Example 1: Step 2 analysis predicts that paired metrics of the quality element respond in different ways to the influence of the pressure** (e.g. % sensitive taxa compared to % of impact taxa for benthic invertebrates in rivers and lakes)



Example 2: Step 2 analysis predicts secondary effects as the metric of the quality element becomes increasingly impacted (e.g. increase in phytoplankton biomass leading to secondary effects on macrophytes – normative definitions for phytoplankton in lakes)

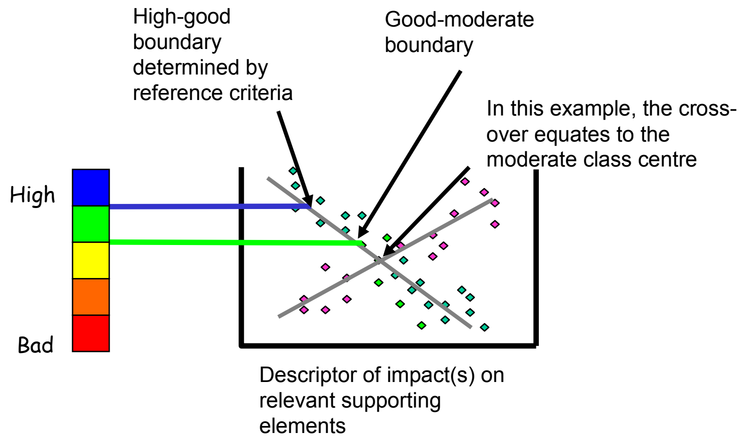


- Assess the relationship between the paired metrics across the gradient of impact represented by the data set.
- If there is an ecologically relevant interaction between paired metrics[17], proceed to Step 7.
- If no relationships are found between any paired metrics, try to obtain better data on the metrics. If this does not improve the situation, proceed to Step 8.

Step 7: Determine whether values derived from the paired metric analysis correspond to class centres or class boundaries.

- Take account of the Step 2 analysis of the normative definitions to decide if the values derived from the paired metric assessments correspond to a class centre or a class boundary, and to which classes they relate (for example, the Good/Moderate boundary for several types in Northern GIG where set at the cross-over point of % phytoplankton sensitive taxa compared to % of phytoplankton impact taxa).

[17] e.g. a cross-over point (example 1) or step changes occurring in a secondary effect at distinct values of the biological element (example 2)



- Specify how the values derived from the paired metric assessments are used to determine the good-moderate class boundary.

- Specify how the error associated with the estimates from the paired metric assessments are taken into account in setting the boundary.

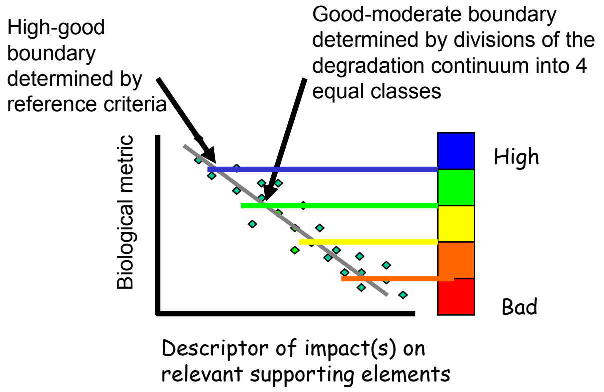
Step 8: Setting class boundaries if the relationship between the quality element and the pressure gradient is a continuum and Step 6 has failed to identify boundaries based on paired metric assessments.

- How should boundaries be identified in this situation?

  Example approach
- As a starting point, divide the continuum of impact below the high-good boundary (established in Step 1) into four equal width classes. If the data set does not cover the full spectrum of impact, divide the data set below the high-good boundary into an appropriate number of equal width classes

- Examine the values of the metric of the quality element represented at the good and moderate status class boundaries and compare the ecological meaning of these values with the Step 2 analysis of the normative definitions (e.g. no major reference taxonomic groups of benthic invertebrates should be absent at good status – normative definitions for rivers and lakes)

- Revise the boundaries until the values represented in the good and moderate status classes are consistent with the descriptions provided by the Step 2 analysis of the normative definitions


High-good boundary determined by reference criteria
Good-moderate boundary determined by divisions of the degradation continuum into 4 equal classes
Biological metric
High
Bad
Descriptor of impact(s) on relevant supporting elements

ANNEX V: Definition of comparability criteria for setting class boundaries

1. Introduction

The Water Framework Directive requires the intercalibration of the results of the Member States' ecological assessment methods. Section 1.4.1 on the comparability of biological monitoring results in the WFD Annex V explains that the Commission shall facilitate this intercalibration exercise in order to ensure that these class boundaries are established consistent with the normative definitions [...] and are comparable between Member States.
This Intercalibration Guidance annex aims to explain better what is meant by comparability between Member States and provides a consistent approach and an improved version of the comparisons that were used in the first phase of intercalibration. Scientific implications of the question on comparability were initiated and developed in the first phase, but required a consistent and more detailed approach to increase transparency and quality of the comparison exercises. The need for an improved and harmonized concept at the beginning of the analytical process of the second intercalibration phase was expressed clearly by the Water Directors and the European Commission.

Why is improvement needed?

In 2007 the European Commission addressed questions on the degree of comparability between specific results of the intercalibration exercise to Working Group A ECOSTAT and the Water Directors and some results were excluded from the Commission Decision on the intercalibration results. The level of comparability seemed to be too variable between GIGs or could not be properly assessed, since they used different criteria to evaluate comparability. Therefore it was concluded that there should be common criteria to assess the level of agreement between the results of the different methods of the Member States over all the GIGs. There was general agreement that further work was required on developing comparability criteria.

As part of the first phase of the intercalibration process (2004-2007) three Options were developed to assist in the comparison of national classifications. The first option considered comparability between Member States when they were all using the same assessment method (but different class boundaries). When Member States were using different methods, two different ways were used to assess comparability. These two intercalibration options were Option 2, where the monitoring data were so different that not all national methods could be applied to all the sites in a common database, and Option 3 where all the national methods could be applied to calculate ecological status for each site in a common database.
For Option 2 the national methods were related to a common metric to provide a common yardstick with which to compare the different national methods. For intercalibration this indirect approach focused on the comparison of the placement of the boundaries between the ecological status classes. The common view on the position of the G/M and H/G boundary was determined by plotting all the national classifications against the common yardstick, using a regression approach. This indirect approach of analysis reflected only the bias in classifications through boundary setting, not the level of class agreement. Thus bias may have been low (boundaries being quite similar) even though class agreement was moderate or poor.

For Option 3 another approach was used involving direct comparison of the results of the different national assessment methods on the same sites. Thus the level of class agreement could be checked directly and reflected how well the EQRs of the national methods were related. The relationship and deviation between the positioning of the boundaries by different Member States was not investigated.
However, both these aspects need to be considered when comparing classifications.
The difficulty in the use of the two different approaches (Options 2 and 3) in the first phase was that it was not clear if these two approaches are equally stringent in judging comparability. In fact, these two approaches focus on two different aspects of comparability (class boundary positioning and variability in EQR values). In the review on the methodology it was concluded that these aspects should not be used separately from each other but rather that they should be used in parallel for all the intercalibration options. Separately, these two approaches seem not to adequately cover the two most important aspects of comparability.

The concept that has been developed now

- is applicable to all BQEs of all water categories. So it guarantees general applicability, comparability and transparency of the approach to assess comparability.
- It is in line with the requirements of the Intercalibration Guidance (phase 2: 2008-2011).
- It considers the two main approaches that are applied in the first phase of intercalibration to assess comparability between different assessment methods. The current approach is a mutual extension and harmonization of the two approaches merging them with each other and making it possible to assess comparability in an equally stringent way.

2. General principles of comparability analysis

- The proposed approach considers only **upper class boundaries** (High/Good and Good/Moderate) – classifications are aggregated in classes above or below the respective boundary.
- Comparability is **always** checked through the analysis of **two components**: **boundary bias** and **class agreement**, using EQRs. Sufficient comparability is reached when acceptability criteria on boundary bias are met and class agreement has been checked, as further explained below.
- Boundary bias can be reduced in the intercalibration exercise by adjusting or harmonizing boundaries. Boundary harmonization represents a state of agreement of the upper boundaries (i.e. High/Good and Good/Moderate) between Member States when the boundary bias criteria are met, i.e. the different national boundaries should not differ more than 0.5 class from each other (= the maximum boundary deviation above or below each national boundary is a quarter of a class). At this point the boundaries are considered to be harmonized and the degree of bias amongst classifications to be minimized with the result that all assessment methods display a similar level of ambition in defining good and high ecological status. All Member States can then be considered to have a similar interpretation of the thresholds of deviation from reference conditions that are contained in the normative definitions.
- Class agreement depends very much on how closely the methods are related. Therefore, the relation between different methods is checked first before analyzing the boundary bias. The class agreement is calculated after boundary harmonization in order to show the performance of the methods after required adjustments are made to the boundaries.

- Presentation of harmonized upper class boundaries must be supported by an **ecological characterization,** describing the biological communities representing the “borderline” conditions between good and moderate ecological status and between good and high ecological status (see point 2.6 and Figure 1 flow chart of the main steps of the intercalibration process in the Intercalibration Guidance).
- In case the assessment methods developed by a Member State differ so much that the data cannot be compared, the assessment method cannot be intercalibrated by one of the options provided in this guidance. The MS (in collaboration with the GIG) will need to find an alternate intercalibration approach (see key principle 9 in the Intercalibration Guidance). In case of not meeting other feasibility criteria such as data requirements for benchmarking or for statistical robustness of the comparison, some technical variations to the provided options can be proposed. This necessity and the preservation of conformity with respect to the content and the sense of the comparability criteria guidance have to be shown to ECOSTAT by the GIG or the MS. The alternate or adapted approach will need to be approved by WG ECOSTAT.

2.1. How is comparability explained in this Intercalibration Guidance annex V?

The comparability of classifications is evaluated by two main components in the intercalibration exercise:

- **Boundary bias** = the deviation in the relative positioning of class boundaries and measured by the magnitude and direction of deviation by a class boundary of one national method relative to the common view of the Member States (i.e. defined by the common metric or by the global mean of all the methods = pseudo-common metric, for the H/G and for the G/M class boundary). This deviation is expressed in class equivalents. It reflects the level of ambition of different methods or how stringent Member States are in defining the good ecological status.
- **Class agreement** = the confidence that two or more national methods will report the same class for a given site, as calculated by the average absolute class difference between all pairs of EQR values across all participating Member States, the proportion of classifications differing by an agreed amount (half a class), and the multi-rater kappa coefficient.

These two aspects are dependent on how well the methods are related. This relatedness of methods is reflected in the scatter in the relationship either between pairs of Member States’ methods or between Member State methods and a common metric. In Figure 1b for example the scatter in the relationship between MS B and the common metric is higher than in the relationship between MS A and the common metric in Figure 1a. The class agreement between MS B and the common metric will always be lower than for MS A and the common metric, even after adjusting the boundaries as much as possible. So, this scatter should be relatively low. Prior to comparing classifications it must be demonstrated that methods are significantly correlated with each other or with a common metric. <u>If this is not the case tests of comparability are not feasible</u>. This is also the reason why the IC feasibility check 1 and IC feasibility check 2 have been included in the main steps of the intercalibration process (see Figure 1 with the flow chart in the Intercalibration Guidance). This annex includes an explanation how a significant correlation between methods or between a method and a common metric has to be tested, to ensure that it is done in the same way for the different intercalibration options. The correlation between methods does not only reflect the relatedness between methods, but depends also on the uncertainty level associated with each national method. The uncertainty of national methods must be reported in the River Basin Management Plan, but it is not analyzed in the intercalibration exercise.

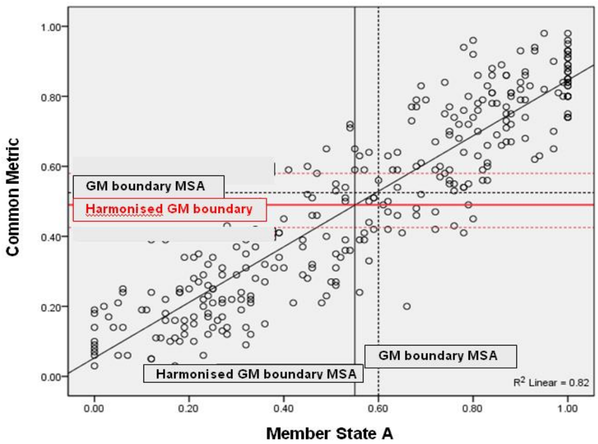


Figure 1a: Level of variability of Member State A relative to the common metric, with the indication of original and adjusted GM boundary

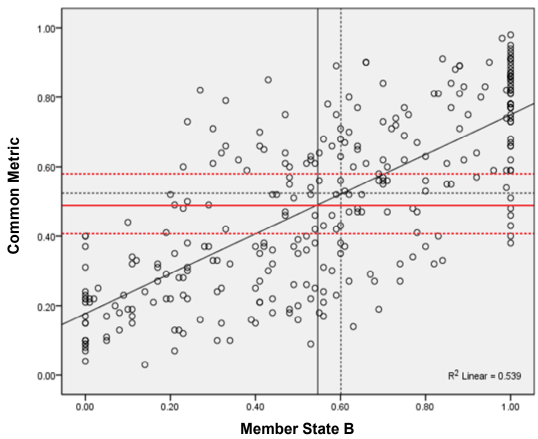


Figure 1b: Level of variability of Member State B relative to the common metric

Therefore, the key issue of how comparable Member States' methods are, can be split into three questions:

1. How closely are the methods related over the whole ecological quality gradient?
2. How comparable are the national definitions of the good ecological status, so how comparable are the boundaries H/G and G/M (= assessment of boundary bias)? This reflects the level of ambition or how stringent Member States are in defining the good ecological status.
3. Do the EQR results of the methods report the same class for the same site (assessment of class agreement)?

The first question is addressed by regression analysis, relating the methods to each other or to a common metric, and the subsequent evaluation of the regression characteristics (steps 3 and 4). In the case a regression is not meaningful, the correlation between methods can be measured using a non-parametric (Spearman's Rank) technique.

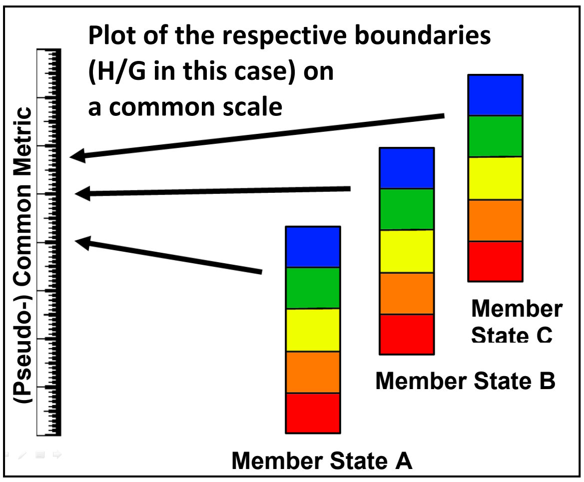


Figure 2: Analysis of boundary bias

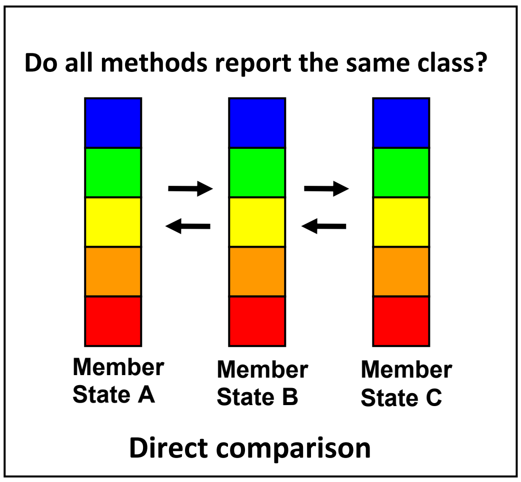


Figure 3: Analysis of class agreement

The second question on boundary bias is addressed by plotting the national boundaries on a common scale and deriving the permitted boundary deviation for each Member State (see Figure 2 and steps 5 to 7). This defines the amount of permitted variation in the positioning of the national boundaries. It is set as a quarter of a class, on the condition that all classes are transformed to have equal widths. In the first phase the value of 0.05 EQR permitted deviation was illustrated to be not equally stringent in all cases, when the class widths were not transformed. In the second phase calculation measures are proposed to ensure that the deviation is expressed in class widths and then constrained to a quarter of a class above and below the boundary. In other words, the different national boundaries should not differ more than 0.5 class from each other. When plotting of boundaries is not possible because a meaningful regression cannot be applied, an alternative in the second phase is to analyze the boundary bias through pairwise differences in EQR results as further explained below.

The third question on class agreement is closely related to the first question and is analyzed by means of direct comparison of the classification results (Figure 3) through an evaluation of (1) the average absolute class difference between all pairs of EQR values across all participating Member States, (2) the proportion of classifications differing by an agreed amount (i.e. half a class), and (3) the multi-rater kappa coefficient. The three class agreement metrics are reported to illustrate a consistent sufficient class agreement, while the evaluation of the acceptability of the class agreement result is based on only one metric, being the average absolute class difference.

2.2. How is comparability assessed in the different options of comparison ?

Figure 4 illustrates how to select the appropriate ways of comparison to compare national methods, corresponding to the options that have been defined in the Intercalibration Guidance.

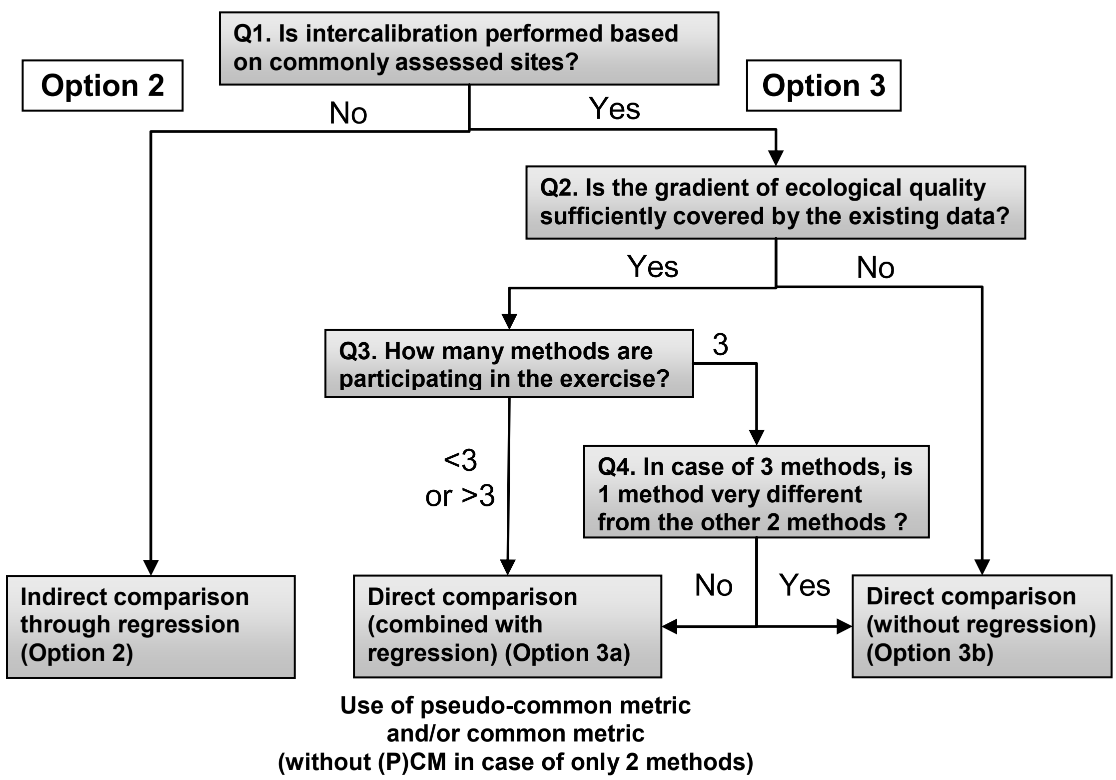


Figure 4: Decision tree as a support to the choice of the appropriate Option and comparison type

Option 2 indirect comparison: Indirect approach for boundary comparison via the regression of a **common metric** against each national EQR. This approach follows the original Phase 1 Option 2 but includes a refinement to convert results to class equivalents. In addition, a class agreement evaluation is included using modeled datasets (see further) to show the performance of methods in a similar way as in Option 3. This ensures that outputs can be inter-compared between all Options. Option 2 allows an indirect comparison of methods via an independent common metric that acts as a yardstick against which the position of national class boundaries can be assessed. This type of comparison is the only option that can be considered in the absence of a set of commonly assessed sites. Thus, if the sampling methods of Member State A and B exclude the application of the method of B to data of A and vice versa there will be no commonly assessed sites.

Whenever it is possible to assess all the sites with all the methods of the MSs, an Option 3 direct comparison should be performed.

Option 3a direct comparison: Direct approach for boundary comparison via the regression of a **pseudo-common metric**, formed from the average of an independent set of Member States, against the national EQR of the remaining Member State. Wherever available, an independent **common metric** can be used in this type of comparison as well, but the creation of an independent common metric is not necessary.
Direct comparison combined with a regression is ideal where there is a large common dataset, covering a wide gradient of ecological quality, where all sites are assessed by at least four different methods (Option 3a) and in 2 more specific cases.

- When there are 3 methods to be assessed, the use of the pseudo-common metric can only be useful when all 3 methods are well correlated. If there is one method deviating considerably from the other 2 methods, it is difficult to tell whether the 2 similar methods are more right than the third deviating one. Therefore, it is recommended to follow Option 3b without regression in this particular case.
- When there are only 2 methods to compare, a regression is possible of one Member State against the other Member State. In this case the construction of a pseudo-common metric reflecting the view of all the other MSs is not possible. The pseudo-common metric just consists of the view of the other single MS.

Whenever it is possible to construct a regression of one national method against the view of the other MSs, a regression has to be provided to illustrate how closely the methods are related. In these cases an Option 3a direct comparison is recommended for further boundary bias and class agreement analysis. The use of Option 3a is not possible in the 2 cases mentioned below under Option 3b.

Option 3b direct comparison without regression: Direct approach for boundary comparison via **multiple pairwise comparisons** of EQRs across a population of commonly assessed sites. The main constraint on the use of direct comparison combined with regression (Option 3a) is those situations where the ecological gradient is not sufficiently covered by the existing data. This involves either small datasets of variable quality or where the existing gradient in ecological gradient is truncated, for example only covering 2 classes. Difficulties also exist in exercises involving only three methods of which one is considerably deviating from the other 2 methods. In such instances simple linear regression cannot be used to establish a model for comparing class boundaries, while the pseudo-common metric is either meaningless or strongly biased by the small number of Member States from which it must be calculated. In such cases methods should be compared according to direct comparison without regression.

Option 1 was also explained in the Intercalibration Guidance when Member States are using the same methods but established different class boundaries. For this option the procedure of Option 3a can be followed to check boundary bias and class agreement. The condition of relatedness of methods is of course clearly fulfilled in this case.

A table is provided for each of the 3 relevant questions in judging whether national methods can be considered comparable for each Option (Table 1, Table 2, Table 3). This relies on minimum acceptable standards for the relationship between methods, the boundary bias and the class agreement.

Table 1 for judging acceptability of comparisons based on related-ness between methods

1. Relatedness methods		Acceptability levels in analyses
Option 2	Parametric regression against common metric	Test the model quality + Pearson's correlation coefficient r ≥ 0.5
Option 3a	Parametric regression against pseudo-common metric (and/or common metric)	
Option 3b	Non-parametric correlation of EQR outcomes of methods	Spearman's Rank correlation coefficient ≥ 0.5

Reaching an acceptable boundary bias means that the difference between the national boundaries of the MSs is not larger than half a class (0.25 class above or below the boundary) (Table 2).

Reaching an acceptable class agreement means that the overall class difference over all the Member States remains lower than 1 class. An overall class difference of less than 0.5 class is considered as a good agreement. It can be accepted to be larger, but a difference equaling 1 class or more is questionable. In the rare cases that the class agreement criterion may not be reached after adjustment of the boundaries, a justification of this result should be provided to agree upon inclusion in the European Commission Decision.

When both conditions on boundary bias and class agreement are fulfilled, it can be concluded that the methods are sufficiently comparable.

Table 2 for judging acceptability of comparisons based on levels of boundary bias

2. Boundary bias[1] evaluation per MS	
All options [2]	≤ 0.25 classes (green)
	> 0.25 classes [3] (red)

Table 3 for judging acceptability of comparisons based on class agreement

3. Class agreement over all MSs	
Mean average absolute class difference	
All options	< 1.0 (green)
	≥ 1.0 [4] (red)

Green: Result accepted. Red: Result not accepted (boundary bias) or questionable (class agreement)

[1] Refers to the deviation from the global mean or median boundary for all national boundaries individually

[2] For option 3b calculated as average class difference (a global mean or median boundary cannot be defined in this case)

[3] In exceptional cases a justification can be given for a boundary bias that can slightly exceed 0.25 class, but it must never be larger than 0.5 class.

[4] In exceptional cases a justification can be given for a mean average absolute class difference that reaches or exceeds 1 class.

The next section of this guidance annex establishes a complete series of steps to perform such comparisons and to achieve comparability through a consistent analysis of quantitative datasets. It explains how comparability metrics for measuring the relation between methods and for boundary bias and class agreement are calculated.

3. Steps of the procedure for comparability analysis

The comparison of national classifications is one of the last steps of the intercalibration process and this can only start when the previous steps of the intercalibration process have been performed, as outlined in the Intercalibration Guidance. These include:

- Good status boundaries have been set by the Member States.
- The national assessment methods have passed the compliance check; hence all the participating methods are compliant with the WFD requirements (5 classes, EQRs, normative definitions etc.).
- The intercalibration process has passed the feasibility checks; hence all the participating methods address similar assessment concepts, the same types and pressures.
- Common intercalibration datasets have been compiled and have been checked against the data acceptance criteria.
- Common intercalibration datasets contain reference/benchmarking sites selected using common criteria (see Annex III).

All necessary steps to conduct the comparison and harmonization of national status classifications are summarized in Table 4. Flow charts depicting these steps are given in Figure 5, illustrating the homogenized procedure that all three Options have to follow. For each Option separately Figures 6 to 9 show how these steps are linked and which conclusions lead to the next step. **The comparison process should be performed first for the G/M boundary and then for the H/G boundary.** Where there are no sufficient existing data to reliably assess the high status, this boundary cannot be intercalibrated. In this case the comparison of the boundary setting protocol between Member States has to be explored to define a common theoretical estimation of the high/good boundary.

Table 4: Overview of process steps

Step	Description
1	*Benchmarking (as a preparation to the comparability criteria analysis)*
2	*Benchmark standardization (as a preparation to the comparability criteria analysis)*
	Comparability criteria analysis
3	*Construction of an ordinary least squares regression*
4	*Assessing how closely the methods are related*
5	*Boundary translation using the regression*
6	*Assessing the level of boundary bias: defining permitted boundary deviation*
7	*Boundary adjustment*
8	*Provide EQR classifications for direct comparison*
9	*Assessing the level of class agreement*
10	*Translation of benchmark standardized EQRs back to the national scale*
11	*Ecological characterization of class boundaries*

Step 1: Benchmarking (as a preparation to the comparability criteria analysis)

Objective: find a common starting point in an ecological gradient in relation to pressures

Intercalibration benchmarks are derived using common criteria. This step has been explained in the IC Guidance and is repeated here, since it is such an important precondition to comparability criteria analysis. See flow chart of the main steps of the intercalibration process (Figure 1 - Q6) and the guidance for deriving reference conditions and defining alternative benchmarks in Annex III of the Intercalibration Guidance.

- Use a population of sites screened against agreed abiotic criteria relevant to the major pressure(s).
- In case there are no (or insufficient) reference sites: define a common benchmark at a lower threshold for which all Member States can provide data.
- Process MUST be independent of national classifications (high status sites cannot just be accepted, has to be illustrated with abiotic data and common agreement).

Step 2: Benchmark standardization (as a preparation to the comparability criteria analysis)

Objective: correct any sub-typological differences that can cause incomparability, minimize biogeographical differences within a common dataset. This allows focusing the intercalibration strictly on the relative positioning of class boundaries and avoiding larger adjustment to methods than needed. This is not needed when either (i) the values of a common metric of benchmark sites are not significantly different between different biological or geographical subtypes (Option 2), (ii) the EQR values of a given national method for the benchmark sites does not differ between these subtypes (Option 3) or (iii) examples of all subtypes occur in all Member States participating in an exercise.

- **Option 2 indirect comparison**: Apply the common metric to the benchmark dataset of each MS, take the median value of the benchmark sites for each MS. Divide the common metric values applied to all the sites of each MS by the median value of their corresponding benchmark sites. This also converts the common metric values to an EQR, if this would not have been the case yet.
- **Option 3a and 3b direct comparison**: Each MS must apply its national method to the benchmark dataset of every other MS, to see if the benchmark of the other MS is higher or lower on its own national scale. Each EQR value, that is the result of the application of a specific national method on each site, must be divided by its corresponding median benchmark value to get the same scale in the whole common dataset.

Step 3: Construction of an ordinary least squares regression

Objective: establish the relationship between the national method and the common view of other Member States.

- **Option 2 indirect comparison**: relate the EQRs of the sites of a MS to the benchmark standardized Common Metric EQR (Figure 10).

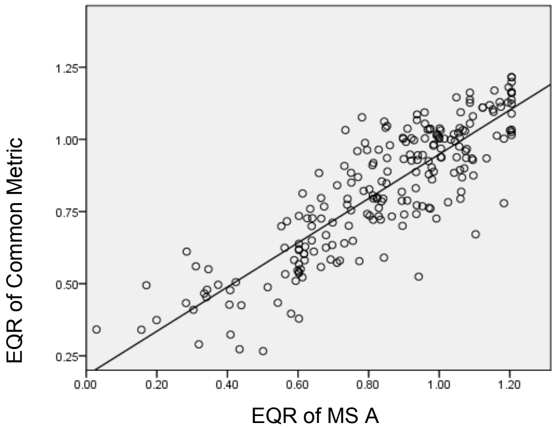


Figure 10: Linear regression of national EQRs of Member State A against a common metric

- **Option 3a direct comparison**: the benchmark standardized EQR for a site of a MS has to be plotted against the average benchmark standardized EQR of each combination of independent MSs for the same site = Pseudo-common metric.
- This step makes it possible to apply a regression approach to Option 3 and to define a global mean or median boundary to optimize the placement of the class boundaries also in Option 3.
- **Option 3b direct comparison**: In case of a direct comparison of 3 methods without regression the relation between methods is judged by a non-parametric correlation analysis of the EQR results of the different Member States' methods.

Step 4: Assessing the relation between methods

Objective: check how closely the methods are related.

- Any Member State not significantly correlated with the average view of the Member States must be excluded from the process and requested to improve its method (see Intercalibration feasibility check 1 and 2 in Figure 1 flow chart of the Intercalibration Guidance).
- To ensure a minimum acceptable level the size of the correlation rather than its probability is critical. The validity of the model should also be carefully checked.
- The regressions in the indirect (Option 2) and direct comparisons (Option 3a) provide a simple means of establishing the bias of individual methods relative to the common metric (Option 2) or the pseudo-common metric (Option 3a) to see how closely they are related.
- **Option 2 indirect comparison & Option 3a direct comparison**: Regression characteristics that have to be checked:
  - Relationship must be significant depending on the size of the dataset (from $p \leq 0.05$ to $p \leq 0.001$).
  - (Pseudo-) common metric must adequately represent all methods: correlation between each method and the (pseudo-) common metric should be $\geq r = 0.5$ (Pearson's correlation coefficient).
  - Slope of the regression should be tested to be significantly different from 0 and should preferably lie between 0.5 and 1.5.
  - Observed minimum r^2 should be at least half of the observed maximum r^2.
  - All necessary assumptions of a linear regression need to be checked (normally distributed error and variance (homoscedasticity) and independence of model residuals of the regression).
- **Option 3b direct comparison** without regression: correlation characteristics are checked with the non-parametric Spearman's Rank correlation coefficient.

Step 5: Boundary translation using the regression

Objective: translate the national boundary positions to a common metric scale using the regression.

- **Option 2 direct comparison & Option 3a indirect comparison**: Use the regression formula (P)CM = m*(EQR Member State) + c for H/G and G/M boundaries for all the Member States to translate the EQR values of the national H/G and G/M boundaries onto the (pseudo-) common metric scale (Figure 11).

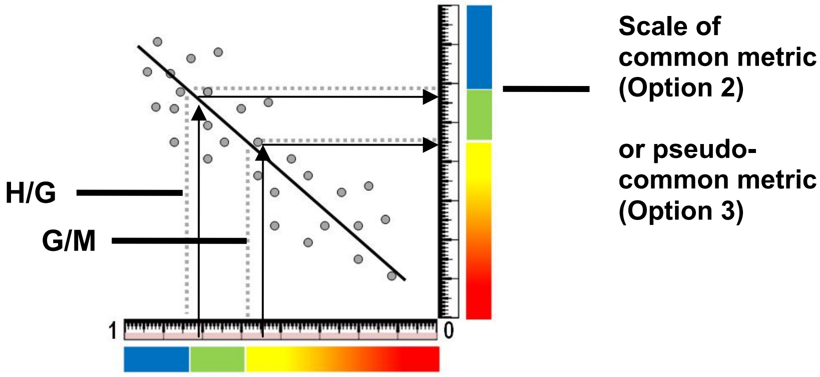


Figure 11: Translation of the H/G and the G/M boundary on a common scale.

- **Option 3b direct comparison** without regression: Perform a piecewise linear transformation of the EQRs of the national classification onto a common scale, to make all the classes equal to 0.2 (H/G boundary corresponding to 0.8, G/M boundary to 0.6, see Figure 12). Take the position of the reference, the H/G and the G/M boundary on each national scale and use these values to define the different transformation formulae for each MS.

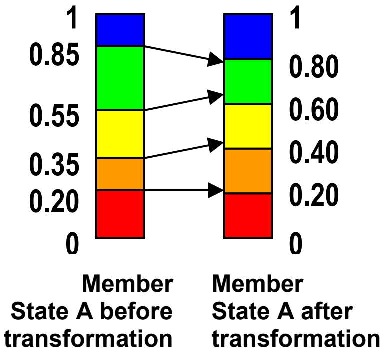


Figure 12: Piecewise linear transformation of EQRs

Step 6: Assessing the level of boundary bias: defining the permitted boundary deviation

Objective: derive the common view of boundary position and the range of permitted deviation.

- **Option 2 indirect and Option 3a direct comparison**: Define the global mean or median of all the MS predicted EQR values of the respective boundary on the (pseudo-) common metric scale.
  Define the difference between the global mean or median and the national boundary values on the (pseudo-)common metric scale and convert these differences into class equivalents. Define the permitted boundary deviation as a difference from the global mean or median that should not exceed 0.25 class equivalents for each national method. In case of exceeding this threshold, the respective boundary has to be adjusted (Figure 13). Where the comparison involves a low number of methods and the mean and median are not meaningful, the permitted boundary deviations of 0.25 class equivalents around the national boundaries should all overlap with each other.

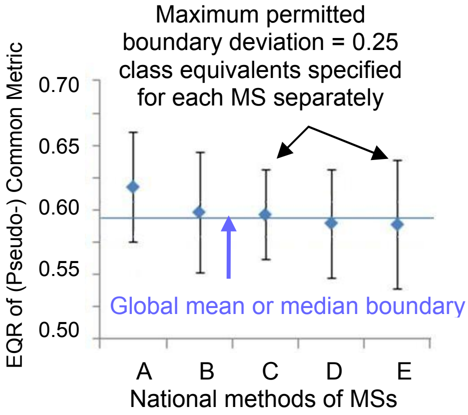


Figure 13: Boundary deviations of 0.25 classes of the global mean or median boundary for 5 Member States

- **Option 3b direct comparison** without regression: the definition of a global mean or median boundary is not meaningful in this case. The agreement in boundary setting is checked through the calculation of the average class difference: the EQR differences between each method and every other method for all commonly assessed sites are calculated and an average of the overall population of pairwise differences is determined for each method, separately for the G/M and the H/G boundary. This provides a measure of the difference in EQR values between one method and all other methods, which can be converted directly to class equivalents (dividing by 0.2). To ensure a meaningful result it needs to be tested if the differences are homogeneously distributed and if the median is close enough to the average class difference.

Step 7: Boundary adjustment

Objective: translate the adjusted boundary that does not exceed the permitted boundary deviation (= the harmonized boundary) on the scale of the national method and adjust the national classifications where necessary.

- **Option 2 indirect and Option 3a direct comparison**:
  - Define the lower (and upper) acceptable class boundary by subtracting (adding) the permitted boundary deviation of 0.25 class in the respective class equivalents of the Member State from (to) the global mean or median boundary.
  - By inverting the regression model between the (pseudo-)common metric and the national EQR it is straightforward to determine where national class boundaries should be positioned in order to secure an acceptable level of bias in the boundary comparison.
  - To translate the adjusted boundary to the benchmarked national scale, invert the formula of the regression previously established: EQR of the Member State for the boundary = $(y_{harmonized}-c)/m$.
- **Option 3b direct comparison**: class boundaries are adjusted by an iterative process, changing the national class boundary of the most biased Member State by 0.01 EQR units each step until all Member States obtain an acceptable level of bias. After each time of iteratively modifying the class boundaries of a specific method, the result for the average of the differences in agreement is calculated. In the case of a comparison of only 2 or 3 methods, adjust the class boundaries of the methods simultaneously, by an amount of 0.01 EQR units on the benchmark standardized scale, until the bias value is below the agreed threshold.
- Member States whose boundary falls above the global mean or median boun-dary (so are more stringent) are not obliged to adjust their boundaries.
- Lower national class boundaries (M/P and P/B) may have to be adjusted as well to be compliant with the normative definitions if the G/M boundary had to be adjusted, although this is a national issue.

Step 8: Provide EQR classifications for direct comparison

Objective: prepare the dataset for direct comparison of class agreement.

- Data need to cover the full range of ecological quality.
- For the G/M boundary the classification of the sites simply distinguishes between those above or below good.
- The direct comparison of class agreement does not use the categorical class identification (high, good, moderate,…), but uses the numerical EQR values.
- **Option 2 indirect comparison**: Generate a set of modeled national EQR values for each Member State across 300 sites, using the relationship between the benchmark standardized Common Metric and the national EQR for each Member State, with an error distribution at each predicted point specified as follows: normally distributed, mean = 0, sd = regression prediction error, to get results for all MSs for the 300 sites. These datasets are easy to compile and can be done centrally as a support to the GIGs.
- **Option 3a direct comparison**: All EQR values must be transformed to a common scale, to make all the classes equal to 0.2 (H/G boundary corresponding to 0.8, G/M boundary to 0.6). Take the position of the reference, the H/G and the G/M boundary on each national scale and use these values to perform a piecewise linear transformation of the EQRs of the national classification onto a common scale, with different transformation formulae for checking the H/G boundary class agreement or the G/M boundary class agreement (Figure 12).
- **Option 3b direct comparison**: this transformed dataset had to be obtained already for checking boundary bias, but the transformation has to be repeated after boundary adjustment.

Step 9: Assessing the level of class agreement

Objective: check the class agreement after the boundary harmonization.

Class agreement is a measure of the confidence that two or more national methods will report the same class for a given site, as assessed by the average absolute class difference between all pairs of EQR values across all participating Member States, the proportion of classifications differing by an agreed amount (i.e. half a quality class), and the multi-rater kappa coefficient.

- **Option 2 indirect comparison and Option 3a and 3b direct comparison**: Calculate the unsigned (i.e. absolute) differences between each method and every other method for all commonly assessed sites and define the average absolute class difference. Define the mean percentage of deviations of ≤0.5 classes. Obtain also the multi-rater kappa coefficient distinguishing the classification in above or below the considered boundary.

Step 10: Translation of benchmark standardized EQRs back to the national scale

Objective: determine where on the original national EQR scale the benchmark standardized EQR value should lay after harmonization. The harmonized boundary still has to be translated to the original (non-standardized) national scale when it is clear what the harmonized boundary on the benchmark standardized scale has to be.

- Multiply national class boundaries by the median of the population of benchmark sites (previously used for benchmark standardization by division)

Step 11: Ecological characterization

To support the quantitative comparison of class boundaries it is essential that GIGs produce a narrative for each intercalibration type that clearly establishes where the harmonized High/Good and Good/Moderate boundaries lie in terms of their ecological characteristics. As was clarified in the IC Guidance, the biological communities representing the “borderline” conditions between good and moderate ecological status and between good and high ecological status have to be described. This shall be done using sites of the common dataset that fall into a selected boundary range (e.g. harmonization band of national good-moderate boundaries expressed in common metric scale) (see Intercalibration Guidance p.28).

4. Data considerations

Objective: reduce the effect of dataset characteristics on comparability

- Need to work with EQRs (not truncated at 1)
- Minimum requirements for meaningful quantitative comparisons:
  - Size of datasets as large as possible for all the intercalibration options (sample – water body level: repackage samples to increase data availability, aggregation leads to reduction in the gradient of ecological quality): a quantitative comparison is difficult with <20-25 discrete cases classified by every Member State in option 3.
  - Data should cover the widest possible gradient in ecological quality.
  - Bilateral comparisons are preferably avoided.

5. Glossary

- **Site** = the smallest unit to which you can assign an EQR. It has to be representative for the water body. It can be one sampling station or a group of sampling stations, dependent how the assessment method is constructed and if it needs aggregation of data to assign an EQR or not.
- **Boundary bias** = the deviation in the relative positioning of class boundaries and measured by the magnitude and direction of deviation by a class boundary of one national method relative to the common view of the Member States (i.e. defined by the common metric or by the global mean of all the methods = pseudo-common metric, for the H/G and for the G/M class boundary). This deviation is expressed in class equivalents. It reflects the level of ambition of different methods or how stringent Member States are in defining the good ecological status.
- **Class agreement** = the confidence that two or more national methods will report the same class for a given site, as calculated by the average absolute class difference between all pairs of EQR values across all participating Member States, the proportion of classifications differing by an agreed amount (half a class), and the multi-rater kappa coefficient.
- **Pseudo-common metric** = the average benchmark standardized EQR of each combination of independent MSs for the same site. So the pseudo-common metric value for MS A compared to MS B, C & D for 1 site is the average of 3 EQR values: the EQR assigned by MS B, the EQR assigned by MS C and the EQR assigned by MS D for this site (so not including the EQR assigned by MS A for this site).
- **Benchmark standardization of the EQRs** = adjustment of all the benchmark median values to 1 in order to eliminate the differences in benchmark values that may exist between Member States or subtypes in the common dataset.
- **EQR classifications with transformed classes of equal widths** = classifications for which all the class widths are equal to 0.2 (H/G boundary corresponding to 0.8, G/M boundary to 0.6) in order to compare the EQR results directly. This is obtained by a piecewise linear transformation of the EQRs of the national classifications.

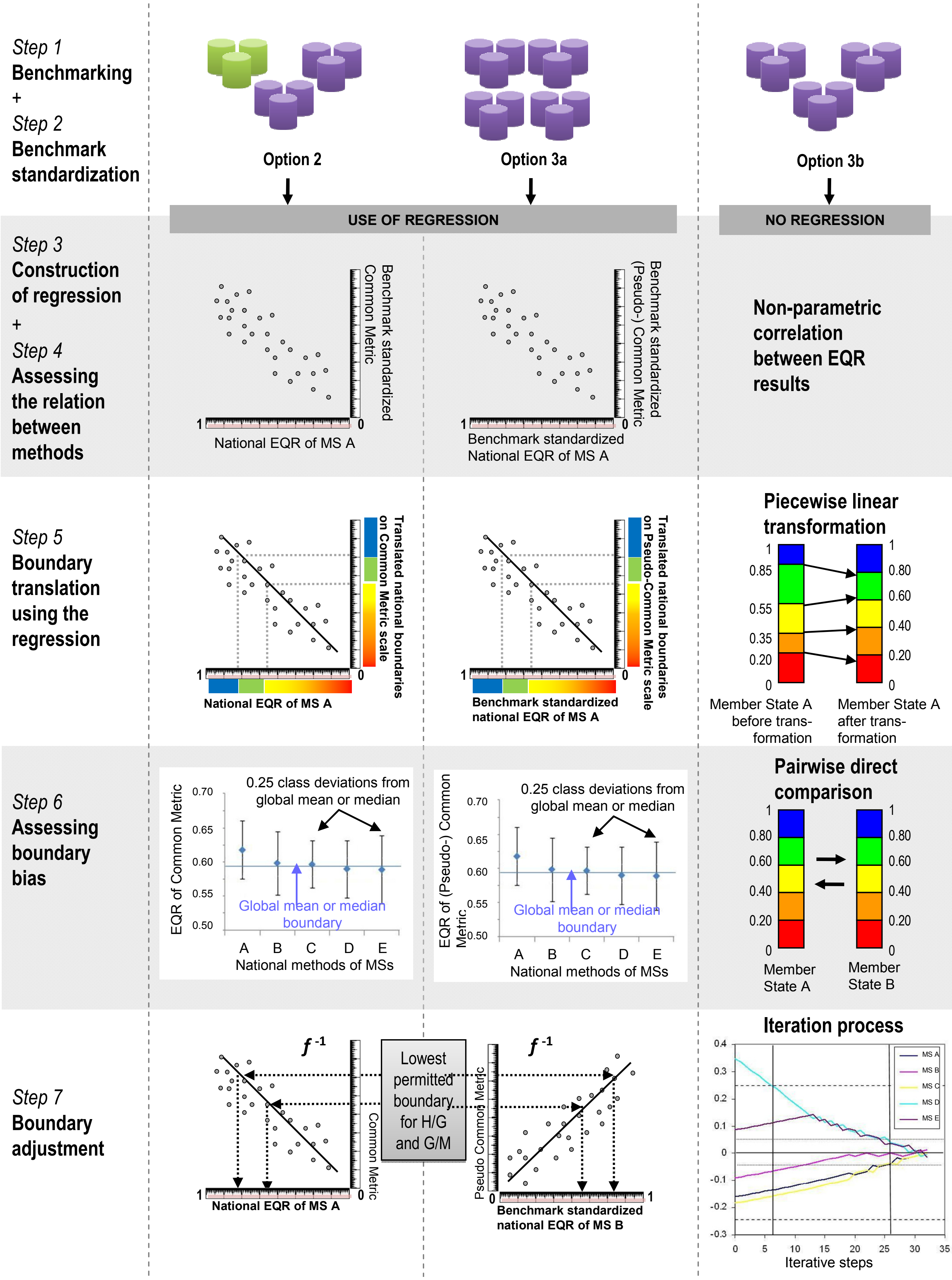


Figure 5: Process flowchart for the intercalibration Options 2, 3a and 3b

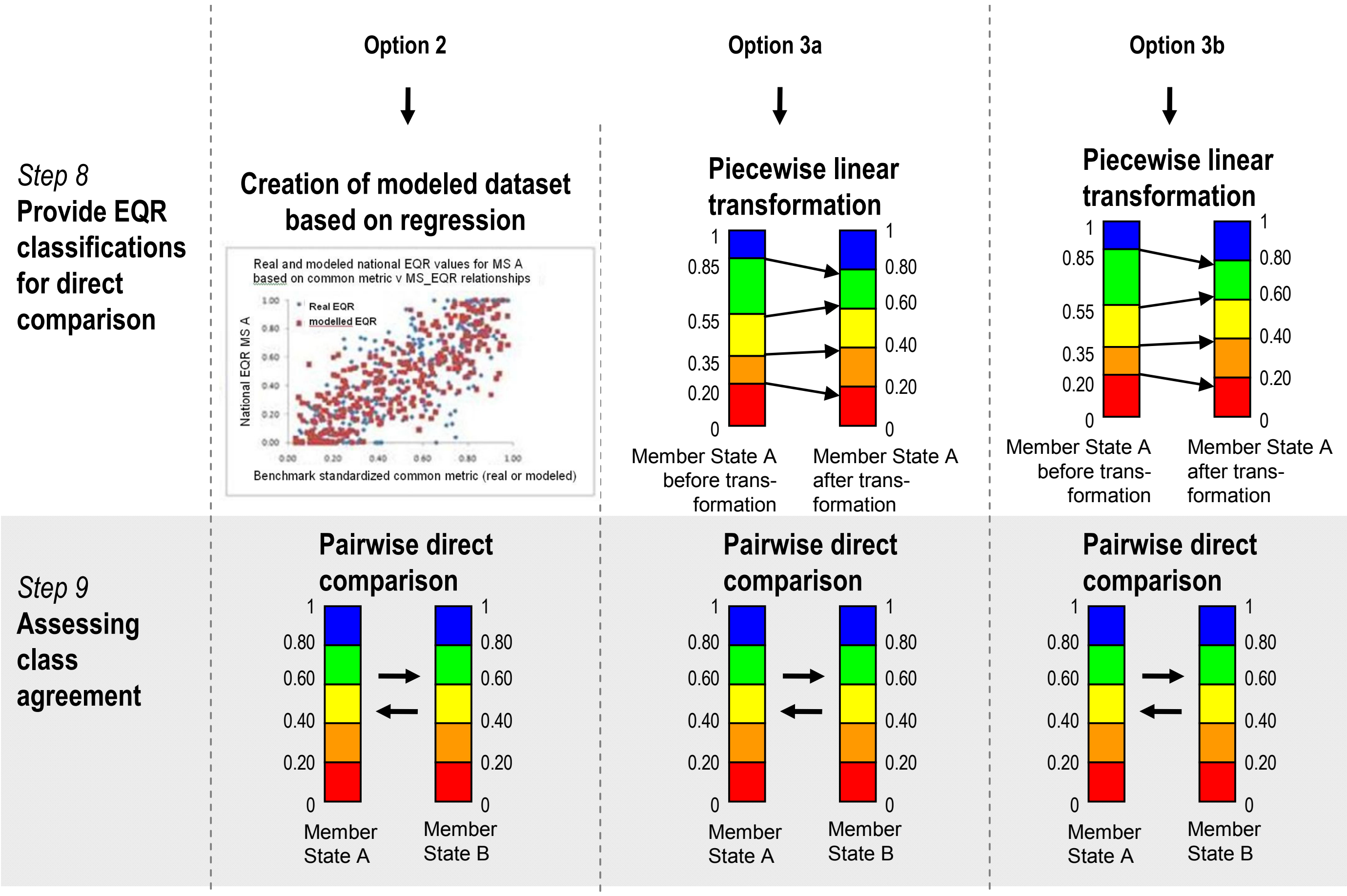


Figure 5 (continued): Process flowchart for the intercalibration Options 2, 3a and 3b

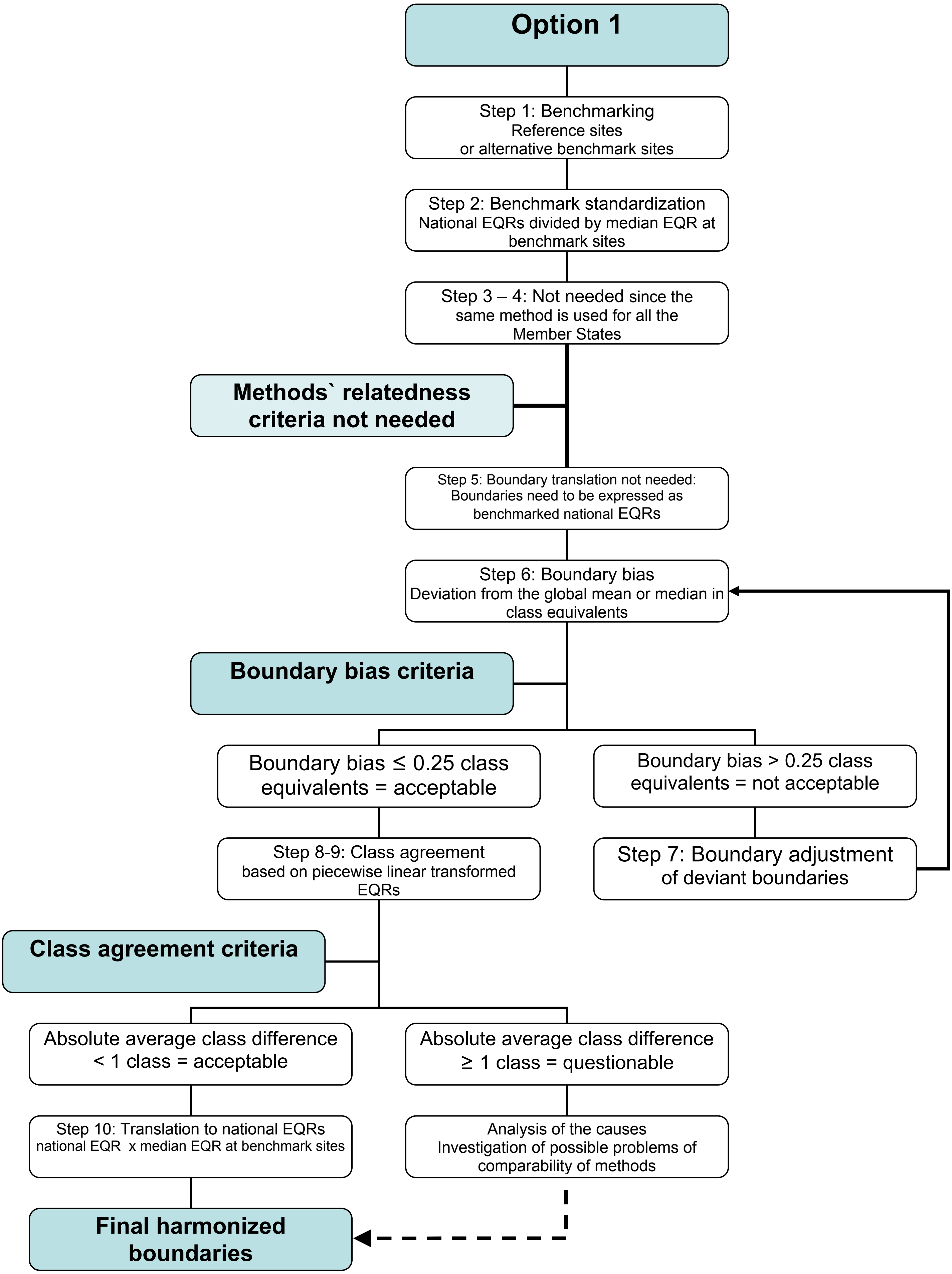


Figure 6: Flow chart of the main steps for Option 1

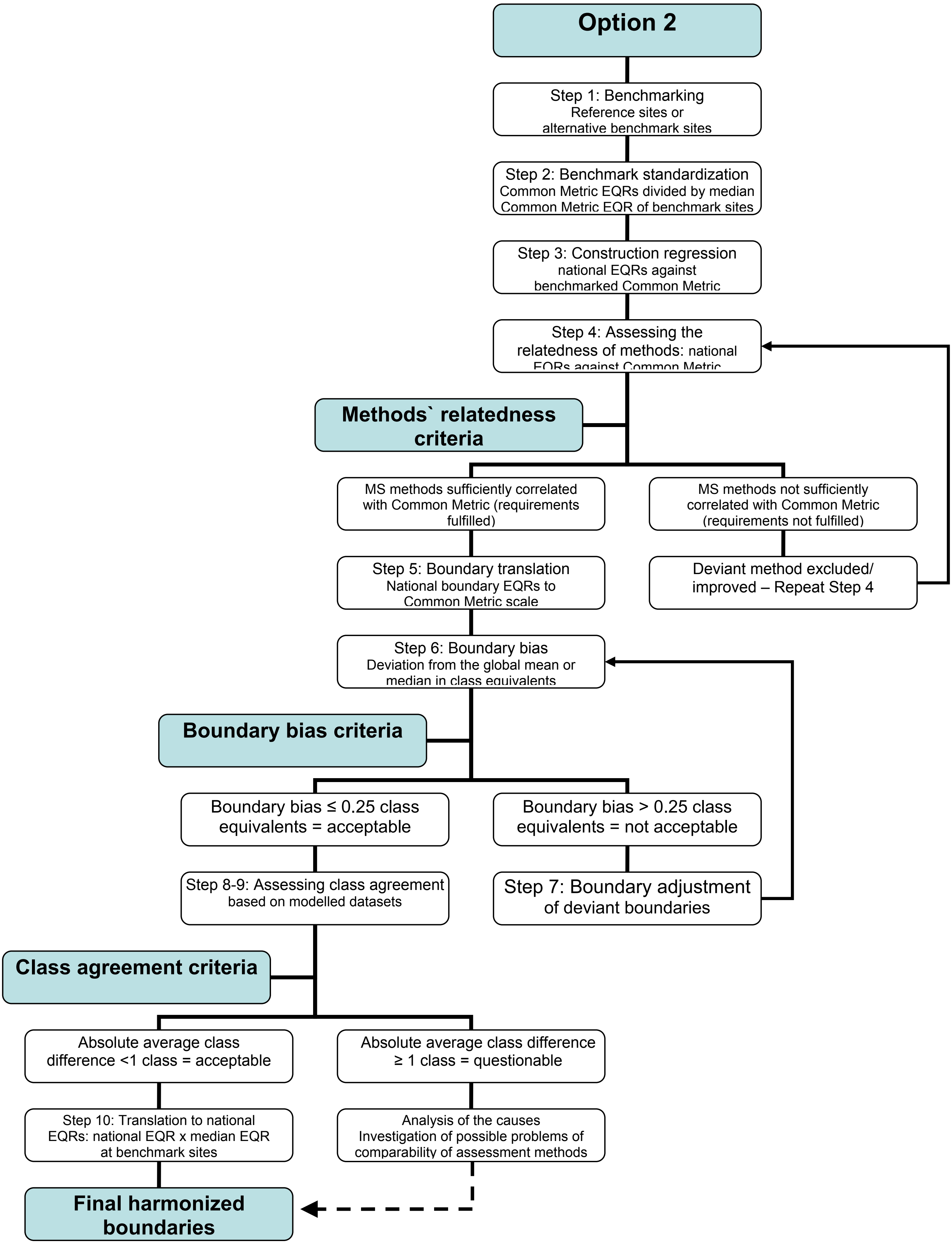


Figure 7: Flow chart of the main steps for Option 2

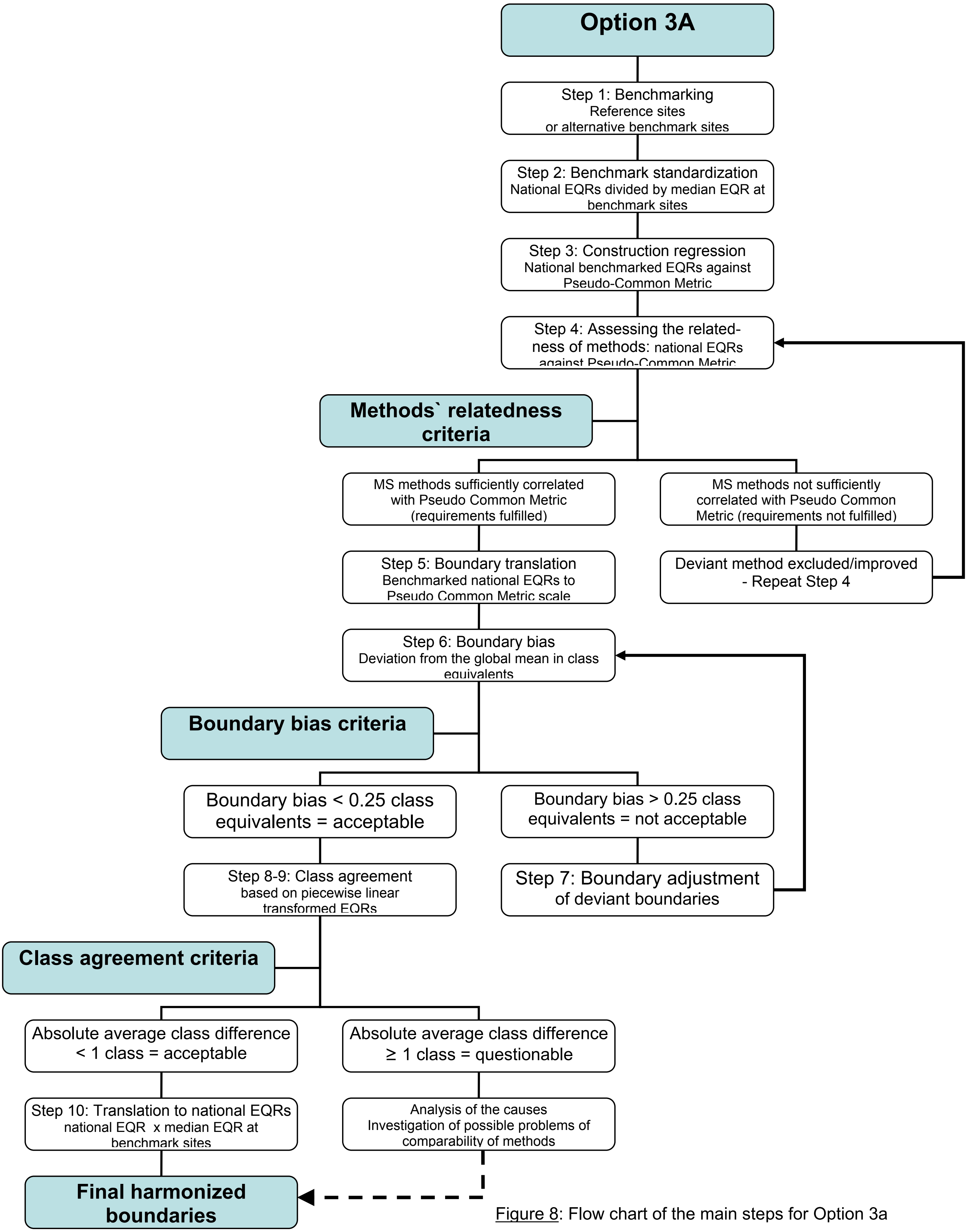


Figure 8: Flow chart of the main steps for Option 3a

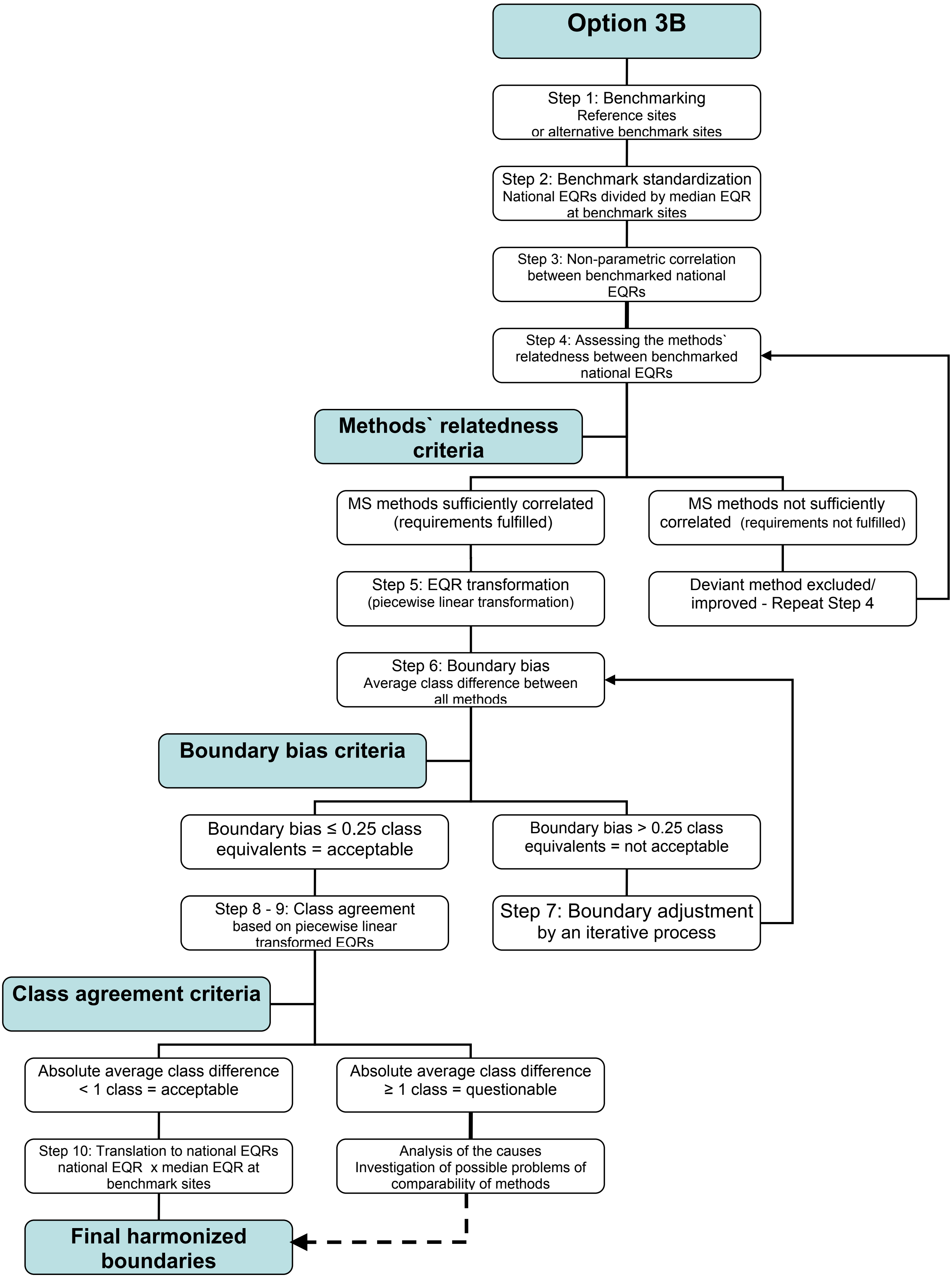


Figure 9: Flow chart of the main steps for Option 3b

Annex VI: Reporting template for the milestone reports

Intercalibration groups are responsible to regularly report the progress of the intercalibration process to the IC Steering group and WG ECOSTAT.
Reporting milestones are related to the major steps of the IC flowchart (Figure 1) and linked to the WG ECOSTAT meetings. Altogether five Milestones are foreseen for 2009-2011 with the following key elements:

Contents of the milestone reports

Milestone 1 (October 2009):
- Progress on WFD compliance checking (do all national assessment methods meet the requirements of the Water Framework Directive?);
- Progress on Feasibility checking (do all national methods address the same common types(s) and pressures(s) and follow a similar assessment concept?);
- Progress on Collection of IC dataset and Design the work for IC procedure;
- Review common IC types and description of pressures or pressure combinations to be intercalibrated;

Milestone 2 (April 2010):
- Update of info provided in Milestone 1
- Description of national assessment methods;
- Results of WFD compliance and feasibility check;
- Data set collected;
- Comparability of sampling and data processing;
- Selection of IC option;
- Development of IC common metric;
- Progress on Benchmarking Boundary comparison/setting.

Milestone 3 (October 2010):
- Update of info provided in Milestone 2
- Results of Benchmarking Boundary comparison/setting;
- Progress on Boundary harmonisation.

Milestone 4 (April 2011):
- Update of info provided in Milestone 3
- Boundary harmonisation completed;
- Proposal of class boundaries to be included in the IC Decision.

Milestone 5 (June 2011):
- Update of info provided in Milestone 4
- Final IC group reports.
- Finalised proposal of class boundaries to be included in IC Decision.

This annex VI outlines the main questions included in the Milestone reports to be reported to the IC Steering group and WG ECOSTAT.

Overview of deliverables of the Reporting template for the milestone reports	
Section	**Delivery deadline**
1. Organisation	Continuously
2. Overview of Methods to be intercalibrated + Information submission for Table 1 on national method descriptions (through method description questionnaires)	April 2010
3. Checking of compliance of national assessment methods with the WFD requirements	April 2010 + update in October 2010
4. Methods' intercalibration feasibility check	
5. Collection of IC dataset	
6. Benchmarking: Reference conditions or alternative benchmarking	October 2010
7. Design and application of the IC procedure	April 2010 + update in October 2010
8. Boundary setting / comparison and harmonization in common IC type	October 2010 + updates in April and June 2011
9. IC results	April 2011 + update in June 2011

Template for the milestone reports

Water category/GIG/BQE/ horizontal activity:	
Information provided by:	

1. Organisation (October 2009 + later updates)

1.1. Responsibilities

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

1.2. Participation

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

1.3. Meetings

List the meetings of the group:

2. Overview of Methods to be intercalibrated (April 2010 + later updates)

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

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

Member State	Method	Status

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

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

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

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

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

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

4. Methods' intercalibration feasibility check (April 2010 + update in October 2010)

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

4.1. Typology

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

Common IC type	Type characteristics	MS sharing IC common type
IC type 1		Member State A – yes Member State B - no
IC type 2		Member State A - yes Member State - yes

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

Method	Appropriate for IC types / subtypes	Remarks
Method A	IC type 1 IC type 2	
Method B	IC type 1 IC type 2	
Conclusion Is the Intercalibration feasible in terms of **typology ?**		

4.2. Pressures

Describe the pressures addressed by the MS assessment methods

Method	Pressure	Remarks
Method A		
Method B		
Conclusion Is the Intercalibration feasible in terms of **pressures** addressed by the methods?		

4.3. Assessment concept

Do all national methods follow a similar assessment concept?

Examples of assessment concept:

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

Method	Assessment concept	Remarks
Method A		
Method B		
Conclusion Is the Intercalibration feasible in terms of **assessment concepts?**		

5. Collection of IC dataset (April 2010 + update in October 2010)

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

Make the following table for each IC common type

Member State	Number of sites or samples or data values		
	Biological data	Physico- chemical data	Pressure data
MS A			
MS B			
MS C			

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

Data acceptance criteria	Data acceptance checking
Data requirements (obligatory and optional)	Member State A Member State B
The sampling and analytical methodology	Member State A Member State B
Level of taxonomic precision required and taxalists with codes	
The minimum number of sites / samples per intercalibration type	
Sufficient covering of all relevant quality classes per type	
Other aspects where applicable	

6. Benchmarking: Reference conditions or alternative benchmarking (October 2010 + later updates)

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

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

6.1. Reference conditions

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

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

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

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

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

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

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

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

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

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

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

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

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

7. Design and application of the IC procedure (April 2010 + update in October 2010)

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

Which IC option did you use?

- IC Option 1 - Same assessment method, same data acquisition, same numerical evalua-tion (Y/N)
- IC Option 2 - Different data acquisition and numerical evaluation (Y/N)
- IC Options 3 - Similar data acquisition, but different numerical evaluation (BQE sampling and data processing generally similar, so that all national assessment methods can reasonably be applied to the data of other countries) → supported by the use of common metric(s) (Y/N)
- Other (specify) (Y/N)

Explanation for the choice of the IC option:

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

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

Describe the IC Common metric:

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

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

Member State/Method	r	p
A		
B		

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

8. Boundary setting / comparison and harmonization in common IC type (October 2010 + later updates)

Clarify if

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

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

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

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

- Provide a description how you applied the full procedure (use of discontinuities, paired metrics, equidistant division of continuum)
- Provide pressure-response relationships (describe how the biological quality element changes as the impact of the pressure or pressures on supporting elements increases)
- Provide a comparison with WFD Annex V, normative definitions for each QE/ metrics and type

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

8.3. Boundary comparison and harmonisation

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

- Do all national methods comply with these criteria ? (Y/N)
- If not, describe the adjustment process:

9. IC results (April 2011 + update in June 2011)

- Provide H/G and G/M boundary EQR values for the national methods for each type in a table

Member State	**Classification**	**Ecological Quality Ratios**	
	Method	**High-good boundary**	**Good-moderate boundary**
	Common metric		
MS1	Method 1		
MS2	Method 2		
MS3	Method 3		

- Present how common intercalibration types and common boundaries will be transformed into the national typologies/assessment systems (if applicable)

- Indicate gaps of the current intercalibration. Is there something still to be done ?

Glossary

Term	Explanation
Method acceptance criteria	List of criteria evaluating whether assessment methods can be included in the intercalibration exercise, e.g. address the same common type(s) and anthropogenic pressure(s), and follow a similar assessment concept
Alternative benchmark	Trans-national reference point for intercalibration that is different from near-natural conditions, e.g. representing a similar level of least disturbed conditions
Artificial Water Body (AWB)	A body of surface water created by human activity (Article 2(8)). An artificial water body is a surface water body which has been created in a location where no water body existed before and which has not been created by the direct physical alteration or movement or realignment of an existing water body
Assessment method	The biological assessment for a specific biological quality element, applied as a classification tool, the results of which can be expressed as EQR. Since the assessment method evaluates different required parameters indicative of a biological quality element, it is a combination of biological metrics (each designed for a specific parameter)
Biological metric	A metric quantifies some aspects of the biological population's structure, function or other measurable characteristic that changes in a predictable way with increased human influence
Biological quality element (BQE)	Particular characteristic group of animals or plants present in an aquatic ecosystem that is specifically listed in Annex V of the Water Framework Directive for the definition of the ecological status of a water body (for example phytoplankton or benthic invertebrate fauna)
Common Implementation Strategy (CIS)	The Common Implementation Strategy for the WFD was agreed by the European Commission, Member states and Norway in May 2001 with the main aim to provide support in the implementation of the WFD by developing common understanding and guidance on key elements of the Directive. A series of working groups has been developed to help carry out the appointed tasks: - Raise awareness and exchange information; - Develop Guidance documents on various technical issues; Carry out integral testing in pilot river basins
Class boundary	The Ecological Quality Ratio value representing the threshold between two quality classes
Classification	The arrangement of similar entities into classes according to established criteria related to the environmental conditions of a water body
Boundary setting protocol	A procedure for defining boundaries of ecological status classes between different Member States in the framework of WFD intercalibration

Borderline conditions	Description of the biological communities representing the “borderline” conditions between good and moderate ecological status. This shall be done using sites of the common dataset that fall into a selected boundary range (e.g. harmonisation band of national good-moderate boundaries expressed in common metric scale).
Common metrics	A biological metric widely applicable within a GIG or across GIGs, which can be used to derive a comparable understanding of reference conditions/alternative benchmark and boundary setting procedure among different countries/water body types
Comparability criteria	Criteria for evaluating sufficient comparability of good ecological status between different national assessment methods
Commission Decision	Legally binding decision of the European Commission. The Commission Decision on the WFD intercalibration includes the results of the intercalibration exercise and the values established for the national assessment methods
Common intercalibration type	A type of surface water differentiated by geographical, geological, morphological factors (according to WFD Annex II) shared by at least two Member States in a GIG
Compliance criteria	List of criteria evaluating whether assessment methods are meeting the requirements of the WFD, e.g. • Ecological status is classified by one of five classes (high, good, moderate, poor and bad); • High, good and moderate ecological status are defined in accordance with the normative definitions of WFD (Annex V); • All relevant parameters indicative of the biological quality element are covered (Annex V)
Data acceptance criteria	Minimum data requirement and data quality criteria in order to obtain comparable datasets
Ecoregion	The geographical area illustrated in WFD Annex XI Maps A (rivers and lakes) and B (transitional and coastal waters)
Ecological potential	The status of a heavily modified or artificial water body measured against the maximum ecological quality it could achieve given the constraints imposed upon it by those heavily modified or artificial characteristics necessary for its use
Ecological status	Expression of the structure and functioning of aquatic ecosystems associated with surface waters. In practice, ecological status is determined by biological quality elements, supported by hydromorphological and physico-chemical quality elements (Annex V)
ECOSTAT	WFD CIS Working Group Ecological Status that was established 2002 with the main objective to provide Member States and Candidate Countries with guidance on the intercalibration of the ecological status classification
Ecological Quality Ratio (EQR)	Calculated from the ratio observed value/reference value for a given body of surface water. The ratio shall be represented as a numerical value between zero and one, with high ecological status represented by values close to one and bad ecological status by values close to zero

Geographic Intercalibration Group (GIG)	Organizational unit for the intercalibration consisting of a group of Member States sharing a set of common intercalibration types
Harmonisation	The process by which class boundaries should be adjusted to be consistent with the normative definitions (Annex V Section 1.2) of the Water Framework Directive and comparable among Member States sharing the same type. It must be performed for HG and GM boundaries
Heavily modified water body (HMWB)	A body of surface water which as a result of physical alterations by human activity is substantially changed in character, as designated by the Member State in accordance with the provisions of Annex II
Impact	Effects of pressures on the status of surface water and groundwater (e.g. decrease in taxa richness due to habitat alteration)
Intercalibration	An exercise facilitated by the Commission to ensure that the high/good and good/moderate class boundaries are consistent with Annex V Section 1.2 of the Water Framework Directive and comparable between Member States
Intercalibration Steering Group	This group consists of the water category leads as well as other experts that are able to contribute, e.g. GIG leads and/or BQE leads. Potential external experts can be added. The review panel will have such tasks as checking WFD compliance of the methods and approving the results of the intercalibration.
Joint Research Centre (JRC)	European Commission Joint Research Centre which provides scientific and technical support for EU policy-making
Pressure	Human activities such as organic pollution, nutrient loading or hydromorphological modification that have the potential to have adverse effects on the water environment.
Reference conditions	For any surface water body type reference conditions or high ecological status is a state in the present or in the past where there are no, or only very minor, changes to the values of the hydromorphological, physico-chemical, and biological quality elements which would be found in the absence of anthropogenic disturbance.
River Basin Management Plan (RBMP)	A plan that must be produced for each River Basin District within a Member State under Article 13. The plan shall include the information detailed in Annex VII, including the programme of measures.
Strategic Co-ordination Group	A group led by the European Commission with participants from all Member States which was established to co-ordinate the work of the different working groups of the Common Implementation Strategy
Type specific reference conditions	Reference conditions representative for a specific water body type

Water body	Distinct and significant volume of water. For example, for surface water: a lake, a reservoir, a river or part of a river or a coastal area
WFD Water Framework Directive	Directive 2000/60/EC establishing a framework for Community action in the field of water policy

KH-AN-11-014-EN-N

ISBN 978-92-79-18997-5