



Technical Report - 2015 - 085

Procedure to fit new or updated classification methods to the results of a completed intercalibration exercise

Guidance Document No. 30

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

New freephone number:
00 800 6 7 8 9 10 11

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>).
Luxembourg: Office for Official Publications of the European Communities, 2015

ISBN 978-92-79-38434-9

ISSN 1725-1087

doi: 10.2779/158259

© European Union, 2015

Reproduction is authorised provided the source is acknowledged.

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.

**Procedure to fit new or updated classification
methods to the results of
a completed intercalibration exercise**

Guidance Document No. 30

Summary

The European Water Framework Directive (WFD) requires the national classifications of good ecological status to be harmonised through an intercalibration exercise. Most of the Geographical Intercalibration groups have finalized Intercalibration results but many Member States have not joined the group or have not intercalibrated the methods due to some reasons.

Those Member States will have to show that their methods are compliant with the WFD normative definitions and that their class boundaries are in line with the results of the intercalibration exercise.

This document provides a workflow to fit new or revised national classification methods to the harmonised definition of good ecological status established in the completed intercalibration exercise.

The document was drafted by Sebastian Birk (DE), Nigel Willby (UK), Sandra Poikane and Wouter van de Bund (EC Joint Research Centre), with the help of a drafting group consisting Geoff Phillips (UK), Martin Kelly (UK), Libuse Opatrilova (CZ) Jürgen Böhmer (DE) and Marcel van den Berg (NL). The first version of this document has been circulated to various intercalibration experts and was revised following the comments of Torsten Berg (SE), Paul Logan (UK) and Libuse Opatrilova (CZ).

It was agreed at WFD Working Group ECOSTAT meeting 23-24 October 2013 and endorsed as a guidance document in the WFD Common Implementation Strategy by EU Water Directors on 24 November 2014.

Table of contents

1	Introduction	3
2	Basics	4
2.1	General principles	4
2.2	Workflow	4
2.3	Key terms	5
2.4	Data basis for the intercalibration analysis	8
3	What needs intercalibrating?	9
3.1	Intercalibration requirements	9
3.2	Selection of the fitting procedure	9
4	Intercalibrating revised classification methods	11
5	Intercalibrating new classification methods	12
5.1	Case A1: IC Option 1 or 2 using <i>reference/benchmark sites</i>	12
5.1.1	Requirements	12
5.1.2	Process	13
5.2	Case A2: IC Option 1 or 2 using <i>continuous benchmarking</i>	16
5.2.1	Requirements	16
5.2.2	Process	16
5.3	Case B1: IC Option 3 using <i>reference/benchmark sites</i>	19
5.3.1	Requirements	19
5.3.2	Process	20
5.4	Case B2: IC Option 3 using <i>continuous benchmarking</i>	23
5.4.1	Requirements	23
5.4.2	Process	23
6	Comparability criteria	26
7	Possible problems and solutions	27
8	Literature	28

1 Introduction

The European Water Framework Directive (WFD) requires the national classifications of good ecological status to be harmonised through an intercalibration exercise. For many Biological Quality Elements (BQE) of selected intercalibration types this intercalibration exercise has been completed. The results are laid down in a Commission Decision (European Commission, 2013) and documented in technical reports (Poikane, 2013, van de Bund, 2013). Each exercise was carried out in a Geographical Intercalibration Group (GIG) and followed the procedure described in the CIS-guidance document on the intercalibration process (European Commission, 2011), summarised in Birk et al. (2013).

Some Member States did not intercalibrate their national classification because the method was not fully developed by the time the respective exercise was completed. Other Member States may wish to revise their already intercalibrated classification, for instance to account for technical improvements. In both cases, these Member States have to show that their new or revised classification is compliant with the WFD normative definitions and that their class boundaries are in line with the results of the completed intercalibration exercise.

The instruction manual at hand provides a workflow to fit new or revised national classification methods to the harmonised definition of good ecological status established in the completed intercalibration exercise. All statistical analyses described in the manual require a qualified national dataset compiled by the Member State seeking intercalibration of its classification method. Selected key parameters from the completed exercise (e.g. *global mean view* of high-good and good-moderate boundary positions; global pressure-impact model) form the basis of the fitting procedure to be performed by the Member State.

This manual comprises (i) the basics of fitting new or revised classification methods (Chapter 2), (ii) a decision-tree to facilitate the selection of an appropriate fitting procedure (Chapter 3), (iii) a simplified scheme to fit revised national classification methods (Chapter 4), and (iv) a step-by-step procedure for the two relevant cases of fitting new national classification methods (Chapter 5). It concludes with a summary of the comparability criteria (Chapter 6) and possible problems and solutions (Chapter 7).

2 Basics

2.1 General principles

Fitting new or revised national classification methods of ecological status to the results of the complete intercalibration exercise is part of the intercalibration process stipulated by the WFD. This implies that the relevant intercalibration types were already successfully intercalibrated by other Member States (as documented in Annex I of the Commission Decision) and that other members of the GIG now wish to join the exercise with newly developed methods, or that a Member State who participated in the completed exercise now wishes to revise its intercalibrated method.

As such, the CIS-Guidance Document on the Intercalibration Process (European Commission, 2011) forms the key reference with regard to intercalibration aims, methods and criteria. The CIS-Guidance defines all necessary steps of the intercalibration process. This manual, especially the flowchart depicted in Chapter 3, refers to these steps but does not go into detail concerning, for instance, the necessary checks and criteria prescribed in the guidance. The successful implementation of the fitting procedure described here requires compliance with the process defined in the CIS-Guidance.

Equally important for the practical application of this manual are the contents of the final Milestone Report that documents the approaches and outcomes of the respective intercalibration exercise. Each finalised intercalibration of national class boundaries represented an individual exercise with specific problems and solutions. Thus, those in charge of fitting new or revised national methods also need to familiarise themselves with the details of the relevant exercise. All necessary information can be found in the volume of technical reports edited by van de Bund & Poikane (2013). However, we advise contacting experts who were involved in the completed exercise to fully understand the background and constraints of the respective exercise.

2.2 Workflow

In principle, every Member State is responsible for intercalibration of its ecological assessment methods. Before the work, Member State has to report the ecological assessment method to ECOSTAT. During the process, Member state has to demonstrate the WFD-compliance and compliance with the standards established in the completed intercalibration exercise (in accordance to the CIS Guidance, IC technical report and Instruction manual).

An IC Review Panel consisting of the water category leads as well as other experts will check the WFD compliance of the new or revised methods and review the results of the fitting process. The CIS-Working Group A on Ecological Status (ECOSTAT) shall approve the outcomes of each individual exercise.

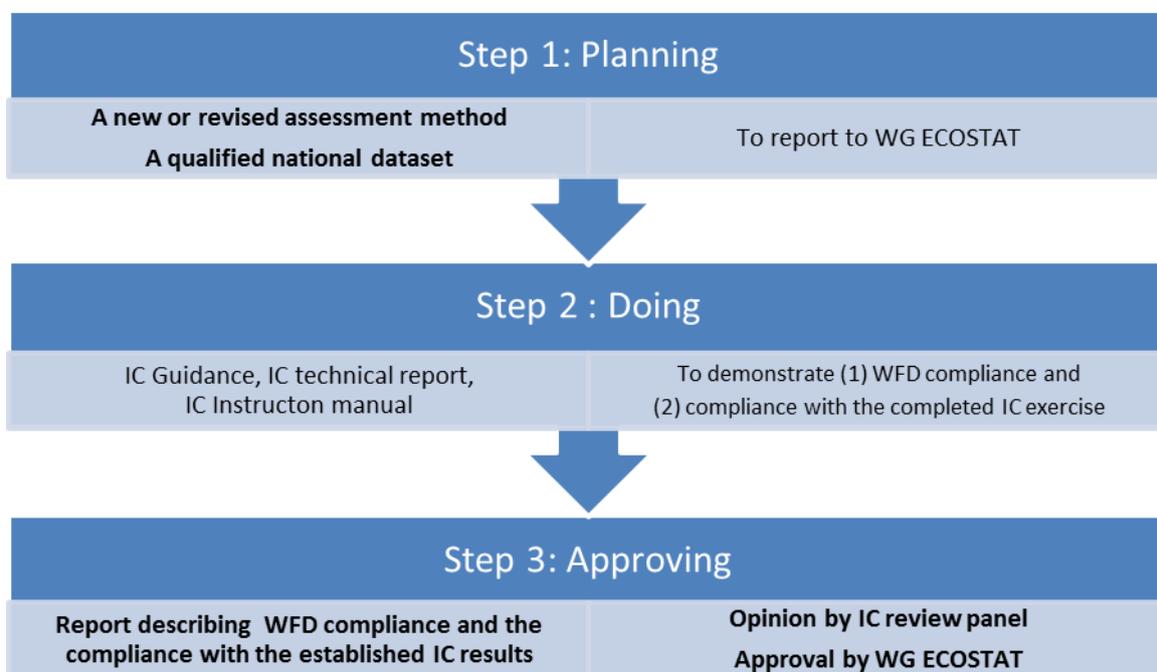


Fig 1. Flowchart describing process of fitting of new or revised national classification methods of ecological status to the results of the complete intercalibration exercise.

2.3 Key terms

Benchmark standardisation	Adjustment of national EQRs on the basis of common abiotic criteria or pressure-biology relationships to allow for a standardised comparison of national status boundaries against the <i>global mean view</i>
BRINC	Best-Related and Intercalibrated National Classification Surrogate metric for the use of a pseudo-common metric (PCM) in IC Option 3
Comparability criteria	Criteria for evaluating sufficient comparability of good ecological status between national classification method and the fixed standard established in the completed exercise
Completed exercise	Intercalibration exercise for a specific BQE of a certain common intercalibration type in a GIG for which intercalibration results were achieved (laid down in Annex I of the Commission Decision)
Continuous benchmarking	Option to perform the benchmark standardisation: Biological differences between national datasets were determined based on the country offsets (i.e. intercept and/or slope deviates) from the global pressure-biology relationship established using general linear models across the combined extent of the pressure gradient afforded by all countries
Data acceptance criteria	Minimum data requirement and data quality criteria in order to obtain a qualified dataset
Global mean view	The common view of the status boundaries by all Member States that participated in a completed intercalibration exercise, expressed in the currency of the (pseudo-) common metric. The <i>global mean view</i> represents the fix standard to which the joining method is fitted.
IC Option	Option to intercalibrate (IC) different national assessment methods in a completed exercise.
Joining method	Newly developed or revised national classification of a MS that joins a completed exercise to adopt the ecological standard established in a completed intercalibration exercise
Method acceptance criteria	Check if the national classification method is applicable to the same common IC types and pressures as addressed in the completed IC exercise, and if its assessment concept is similar to the concept of the methods intercalibrated in the completed exercise
New national classification	National classification method for a specific BQE that is newly developed and has thus not yet been intercalibrated for a certain common intercalibration type

Reference/benchmark sites	Option to perform the benchmark standardisation: Reference sites meet international screening criteria for undisturbed conditions. Benchmark sites meet a similar (low) level of impairment associated with the least-disturbed or best commonly available conditions.
Revised national classification	National classification method for a specific BQE that was intercalibrated for a certain common intercalibration type but has been modified regarding the components: data acquisition (e.g. sampling design, sample treatment), numerical evaluation (e.g. metric selection, indicator scores) or classification (e.g. reference definition, boundary setting)
<i>Site x biology</i> dataset	Qualified national dataset of the joining Member State including WFD monitoring data for several sites/water bodies covering a gradient of anthropogenic pressure (see Chapter 2.3)
WFD compliance criteria	List of criteria evaluating whether assessment methods are meeting the requirements of the WFD, e.g. <ul style="list-style-type: none">• 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).

Abbreviations

BM	Benchmark
BQE	Biological Quality Element
BRINC	Best-Related and Intercalibrated National Classification
BRINC_bm	Benchmark-standardised BRINC values
BRINC_obs	Observed (raw) BRINC values
BRINC_pred	Predicted (modelled) BRINC values
CM	Common Metric
CM_bm	Benchmark-standardised CM values
CM_obs	Observed (raw) CM values
CM_pred	Predicted (modelled) CM values
EQR	Ecological Quality Ratio
GIG	Geographical Intercalibration Group
GM	Good-moderate class boundary
HG	High-good class boundary
MP	Moderate-poor class boundary
MS	Member State
OLS	Ordinary Least Squares regression
PCM	Pseudo-Common Metric
R ²	Coefficient of determination

2.4 Data basis for the intercalibration analysis

The fitting procedure is carried out on the basis of a qualified national dataset. Ideally, this dataset should

- sufficiently cover the geographical area in which the common type occurs within the Member State,
- 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 non-biological data must be contemporaneous with the accompanying biological data in time and space in order to be used for pressure-impact analyses.

3 What needs intercalibrating?

3.1 Intercalibration requirements

This guidance applies to two specific situations: (a) completely new methods which have not previously been subject to intercalibration but for which an exercise has already been completed for the BQE and GIG in question, and (b) methods which were part of a completed exercise but have since been revised in some way. In both cases these methods must be WFD compliant and must meet the feasibility checks set out in the CIS guidance before any of the fitting procedures set out in this manual are applied.

3.2 Selection of the fitting procedure

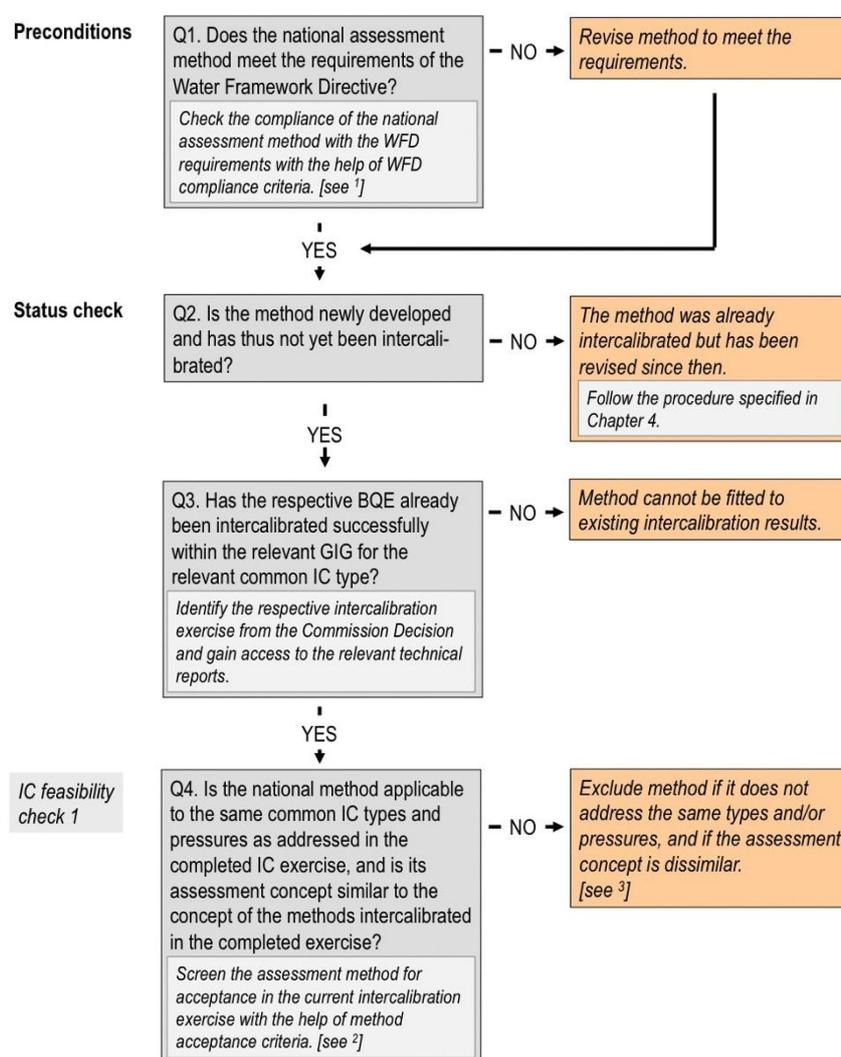


Figure 1: Flow diagram to select the appropriate fitting procedure
(to be continued on the next page)

¹ The WFD compliance criteria are specified in the reporting template for milestone reports (Annex VI of European Commission 2011). This template shall be used to document compliance.

² Method and data acceptance criteria are specific to the completed IC exercise. They have been defined in the relevant technical reports and need to be considered accordingly.

³ In specific cases an intercalibration is still possible, e.g. by compiling a specific dataset or considering only parts of the national classification method (but still demonstrating full comparability).

continued from previous page

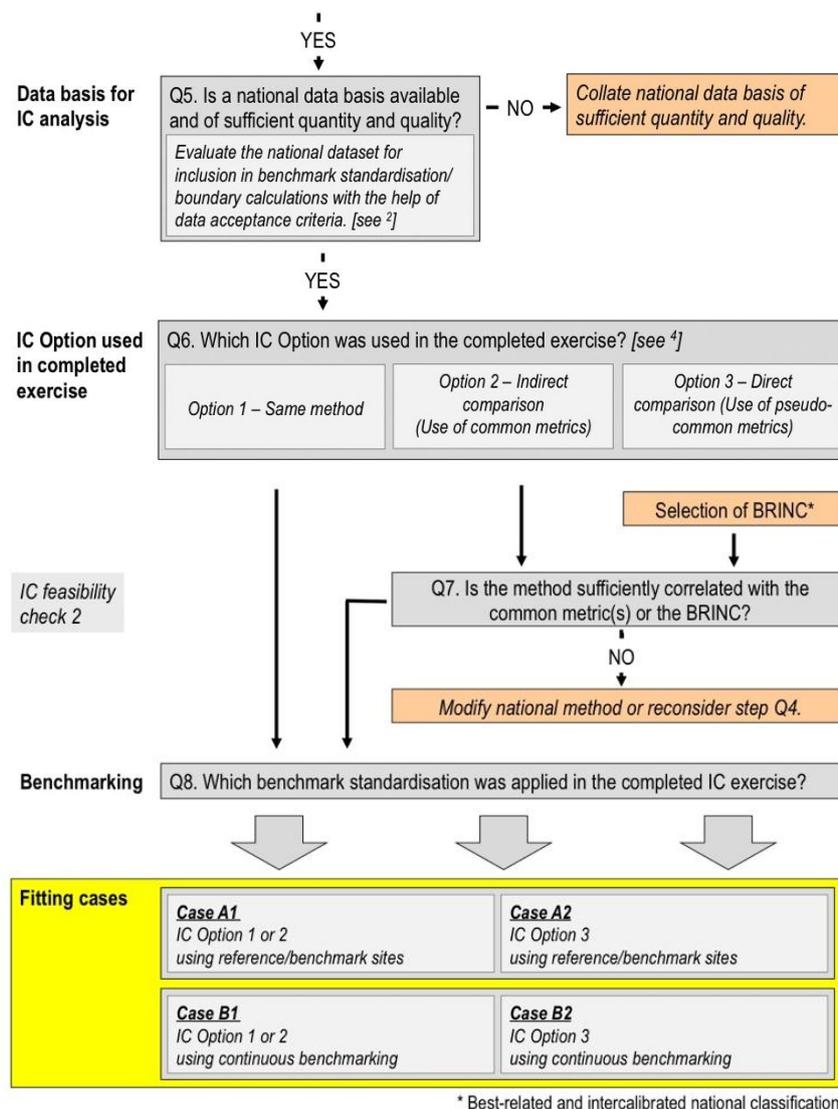


Figure 1: Flow diagram to select the appropriate fitting procedure
(continued from the previous page)

⁴ The fitting procedure laid down in this manual actually allows for selecting an IC Option different from the completed exercise. As long as the criteria of (i) IC feasibility and (ii) comparability are met, joining countries can choose an option most suitable to their requirements.

4 Intercalibrating revised classification methods

A “revised national classification method” is a method that was already intercalibrated for a certain common intercalibration type but has since been modified with regard to: data acquisition (e.g. sampling design, sample treatment), numerical evaluation (e.g. metric selection, indicator scores, combination rules) or classification (e.g. reference definition, boundary setting). Changes to any of these components may affect the comparability with the intercalibrated standard. You must follow the procedure below to fit boundaries to the revised methods (Figure 2):

1. Relate the ecological quality ratios (EQRs) of the old method with the EQRs of the revised (“new”) method using a qualified national dataset (see Chapter 2.4). In cases where the relation between the old and new method equates to $R^2 < 0.8$, intercalibration feasibility should be checked using the common dataset from the completed exercise by relating the EQRs of the new method to the common or pseudo-common metric used in that exercise. This step is not necessary in cases of $R^2 \geq 0.8$.
2. Translate the high-good (HG) and good-moderate (GM) boundaries of the old method into EQR-values of the new method using the ordinary least squares (OLS) regression equation in which:

$$\text{New method EQR} = c + m * \text{Old method boundary EQR}.$$

3. Compare the position of old and new methods’ boundaries on the EQR scale of the new method.
4. If the boundaries of the new method are higher (i.e. more precautionary) than or equal to the old boundaries, the fitting process is accomplished, since this implies that the criteria for boundary bias must have been met¹.
5. If the new boundaries are lower than the old boundaries, the comparability with the intercalibrated standard needs to be checked since the criteria for boundary bias might no longer be satisfied. In these cases, the procedure for fitting new classification methods must be followed.

¹ Member States are allowed to adopt class boundaries more stringent than the + 0.25 class bias relative to the *global mean view*.

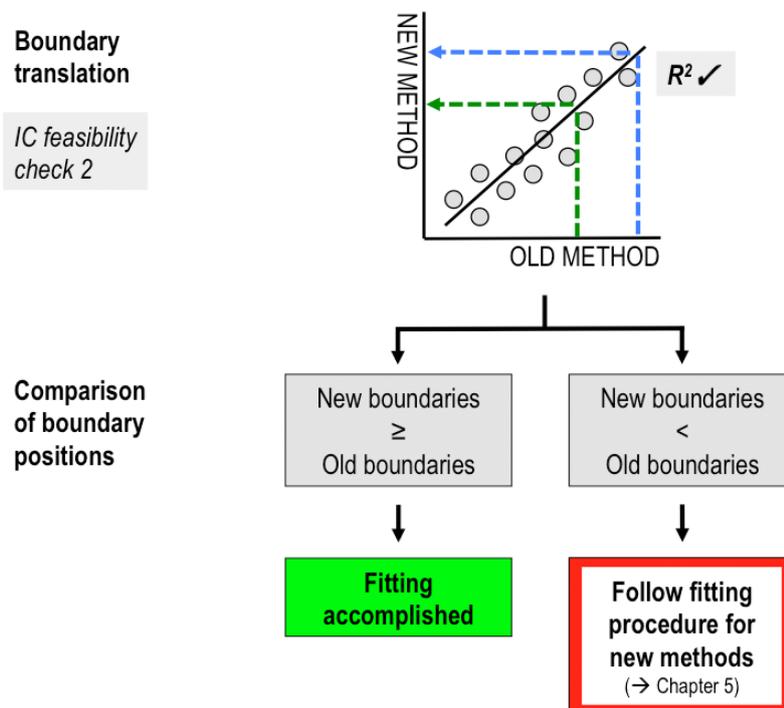


Figure 2: Procedure to fit revised classification methods

5 Intercalibrating new classification methods

Use the flow diagram (Chapter 3) to identify the correct procedure to use for intercalibrating new classification methods. This will depend on the procedure used for the completed intercalibration, described as Cases A1, A2, B1, and B2 in the following sections.

5.1 Case A1: IC Option 1 or 2 using *reference/benchmark sites*

This is the simplest case by which new or revised methods can be intercalibrated.

The key difference between case A1 and A2 is the means of benchmark standardisation with A2 relying on continuous benchmarking. Case A1 relies on a sufficient number of reference condition or other high quality sites that can be used for benchmarking, i.e. a minimum of three sites and preferably no fewer than the smallest number provided by any MS in the completed exercise

5.1.1 Requirements

- Full details of the common metric (e.g. species scores and metric weights).
- A suitable *site x biology* dataset covering a range of environmental quality from which the national EQR and common metric can be calculated.

- Accompanying pressure data in the same format as that used in the completed exercise.
- Information on the specific thresholds already used in the completed exercise to define reference or alternative benchmark sites (e.g. human population density, extent of agricultural land in catchment, nutrient concentrations etc.).
- Details of exactly how the benchmarking was undertaken in the completed exercise (e.g. creation of a common metric EQR by dividing the observed value by the median common metric value of a set of national reference or benchmark sites). If the completed exercise concluded that benchmarking was not necessary the mean value of the benchmark sites from each country must be provided so that the joining Member State can also judge the need to benchmark its own method.
- Values of the *global mean view* of the HG and GM boundaries on the common metric scale for Member States who participated in the completed exercise.

5.1.2 Process

1. Calculate the common metric (CM) on the national dataset.
2. Use the associated pressure data to identify sites in the national dataset that meet the criteria established by the GIG for the selection of benchmark or reference sites.
3. Standardise the common metric (CM_bm) against the benchmark according to the approach used in the completed exercise. If benchmark standardisation was concluded not to be required in the completed exercise the mean CM value of the joining method's benchmark sites must lie *inside* the range of mean values of the benchmark sites of the methods already intercalibrated for this conclusion to remain applicable. If the joining method's benchmark sites lie *outside* of this range the joining method must benchmark standardise its sites relative to the global mean CM value of the benchmark sites included in the completed exercise. These scenarios are illustrated in Table 1 and 2.
4. Use OLS regression to establish the relationship between CM_bm (y) and the EQR of the joining method (x). A specialist case is that when a joining method relies exclusively on the common metric developed in the completed exercise for its classification rather than devising an original method (then being more like Option 1). In such cases a regression would be meaningless as y is directly dependent on x. The goal for an MS choosing to use the CM as the basis for their method is simple – after any benchmarking their boundaries must simply lie within one quarter of class of the global mean view.

5. Predict the position of the national class boundaries (MP, GM, HG and reference) on the CM_bm scale.
6. Apply the comparability criteria as summarised in Chapter 6.

Table 1: Scenario 1 – Benchmark standardisation used in completed exercise

	MS	No. BM sites	Mean PCM	offset	Conclusion	
Completed exercise	A	3	1.05	0.02	MSs benchmark-standardised the PCM by subtracting the offset of their benchmark sites.	
	B	7	1.01	-0.01		
	C	4	1.06	0.03		
	D	10	0.99	-0.04		
	E	6	1.04	0.01		
	Global mean		1.03			
	Range		0.99 - 1.06			
Joining methods	F	5	1.05	0.02	Joining methods benchmark-standardise the PCM by subtracting the offset from the global mean in accordance with the completed exercise.	
	G	4	0.97	-0.06		

Table 2: Scenario 2 – No benchmark standardisation used in completed exercise

	MS	No. BM sites	Mean PCM	Conclusion	
Completed exercise	A	3	1.05	MSs decided that no benchmark standardisation was required.	
	B	7	1.01		
	C	4	1.06		
	D	10	0.99		
	E	6	1.04		
	Global mean		1.03		
	Range		0.99 - 1.06		
Joining methods	F	5	1.05	BM sites lie inside range. No standardisation required.	
	G	4	0.97	BM sites lie outside range. Standardisation against Global mean required.	

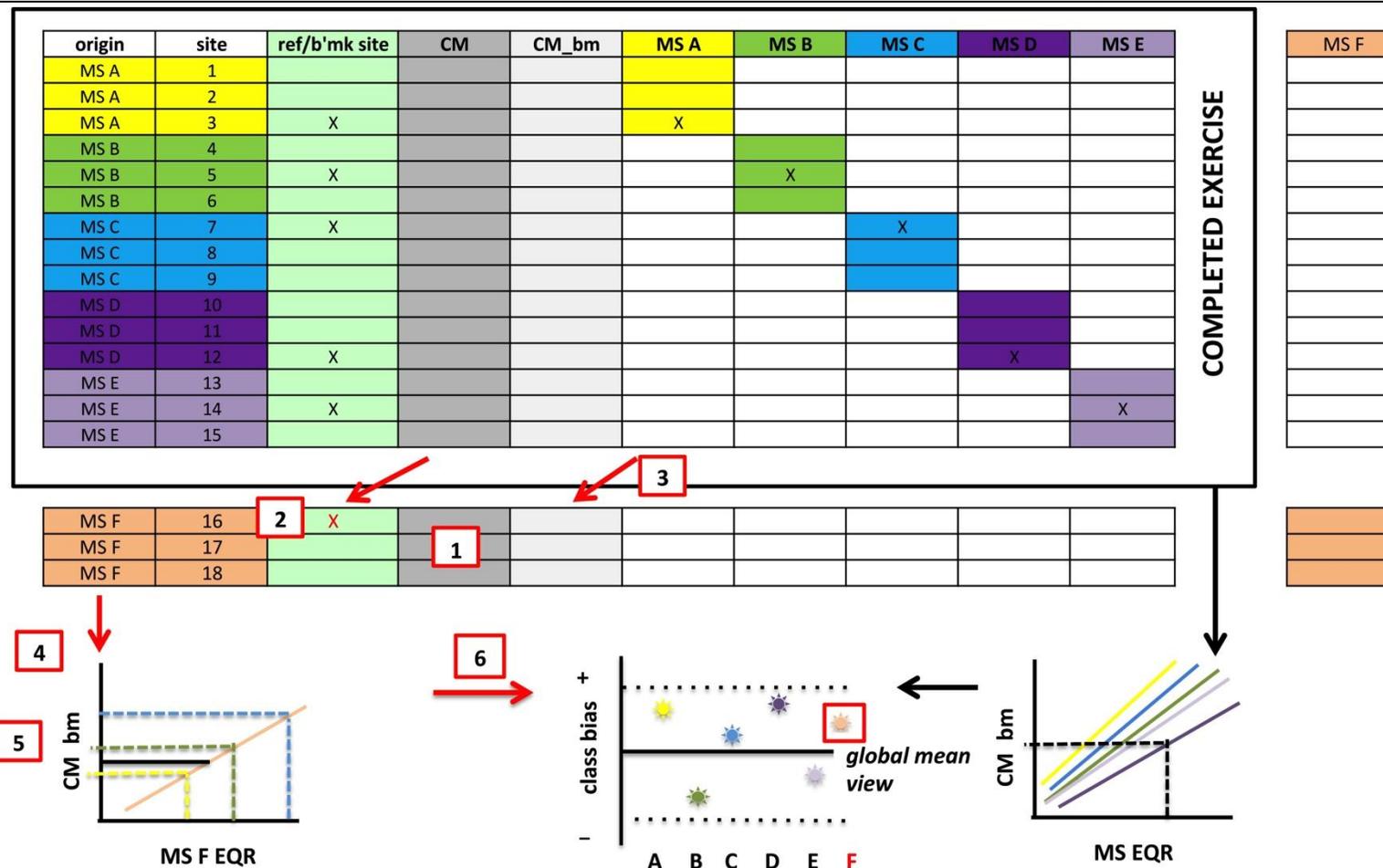


Figure 3: Case A1 - Summary of process in joining a new or revised method for Member State F (MS F). Numbered steps follow the document text in section 5.1.2.

1. Calculate common metric for data of joining country; 2. Identify reference or benchmark sites meeting pressure criteria from the completed exercise; 3. Calculate CM_{bm}; 4. Establish the relationship between CM_{bm} and MS F EQR using OLS regression; 5. Predict the position of the class boundaries of the joining method on the CM_{bm} scale; 6. Calculate the class bias of the joining method for HG and GM boundaries relative to the *global mean view* on the CM_{bm} scale already established for the completed exercise (see CIS Guidance document). Note that the *global mean view* is not altered by the inclusion of new or revised methods.

5.2 Case A2: IC Option 1 or 2 using *continuous benchmarking*

This approach should be applied when a common biological metric is available from the completed exercise, but the range of the pressure gradient available in different Member States overlaps insufficiently, thus preventing the use of reference or alternative benchmarks. This is more complex than Case A1 because the global relationship between the common metric and the pressure from the completed exercise must be used as the basis for benchmarking rather than simply using new sites that fall within a predefined window of pressure.

5.2.1 Requirements

- Full details of the common metric (e.g. species scores and metric weights).
- A suitable *site x biology* dataset from the Member State covering a range of environmental quality from which the national EQR (joining method) and the common metric can be calculated.
- Accompanying pressure data in the same format as those used in the completed exercise (i.e. same parameters measured using similar methods and at the same frequency and spatial scale).
- The OLS regression equation of the global model used to relate the common metric to pressure in the completed exercise. This information is needed even if the completed exercise concluded that benchmark standardisation was not necessary (i.e. in a model relating the CM to pressure the effect of 'country' was non-significant).
- Values of the *global mean view* of the HG and GM boundaries on the common metric scale from the completed exercise.

5.2.2 Process

1. Calculate the value of the common metric (CM_obs) for sites in the national dataset.
2. Using the global relationship between the common metric and pressure established in the completed exercise, calculate the *expected* values of the common metric (CM_pred) for the joining method's national dataset from its associated pressure data.
3. Use OLS regression to define the relationship between CM_pred (y) and CM_obs (x). From this relationship create CM_bm by projecting CM_obs onto CM_pred. This will eliminate any systematic bias in CM_obs relative to CM_pred.

An alternative is to calculate the mean residual between (CM_pred - CM_obs) and then create $CM_{bm} = CM_{obs} + \text{residual}$.

4. Use OLS regression to establish the relationship between CM_{bm} (y) and the joining national EQR (x).
5. Predict the position of the national class boundaries (MP, GM, HG and ref) on the CM_{bm} scale.
6. Apply the comparability criteria as summarised in Chapter 6.

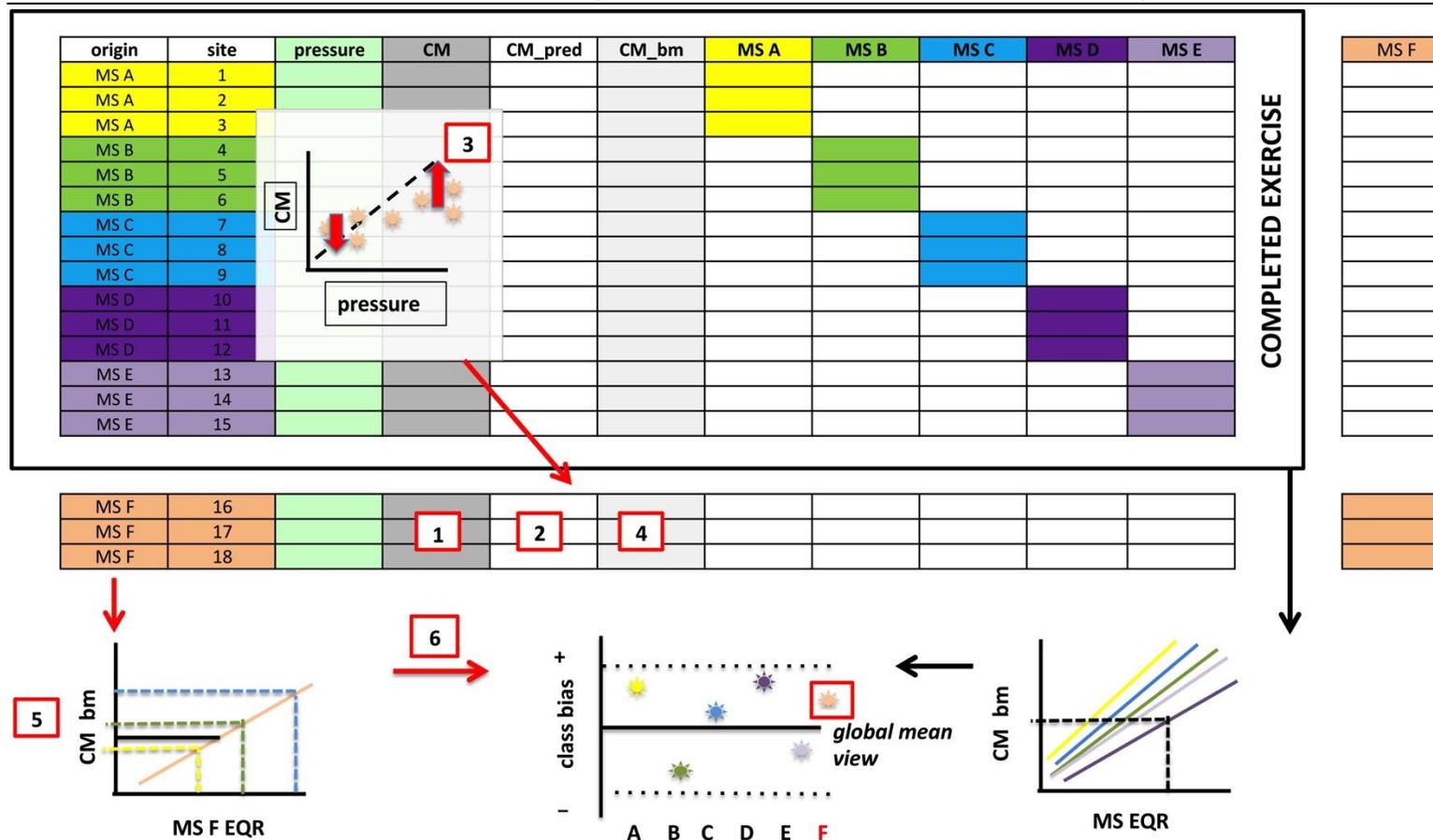


Figure 4: Case A2 - Summary of process in joining a new or revised method for Member State F (MS F). Numbered steps follow the document text in section 5.2.2.

1. Calculate common metric for data of the joining method (MS F);
2. Use the global relationship between the common metric and pressure established in the completed exercise to predict the values of the common metric (CM_pred) for the joining method using its pressure data;
3. Use OLS regression to define the relationship between CM_pred (y) and CM_obs (x);
4. Create CM_bm by projecting CM_obs onto CM_pred;
5. Use OLS regression to establish the relationship between CM_bm (y) and joining method (MS F) (x) and predict the position of the national class boundaries (MP, GM, HG and ref) on the CM_bm scale;
6. Calculate the class bias of the joining method for HG and GM boundaries relative to the *global mean view* on the CM_bm scale already established by the completed exercise (see CIS Guidance). Note that the *global mean view* is not altered by the inclusion of new or revised methods.

5.3 Case B1: IC Option 3 using *reference/benchmark sites*

This is the simplest case to apply when a common biological metric was not used in the completed exercise, and when there is likely to be a sufficient number of reference or other high quality sites available at a national level (i.e. a minimum of three sites and preferably no fewer than the smallest number provided by any MS in the completed exercise). However, it is more complex than Cases A1/2 because the common metric to be used needs to be identified and replaces the use of a pseudo-common metric (PCM) previously used in IC Option 3. To simplify the process for joining methods Case B relies instead on the use of an already intercalibrated method as the common metric (i.e. BRINC = Best-Related and Intercalibrated National Classification) rather than requiring the calculation of all the intercalibrated methods on the data of the joining country, as would be required for the calculation of the PCM. Case B also deviates from the original principles of IC Option 3 because the intercalibration is now undertaken entirely on the basis of national data rather than via a common international dataset. The distinction between Cases B1 and B2 is based on the method of benchmark standardisation with B2 requiring continuous benchmarking.

5.3.1 Requirements

- Full details of the national methods of already intercalibrated Member States in the relevant GIG (e.g. species scores and metric weights).
- A suitable *site x biology* dataset covering a range of national quality from which the national EQR and the intercalibrated methods of other Member States can be calculated.
- Accompanying pressure data in the same format as that used in the original exercise (i.e. same parameters measured using similar methods and at the same frequency and spatial scale). If the completed exercise concluded that benchmarking was not necessary the mean value of the benchmark sites from each country on the BRINC scale must be provided so that the joining Member State can also judge the need to benchmark its assessment method.
- Details of exactly how the benchmarking was undertaken in the completed exercise (e.g. by applying each method to a common international dataset and dividing the national EQR by the median value of the reference or benchmark sites of each country).
- All coefficients of the linear regression model relating the PCM to the BRINC.
- Values of the *global mean view* of the HG and GM boundaries on the PCM scale from the completed exercise.

5.3.2 Process

1. Consult the detail of the national methods and select the method that is most similar to the joining method in terms of concept, scores of indicator species to calculate assessment index etc..

The BRINC should be highly correlated² with the joining method and forms the “common metric” for the specific purposes of intercalibrating the joining method. If there is little to choose between any of the already intercalibrated methods, use the method which had the strongest correlation with the PCM in the completed exercise.

2. Identify which sites in the national dataset meet the criteria established by the GIG for selection of benchmark or reference sites.
3. Benchmark-standardise the common metric (=BRINC_bm) according to the approach used in the completed exercise. If benchmark-standardisation was not required in the completed exercise, the joining Member State must show that the mean BRINC value of its own benchmark sites lies *inside* the range of the mean BRINC values of the benchmark sites of the methods already intercalibrated. In the event that the mean BRINC value of the benchmark sites of the joining method lies *outside* this range the joining method must benchmark standardise the BRINC relative to the global average of the mean values of the benchmark sites of the already intercalibrated methods. The tables in section 5.1 with respect to the CM clarify this information and are interchangeable with the use of a BRINC.
4. Use OLS regression to establish the relationship between the BRINC_bm (y) and the national EQR (x) of the joining method.
5. Predict the position of the national class boundaries (MP, GM, HG and reference) on the BRINC_bm scale.
6. Determine the position of the global view on the BRINC_bm scale using the relationship between the PCM and the relevant benchmark-standardised BRINC established in the completed exercise. The value for the PCM is the *global mean view* of each upper class boundary established in the completed exercise (it is not simply the position on the PCM scale of the class boundaries of the BRINC, which already carries a degree of bias). This relationship will need to be inverted (as it was originally used to predict the PCM from the benchmark-standardised national

² European Commission (2011) defined the minimum level for judging acceptability of comparisons based on related-ness between methods as r (i.e. correlation coefficient) ≥ 0.5 .

method, now nominated as the BRINC_bm). The aim is now to find the value on the benchmark-standardised national metric (BRINC_bm) that equates to the known (i.e. global mean) PCM value.

7. Having derived the value of the benchmark-standardised national metric (BRINC_bm) that equates to the global mean PCM for the HG and GM boundaries, apply the comparability criteria as summarised in Chapter 6.

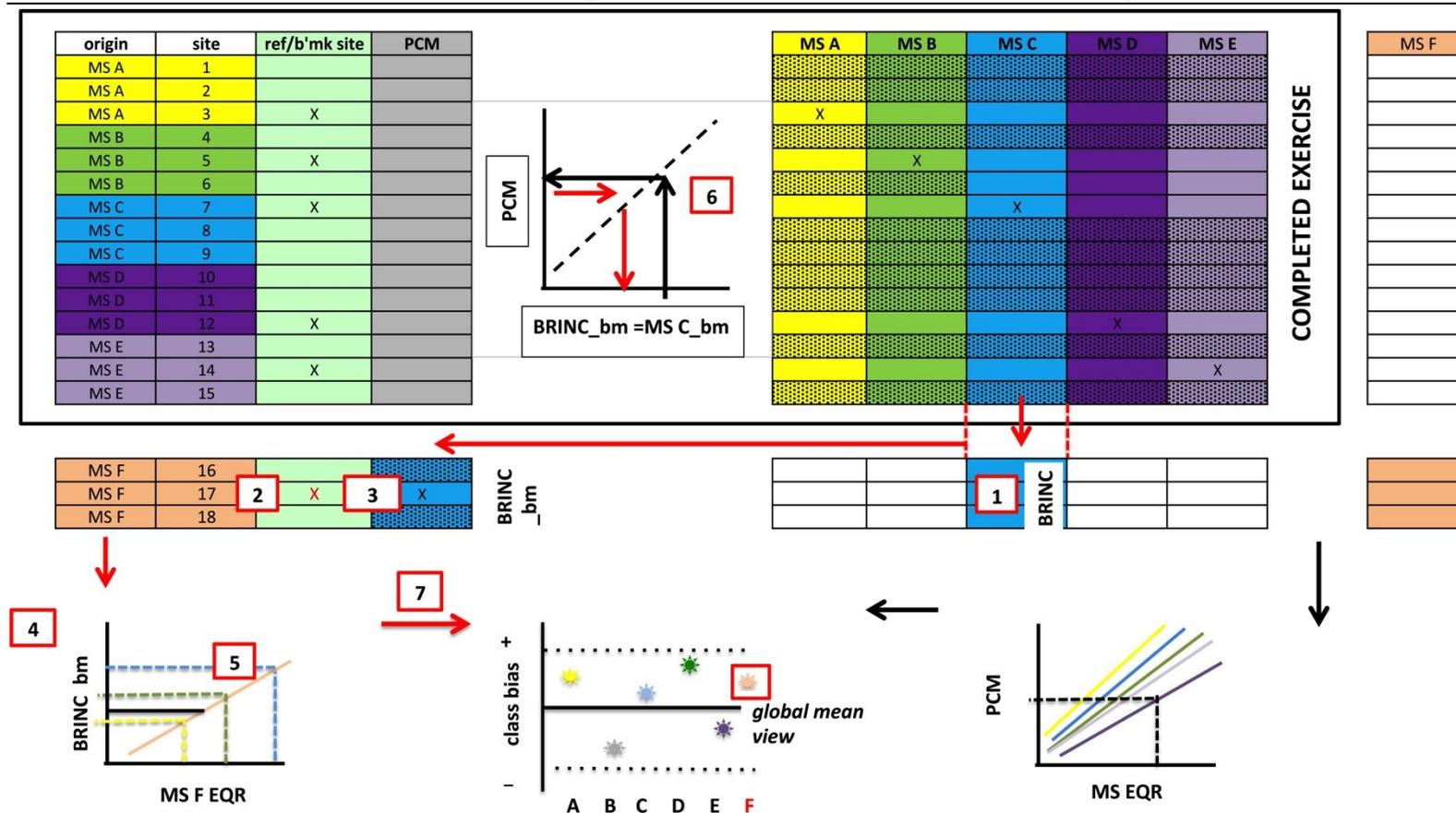


Figure 5: Case B1 - Summary of process in joining a new or revised method for Member States F (MS F). Numbered steps follow the document text in section 5.3.2.

1. Determine which of the already intercalibrated national methods will form the common metric (=BRINC). Calculate this common metric for data of joining method (MS F); 2. Identify reference or benchmark sites meeting criteria from completed exercise; 3. Benchmark-standardise the common metric (=BRINC_b m) according to the approach already used in the completed exercise; 4. Use OLS regression to establish the relationship between BRINC_b m (y) and joining method (MF F) (x); 5. Predict the position of the national class boundaries (MP, GM, HG and ref) on the BRINC_b m scale; 6. By inverting the established relationship between the PCM and the benchmark-standardised version of the national method that forms the BRINC determine the position on the BRINC_b m scale equivalent to the *global mean view* on the PCM scale; 7. Calculate the class bias of the joining method for HG and GM boundaries relative to the *global mean view* on the BRINC_b m scale. Note that this *global mean view* is not altered by the inclusion of new or revised methods.

5.4 Case B2: IC Option 3 using *continuous benchmarking*

This is the most complex case to deal with. It must be followed when a common biological metric was not used in the completed exercise and when reference or alternative benchmark sites are not commonly available at a national level. In terms of continuous benchmarking the approach deviates from that used in the completed exercise because the intercalibration is undertaken entirely on the basis of national data rather than via a common international dataset. The link to the common dataset is via the model derived from this dataset linking the common metric to the pressure(s).

5.4.1 Requirements

- Full details of all the national methods already intercalibrated in the relevant GIG (e.g. species scores and metric weights).
- A suitable national *site x biology* dataset covering a range of environmental quality from which the national EQR and the methods of other Member States can be calculated.
- Accompanying pressure data in the same format as that used in the completed exercise (i.e. same parameters measured using similar methods and at the same frequency and spatial scale).
- The OLS regression equation used to relate the BRINC to pressure in the common dataset used in the completed exercise. This information is needed even if the completed exercise concluded that benchmark standardisation was not necessary (i.e. the country effect was not statistically significant in a model of BRINC versus pressure).
- The coefficients (i.e. constants and multipliers) of the model relating the PCM to the BRINC.
- Values of the *global mean view* of the HG and GM boundaries on the PCM scale from the completed exercise.

5.4.2 Process

1. Consult the detail of the national methods and use the method that is most similar to the joining method in terms of concept, scores of indicator species to calculate assessment index etc..

This already intercalibrated method (i.e. BRINC = Best-Related and Intercalibrated National Classification) should be highly correlated³ with the joining method and

³ see footnote 1

forms the “common metric” (BRINC_obs) for the specific purposes of intercalibrating the joining method. If there is little to choose between any of the methods intercalibrated already, use the intercalibrated method having the strongest correlation with the PCM in the completed exercise.

2. Using the global relationship between the BRINC and pressure established in the completed exercise, predict the values of the BRINC (BRINC_pred) for the joining method’s national dataset from its associated pressure data.
3. Use OLS regression to define the relationship between BRINC_pred (y) and BRINC_obs (x). From this relationship create the benchmarked version of the BRINC, i.e. BRINC_bm, by projecting BRINC_obs onto BRINC_pred. This will eliminate any systematic bias in BRINC_obs relative to BRINC_pred. An alternative is to calculate the mean residual between (BRINC_pred - BRINC_obs) and then create $BRINC_bm = BRINC_obs + \text{residual}$.
4. Use OLS regression to establish the relationship between BRINC_bm (y) and the joining national EQR (x).
5. Predict the position of the national class boundaries (MP, GM, HG and ref) on the BRINC_bm scale.
6. Determine the position of the *global view* on the BRINC_bm scale using the relationship between the PCM and the relevant benchmark-standardised BRINC established in the completed exercise. The value for the PCM is the *global mean view* of each upper class boundary established in the completed exercise (note that it is not simply the position on the PCM scale of the class boundaries of the BRINC). This relationship will need to be inverted (as it was originally used to predict the PCM from the benchmark-standardised national method, now called the BRINC_bm). The aim is now to find the value on the benchmark-standardised national metric that equates to the known (i.e. global mean) PCM value.
7. Having derived the value of the benchmark-standardised national metric (BRINC_bm) that equates to the global mean PCM for the HG and GM boundaries, apply the comparability criteria as summarised in Chapter 6.

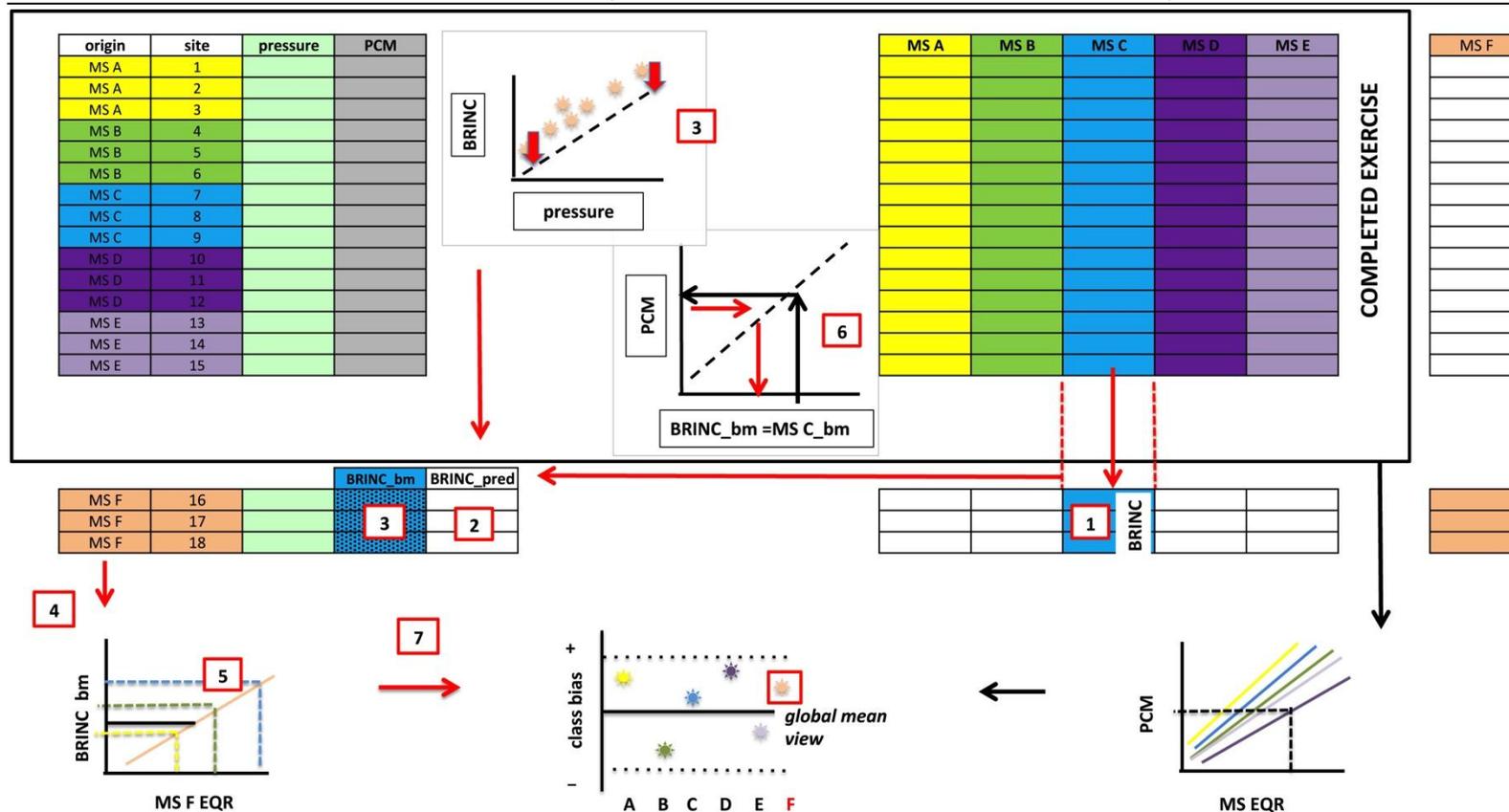


Figure 6: Case B2 - Summary of process in joining a new or revised method for Member States F (MS F). Numbered steps follow the document text in section 5.4.2.

1. Select which of the already intercalibrated methods is going to act as the common metric (=BRINC). Calculate this common metric for data of joining country (BRINC_obs); 2. Use the global relationship between the BRINC and pressure established in the completed exercise to predict the values of the common metric (BRINC_pred) for the joining method using its pressure data; 3. Use OLS regression to define the relationship between BRINC_pred (y) and BRINC_obs and then create BRINC_bm by projecting BRINC_obs onto BRINC_pred; 4. Use OLS regression to establish the relationship between BRINC_bm (y) and MS F EQR (x). 5. Predict the position of the national class boundaries (MP, GM, HG and ref) on the BRINC_bm scale; 6. By inverting the established relationship between the PCM and the benchmark-standardised version of the national method that forms the BRINC determine the position on the BRINC_bm scale equivalent to the *global mean view* on the PCM scale; 7. Calculate the class bias of the joining method for HG and GM boundaries relative to the *global mean view* on the BRINC_bm scale. Note that this *global mean view* is not altered by the inclusion of new or revised methods.

6 Comparability criteria

1. Determine the direction of deviation of the national HG and GM boundaries of the joining method on the common metric scale relative to the *global mean view* defined in the completed exercise⁴.
2. If the national GM boundary on the common metric scale falls below the global view, calculate the amount of this deviation and express it as a proportion of the width of the good status class on the common metric scale. If this value is ≤ 0.25 the boundary meets the comparability criteria. If > 0.25 , the GM boundary must be raised until the deviation between the national GM boundary on the common metric scale and the global view on the same scale is ≤ 0.25 class widths.
3. If the national GM boundary on the common metric scale falls above the global view, calculate the amount of this deviation and express it as a proportion of the width of the moderate status class on the common metric scale. If this value is ≤ 0.25 the boundary meets the comparability criteria. If > 0.25 the GM boundary can be lowered until the deviation between the national GM boundary on the common metric scale and the global view on the same scale is ≤ 0.25 class widths. However, there is no obligation to make this adjustment. If the deviation is equivalent to > 0.5 of the moderate class width, an adjustment is strongly recommended since this implies that the global view of the GM boundary of the countries that completed the exercise is closer to the MP boundary of the joining Member State.
4. These steps should then be repeated for the HG boundary. Thus, if the national HG boundary on the common metric scale falls below the global view, calculate the amount of this deviation and express it as a proportion of the width of the high status class on the common metric scale. If this value is ≤ 0.25 , the boundary meets the comparability criteria. If > 0.25 the HG boundary must be raised until the deviation between the national HG boundary on the common metric scale and the global view on the same scale is ≤ 0.25 class widths. If the national HG boundary on the common metric scale falls above the global view, calculate the amount of this deviation and express it as a proportion of the width of the good status class on the common metric scale. If this value is ≤ 0.25 , the boundary meets the comparability criteria. If > 0.25 , the HG boundary can be lowered until the deviation between the national HG boundary on the common metric scale and the global view on the same scale is ≤ 0.25 class widths. However, there is no obligation to make this

⁴ The approach of the Eastern Continental rivers' GIG to define the comparability criterion of ≤ 0.25 class width using the *global mean view* instead of the individual class widths defined by each national method is equally valid.

adjustment. If the deviation is equivalent to >0.5 of the good class width, an adjustment is strongly recommended since this implies that the global view of the HG boundary of the countries that completed the exercise is closer to the GM boundary of the joining Member State.

5. It should be noted that the class agreement element of comparability cannot (easily) be calculated for joining methods because there is no common dataset or application of the method of every country to the data of all other countries. This step, used in Option 3 of the completed intercalibration exercise, therefore has to be suspended. It is arguable that a more stringent correlation between the joining method and the common metric should be stipulated in the case of joining methods to compensate for the lack of a class agreement test, but this would introduce an inconsistency with the completed exercise. In reality, the goal for any joining method is clear and it should be simpler to achieve the necessary feasibility in these annex exercises than it was in earlier rounds of intercalibration, where Member States developed their methods largely in isolation.

7 Possible problems and solutions

This section examines various problems that we envisage may appear in the application of Cases A1 to B2 and proposes some solutions.

- i. The common metric is poorly related to the relevant pressure(s) within the data of the joining Member State, or the joining method is too poorly related to the common metric. A variety of factors may be responsible in this case but one of the commonest is that the data of the joining Member State covers an insufficient pressure gradient.

The best approach will be to enhance the data by applying the joining method to data in the common dataset belonging to another country with an already intercalibrated method and for which a more extensive quality gradient is available. A single statistical model can then be developed using national EQR and country as predictors and the common metric as the dependent variable. This solution relies on the ability to apply the joining method to data collected by another Member State.

- ii. The joining Member State cannot meet the criteria used by the existing Member States in defining either benchmark or reference sites (depending on the approach originally used). This may occur if the joining Member State has mainly high quality sites whilst the original Member States were forced to rely on an alternative

benchmark, or if most of its sites are too impacted to meet either reference or alternative benchmark criteria that all previously participating Member States could achieve.

One solution to this problem is a limited extrapolation (~10%) of the national method-pressure relationship into the pressure range that was achieved by other GIG members and was used as the basis for setting reference or benchmark conditions. This approach is only appropriate if there is a significant pressure-response relationship over the range of pressure data currently available. The validity of any extrapolation could be investigated by applying the joining method to the biological data from another Member State for which a longer pressure gradient was available as suggested in point (i) above.

- iii. None of the previously intercalibrated methods can be satisfactorily applied to the data of the joining Member State. This applies specifically to Cases A2 and B2. In this event, it will be the duty of the joining Member State to investigate any bias associated with the full or partial use of a given method on data pertaining to that original method (e.g. if country X can only utilise metrics *a* and *b* in a three-metric system developed by country Y it must determine the bias associated with using metrics *a* and *b* only as opposed to the full method, based on data already in the common dataset provided by country Y).
- iv. A joining Member State has a biota that includes many species for which common metric scores are unavailable, thus compromising application of the common metric. This is most likely when the joining method is associated with a Member State located near the periphery of a GIG. Calculate scores for missing species based on site scores calculated from known species. Use constrained ordination to determine empirical scores for all taxa. Relate these to common metric scores. Use this relationship to predict common metric scores for additional species.

8 Literature

Birk, S., Willby, N., Kelly, M., Bonne, W., Borja, A., Poikane, S., & Van de Bund, W. (2013). Intercalibrating classifications of ecological status: Europe's quest for common management objectives for aquatic ecosystems. *Science of the Total Environment*, 454-455, 490-499.

European Commission (2013). Commission Decision of xx establishing, pursuant to Directive 2000/60/EC of the European Parliament and of the Council, the values of

the Member State monitoring system classifications as a result of the intercalibration exercise.

European Commission (2011). Guidance document on the intercalibration process 2008–2011. Guidance Document No. 14. Implementation strategy for the Water Framework Directive (2000/60/EC). Technical report-2011-045.

Van de Bund W. (ed.)(2014) Water Framework Directive River intercalibration technical report. Luxembourg: Office for Official Publications of the European Communities.

Poikane S. (ed.) (2014). Water Framework Directive intercalibration technical report. JRC scientific and technical reports. Luxembourg: Office for Official Publications of the European Communities.

