ENVIRONMENTAL SCIENCE & POLICY 44 (2014) 237–246

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/envsci

Intercalibration of aquatic ecological assessment methods in the European Union: Lessons learned and way forward

Sandra Poikane [a,*], *Nikolaos Zampoukas* [a], *Angel Borja* [b], *Susan P. Davies* [c], *Wouter van de Bund* [a], *Sebastian Birk* [d]

[a] European Commission, Joint Research Centre, Institute for Environment and Sustainability, 21027 Ispra, Italy
[b] AZTI-Tecnalia, Marine Research Division, Herrera Kaia, Portualdea s/n, 20110 Pasaia, Spain
[c] Midwest Biodiversity Institute, 4673 Northwest Parkway, Hilliard, OH 43026, USA
[d] Aquatic Ecology, Faculty of Biology, University of Duisburg-Essen, Universitätsstraße 5, 45141 Essen, Germany

ARTICLE INFO

Article history:
Received 17 December 2013
Received in revised form
28 July 2014
Accepted 8 August 2014
Available online 16 September 2014

Keywords:
Water Framework Directive
Intercalibration
Rivers
Lakes
Transitional waters
Coastal waters
Ecological status

ABSTRACT

The Water Framework Directive requires that European Union (EU) Member States ensure that their surface waters are in at least *good ecological status* by 2015 or at the latest by 2027. The *good ecological status* objective has been described and operationally defined in the Water Framework Directive. Member States develop their own ecological assessment methods but they must demonstrate that their methods and resulting classifications are comparable to other Member States across the EU. Comparability of assessment results is determined through an intercalibration exercise, the subject of this article. In 2013 The European Commission issued an updated Commission Decision on the results of intercalibration of assessment results across Europe. We present an overview of the process, discuss critical issues and good practices, and recommend approaches for a successful completion of the exercise.

1. Introduction

Aquatic biological standards, established in law, can be an effective mechanism to promote restoration and ensure the ecological sustainability of aquatic resources (Adler, 2003; Hering et al., 2010). Several countries around the world have established legislation or policies to promote the restoration and maintenance of aquatic ecosystems (CWA, 2006; ANZECC, 2000; EC, 2000). However, the effectiveness of such policy initiatives depends upon the technical clarity of ecological goal statements, and the political clarity of intent that is written into the law. The United States Clean Water Act, for example, states a long-term, national objective to ''restore and maintain the ... biological integrity of the Nation's waters'' (section 1251). However, the Act does not define the ecological components, or attributes, that constitute biological integrity. Neither does the Act recommend scientific methods to measure the condition of aquatic biota. Rather, the U.S. Clean Water Act delegates the technical implementation of the

* *Corresponding author*. Tel.: +39 0332789720
E-mail address: sandra.poikane@jrc.ec.europa.eu (S. Poikane).
http://dx.doi.org/10.1016/j.envsci.2014.08.006
1462-9011/

biological integrity objective to U.S. states. Consequently, implementation of the biological integrity objective by U.S. states is neither standardized for assessment methodology, nor for the definition of boundaries representing acceptable biological condition (Adler, 2003; Davies and Jackson, 2006; Yoder and Barbour, 2009).

Similarly, Australian and New Zealand governmental guidelines for sustainable ecological status are presented mostly as guidance and suggested protocols (ANZECC, 2000). The guidelines do not prescribe standards but rather provide general descriptions of three categories of ecosystem condition (high value, moderately disturbed, and highly disturbed) and recommend procedures for regions to negotiate the assignment of condition goals for local water bodies.

In contrast, the Water Framework Directive (WFD) of the European Union provides operational definitions for assessing ecological status, setting management objectives, and harmonising EU Member States' ecological assessment systems (EC, 2000). The WFD requires rivers, lakes, transitional (i.e., marine-freshwater interface) and coastal waters of the European Union to be in *good ecological status* in the near future. As prescribed in the WFD, determination of attainment of *good ecological status* is based on assessment of specific aquatic assemblages, termed biological quality elements (BQEs). These include phytoplankton, aquatic flora (including macrophytes, macroalgae and phytobenthos), benthic invertebrates, and fish. The classification is done by national assessment methods developed individually by the Member States along basic standards specified by the WFD such as to include specific biological features, e.g., taxonomic composition and abundance, and to express results as a ratio of the monitored to reference conditions (ecological quality ratios, EQR).

The WFD considers that *good ecological status* is reached if an assessed BQE deviates only slightly from near-natural reference conditions. Member States are responsible for assessing specified BQEs regularly in order to classify the status of their water bodies into one of the five condition classes defined in the WFD: high, good, moderate, poor or bad. The WFD *good ecological status* objective necessarily focuses particular attention on the need to define, characterize and standardize the boundary between *good* versus *moderate* status.

European environmental legislation, including the WFD, is based on the principle of subsidiarity, meaning that every Member State has to develop assessment systems for different ecosystem types, following the general ideas expressed in the WFD. As a result of this approach, a huge number of different assessment methods have been developed and adopted by Member States (Birk et al., 2012a).

To ensure comparability of ecological status boundaries and national assessment methods across Europe the WFD stipulates an intercalibration exercise. Intercalibration is performed separately for rivers, lakes, coastal and transitional waters, and the exercise is further stratified by different anthropogenic pressures, and by BQEs. Intercalibration exercises are carried out within larger geographical units termed geographical intercalibration groups (GIGs, Table 1) which consist of Member States having waters of similar bio-geophysical types (termed 'common intercalibration types'). GIGs are somewhat akin to stratification by ecoregion, as established in the United States, to adjust expectations for aquatic biota by grouping together regions influenced by similar geophysical drivers (Omernik, 1987). Intercalibration provides a mechanism to reconcile apparent errors in the *good status* boundaries of some Member States when they differ significantly from the classification boundaries of most other Member States within the same GIG. Through this process the divergent *good status* boundaries of some national assessment methods can be harmonized and, if necessary, adjusted upward or downward (EC, 2011; Birk et al., 2013).

Because intercalibration operates at the interface of science and public policy, it is not only a question of basic scientific methodology. Successful intercalibration is also central to the fair and balanced achievement of Europe's public goals for the condition of water bodies, as set forth in public law. This exercise also, necessarily, included strong political dimension (Moss, 2008), i.e. reaching agreement

Table 1 – Geographical intercalibration groups and participating countries (EC, 2013). Note that some exercises (e.g. intercalibration of very large rivers or lake phytobenthos) were carried out across groups.

GIG	Water category	Member States included
Alpine	Rivers/lakes	Austria, France, Germany, Italy, Slovenia, Spain[a]
Eastern Continental	Rivers	Austria, Bulgaria, Czech Republic, Hungary, Romania, Slovakia, Slovenia
	Lakes	Bulgaria, Hungary, Romania
Central-Baltic	Rivers	Austria, Belgium, Czech Republic, Denmark, Estonia, France, Germany, Ireland, Italy, Latvia, Lithuania, Luxemburg, Netherlands, Poland, Portugal, Spain, Sweden, United Kingdom
	Lakes	Belgium, Denmark, Estonia, France, Germany, Ireland, Latvia, Lithuania, Netherlands, Poland, United Kingdom
Mediterranean	Rivers	Cyprus, France, Greece, Italy, Portugal, Slovenia, Spain
	Lakes	Cyprus, France, Greece, Italy, Portugal, Romania, Spain
Northern	Rivers/lakes	Finland, Ireland, Norway, Sweden, United Kingdom
Baltic	Coastal and transitional waters	Denmark, Estonia, Finland, Germany, Latvia, Lithuania, Poland, Sweden
Black Sea	Coastal waters	Bulgaria, Romania
Mediterranean	Coastal and transitional waters	Cyprus, France, Greece, Italy, Slovenia, Spain
North-East Atlantic	Coastal and transitional waters	Belgium, Denmark, France, Germany, Ireland, Netherlands, Norway, Portugal, Spain, Sweden, United Kingdom

[a] Only Alpine River GIG.

among social, economic and environmental viewpoints (Daly, 1999) that could neither be avoided nor minimized, since the final objective is the restoration and management of water bodies, with all the attendant economic and environmental implications.

We present an overview of the origins of the EU intercalibration exercise, the development of its concepts, the organization of the exercise, and the results achieved thus far. We also highlight the main scientific concerns and difficulties encountered during implementation, and propose a way forward.

Because intercalibration is often perceived as an arcane and uncertain task by many involved in WFD implementation (Hering et al., 2010) we hope this contribution will improve understanding and shed practical light on the subject. In particular, we hope this presentation of the material will provide a practical guide for Member States that have not fully completed intercalibration, for countries that may join the European Union in the coming years, and for countries in other continents having similar environmental legislation. Ensuring a level playing field across political boundaries is an essential task for any country or federation that aims to establish enforceable ecological goals. Thus we hope this presentation of Europe's progress developing methods to ensure a shared understanding and quantification of *good ecological status* will be of interest beyond Europe.

2. Coordination, communication, and scientific leadership

The Common Implementation Strategy (EC, 2001) sets forth the approach of the European Union to foster uniform WFD implementation. National experts collaborate within the guidance of this framework, to generate, accumulate and consolidate experiences and views into a workable format to implement the WFD and achieve its goals. The Joint Research Centre (the European Union's in-house science service) coordinated the intercalibration effort, which involved hundreds of experts across Europe, and established more than 50 separate working groups. The achievements of these working groups are documented on the Communication and Information Resource Centre for Administrations, Businesses and Citizens of the European Commission (CIRCABC; https://circabc.europa.eu), a freely accessible public resource that includes reports, presentations, meeting agendas and minutes.

3. Evolution and refinement of intercalibration

3.1. *Early days of intercalibration*

The concept of intercalibration as described in the WFD was initially very simple. For each intercalibration type, Member States were asked to nominate two sites representing the upper (''high-good'') and lower (''good-moderate'') boundary of good ecological status according to their national assessment methods (EC, 2003). The initial expectation was for these sites to serve as anchors to characterize boundaries for WFD *good status* so that new sites could be evaluated in relation to type-specific benchmarks of status. These intercalibration sites were compiled in the intercalibration register that was published as a Commission Decision (EC, 2005a). The final register contained ca. 1500 surface water sites in 27 countries.

However, the intercalibration sites were of limited benefit for the actual intercalibration exercise both due to lack of biological data, ecological assessment methods and statistical approaches for boundary comparison (Heiskanen et al., 2004). Consequently, more sophisticated approaches were developed involving compilation of a datasets covering the whole pressure gradient (Buffagni et al., 2007; Borja et al., 2007).

3.2. *The first phase of intercalibration*

Three principal options were defined in the first intercalibration guidance to compare and harmonize the national classifications, depending on how similar were the national assessment methods (EC, 2005b):

(i) If all Member States used the same sampling and assessment method they just had to agree on common reference conditions and common class boundaries. This was the most simple, transparent and straightforward option. However, it was only possible in a few cases (Poikane et al., 2010) because most of the states used different assessment methods (Birk et al., 2012a).

(ii) If Member States had uniform sampling protocols and thus collected the same biological data (e.g. number of individuals of all species), national methods were compared directly by assessing the same sampling sites with different methods (e.g., Borja et al., 2007). The comparability was evaluated using the degree of class agreement, i.e. the EQR differences between each method and every other method for all commonly assessed sites (e.g., Borja et al., 2007).

(iii) If Member States did not have common sampling methods, the results of national assessment methods were translated into a comparable format using common metrics (e.g., Buffagni et al., 2007). In this case methods were compared using boundary bias, i.e. the deviation of class boundaries of one national method relative to the common view of all Member States participating in the exercise.

The first intercalibration phase was completed with the publication of a Commission Decision in 2008 (EC, 2008a), although limited results were reached (Table 2). Moreover, comparability was not considered well-demonstrated and adequate in all cases (Van de Bund et al., 2008).

3.3. *The second phase of intercalibration*

To overcome the shortcomings of the first phase, an updated guidance was completed in 2011 (EC, 2011). It aimed at increased comparability by implementing harmonized criteria into all three options: the most that any national boundary could deviate from the global mean view of all countries was ±0.25 classes and therefore the most widely divergent national methods could not differ from each other by more

Table 2 – Intercalibration of European assessment methods' results in the 1st (2004–2008) and 2nd (2009–2012) phases (see Table 1 for geographical areas).

	Rivers						Lakes				
	ALP	CB	EC	MED	NOR	LaR	ALP	CB	EC	MED	NOR
1st phase											
Phytoplankton											
Macrophytes	NA									NA	
Phytobenthos										NA	
Benthic fauna										NA	
Fish fauna										NA	
2nd phase											
Phytoplankton											
Macrophytes	NA									NA	
Phytobenthos										NA	
Benthic fauna										NA	
Fish										NA	

	Coastal waters				Transitional waters			
1st phase								
	BAL	BLACK	MED	NEA	BAL	BLACK	MED	NEA
Phytoplankton						NA		
Angiosperms						NA		
Macroalgae						NA		
Benthic fauna						NA		
Fish fauna	NA	NA	NA	NA				
2nd phase								
Phytoplankton						NA		
Angiosperms						NA		
Macroalgae						NA		
Benthic fauna						NA		
Fish fauna	NA	NA	NA	NA		NA		

Rivers / lakes: ALP = Alpine; CB = Central Baltic; EC = Eastern continental; MED= Mediterranean; NOR = Northern; LaR = Large rivers
Coastal / transitional waters: BAL= Baltic Sea; Black = Black Sea; MED= Mediterranean Sea; NEA = North Eastern Atlantic

	Results with well demonstrated comparability and no or minor gaps in parameters, countries and types of coverage
	Some results, but with important gaps in parameters, countries and types of coverage or with questionable comparability
NA	Not applicable (intercalibration not required for a Biological Quality Element or not possible e.g. due to too few comparable water bodies)
	No results achieved

than 0.5 classes. Additionally, the concept of benchmark standardization was introduced to account for systematic differences among national water bodies belonging to the same, yet rather broadly defined intercalibration type. The second phase was completed with a second Commission Decision in 2013 (EC, 2013), including new and updated results (Table 2). With some gaps still remaining, a major step forward was achieved with the intercalibration of 230 methods from 28 countries.

The labour of the hundreds of scientists involved, the important progress in understanding the structure and the functioning of aquatic ecosystems and developing an impressive range of assessment methods was highlighted by several authors (e.g. Hering et al., 2010; Birk et al., 2012a, 2013).

4. Balancing WFD requirements against practical and technical concerns

The WFD (Annex V) requires that national assessment methods are BQE-specific and address specific biological parameters (e.g. taxonomic diversity, ratio of sensitive to insensitive taxa). The intercalibration exercise included checks for compliance with these criteria, but legitimate technical concerns necessitated some flexibility in weighing formal requirements against practical concerns. These judgements were made by individual intercalibration groups that discussed the scientific validity of reasonable derogations from WFD-prescribed criteria. For example, Member States in the intercalibration group for Mediterranean lagoons agreed on combined, instead of separate, assessment of macroalgae and angiosperms. For benthic invertebrates in lakes, the abundance parameter was omitted as it showed high variability and, thus, low interpretability (Sandin and Johnson, 2000). Similarly, the cost-effectiveness and ecological information of fish age structure was questioned for the assessment of Alpine lakes (Gassner et al., 2003).

However, some scientifically valid issues remain unsolved. Citing evidence of non-linear pressure–response relationships (e.g. Quintino et al., 2006) some Member States did not address taxonomic diversity in their assessment of macroinvertebrates in coastal waters although it is required by the WFD. Macrophyte abundance is considered a key indicator, especially in shallow lakes (Scheffer, 1998), yet the parameter was ignored by many national methods that instead relied solely on indicators of species composition. Similarly, most Member states for phytobenthos assessment use only diatom composition metrics, thus ignoring filamentous algae which can reach nuisance levels as a result of nutrient enrichment (Kelly, 2013).

More fundamental concern was raised regarding the use of zooplankton in lake assessment. Researchers have argued that this biological group represents a key element for understanding lake ecosystem functioning (Davidson et al., 2011) but is not included in the WFD. Jeppesen et al. (2011) thus advocated the need to develop zooplankton-based lake assessment methods. Building upon the lessons learned and best practices, all these cases indicate the need for a careful revision of the WFD requirements.

5. Defining *good ecological status* class boundaries

Clearly and transparently delineating actionable boundaries between acceptable and unacceptable ecological conditions for waters is an essential step to restore or maintain goal conditions. In the United States, except for extremely general language in the Clean Water Act, the authority to establish numeric biological criteria is delegated to the individual states and there is no penalty for states that do not establish numeric biological criteria (Adler, 2003).

In Europe, to ensure a level playing field among countries, the WFD requires that "the values of boundaries between the classes shall be established through the intercalibration exercise". In practice, most Member States entered intercalibration having already established actionable boundary values nationally, so the EU-wide intercalibration exercise focused on comparing and harmonizing those boundaries. This process raised questions: how did the Member States set their boundaries of good ecological status? Are these boundaries of any ecological relevance? Such questions have also been raised in the United States (Davies and Jackson, 2006).

Most boundaries in the EU have been defined using expert judgement and equidistant division of a continuum of impact (Birk et al., 2012a; Brucet et al., 2013). Nevertheless, using an ecological rationale in boundary setting is feasible and has been advocated and used by some EU member states, and some US states. States and the federal government of the United States collaborated to develop the biological condition gradient model with the explicit intent of bringing greater ecological transparency to decision-making concerning boundary conditions (Davies and Jackson, 2006). Some US states have based their overall approach to water resource management on the biological condition gradient with the result that goals for aquatic life condition are ecologically transparent (USEPA, 2011).

Good examples of boundary-setting from Europe include analysis of different pressure–response relationships for setting boundaries, such as changes in the species composition of phytoplankton (Ptacnik et al., 2008), phytobenthos (Kelly et al., 2008), macrophyte (Penning et al., 2008), and fish (Uriarte and Borja, 2009) along gradients of eutrophication or organic pollution. Recently, an ecosystem-based approach was proposed where status class boundaries are defined based on shifts in ecosystem functioning (Poikane et al., 2014).

The intercalibration guidance required to draft descriptions of type-specific biological communities representing high and *good status* (EC, 2011). The aim was to provide an ecologically meaningful image of the environmental objectives (Willby, 2011) to communicate the condition of aquatic resources in a more compelling way to the public. Unfortunately, only a very few intercalibration groups completed such narratives. Ecological transparency, however, can be crucial when attempting to initiate and fund restoration measures. Stakeholders, for instance, would probably be keener on funding management actions to reduce the risk of noxious cyanobacteria blooms in lakes (Poikane et al., 2014), to avoid excessive growths of filamentous algae in rivers (Kelly, 2012) or summer fish mortality in anoxic estuaries (Uriarte and Borja, 2009), than to raise an EQR from 0.57 to 0.71. The links between the ecological indicators and ecosystem services should be convincingly demonstrated and the outcomes of management have to be expressed as tangible benefits.

6. Pressure–response relationships

The intercalibration process prompted attention to a formal demonstration of pressure–response relationship (i.e. how well did national assessment methods respond to gradients of anthropogenic pressure) though the WFD did not explicitly required it (EC, 2011). Almost one third of ca. 300 assessment methods reviewed in Birk et al. (2012a) failed to demonstrate significant relationships between anthropogenic pressure and biological response, despite a large research investment in the attempt. The challenge of distinguishing between natural variability and anthropogenic pressure complicated interpretation of pressure–response results. Europe's multi-stressor environments additionally challenged the presentation of explicit relationships. However, the effort did result in some success. Several intercalibration groups were able to develop a global pressure index adapted to the pressures relevant for the GIG and BQE assessed (Aubry and Elliott, 2006; Böhmer et al., 2014; HELCOM, 2010).

7. Benchmark standardization

Due to biogeographical and typological reasons, as well as differences in data acquisition, biological data of different countries cannot be compared without concern. For instance, the fish fauna in Ireland is originally very scarce in number of species, mainly dominated by salmonids, whereas in major parts of Finland and Sweden, the species richness is higher (Olin et al., 2014). Because of different sampling methods, the number of taxa might be generally higher in a country than in others, because the sampling covers much more area per site (Böhmer et al., 2014). Therefore, intercalibration demanded "benchmark standardization" to remove intrinsic differences between the participating countries at the start of the exercise.

In the ideal case the comparison is made against extant, minimally disturbed reference conditions (Stoddard et al., 2006). Therefore, the initial concept foresaw deriving intercalibration benchmarks from sites in near-natural reference condition (Pardo et al., 2012). It soon became obvious that the scarcity of truly undisturbed sites precludes this approach for most regions and water body types in Europe.

Alternative benchmarking (Birk et al., 2012b, 2013) was meant to provide a practical solution to this problem by selecting sites with a similar level of impairment. For example, benchmark sites for Danube River were selected using the thresholds of total phosphorus $<200\,\mu g/l$ and dissolved oxygen ≥ 6 mg/l (Birk et al., 2012b). However, this approach was used only in very few cases as it was not possible to find a sufficient number of benchmark sites, especially in cases when countries with highly different level of human impact, as Estonia and Belgium, were involved in the same exercise (Böhmer et al., 2014).

Birk et al. (2013) introduced an approach termed 'continuous benchmarking' for countries with an insufficient number of reference and benchmark sites. This approach sought to standardize background differences by using data from across the full gradient of pressure, not solely from reference conditions. Prerequisite is the development of pressure–response relationships which are adjusted to a common regression curve for all data together (Birk et al., 2013). For example, in lake phytobenthos intercalibration exercise (Kelly et al., 2014) a variation between national datasets was noted but not all countries had reference sites against which national methods could be standardized. This problem was solved by continuous benchmarking: (i) pressure–response relationships between total phosphorus and common metric for each national dataset were developed and (ii) generalized linear modelling was used to control the effect of national differences in datasets. Similar approach was used for lake phytoplankton, macrophytes and benthic invertebrates (e.g., Böhmer et al., 2014).

From an analytical perspective, continuous benchmarking is certainly the preferred option for standardizing assessment methods among Member States with a poor availability of sites in undisturbed or similarly disturbed conditions. However, the added element of statistical complexity inherent in this approach further challenged a comprehensible intercalibration process. Still, this approach most effectively handles differences among biological datasets, thus rendering possible comparison and harmonization of management objectives between countries.

8. Typology

The evidence of bio-geographical differences within a type triggered additional considerations on typology. A recent overview revealed an overwhelming variety of water body types defined by the Member States: in total 2646 national types were delineated, including 1599 river types, 673 lake types, 261 coastal water types and 116 transitional water types (Nixon et al., 2012). Further, the intercalibration types were defined based on expert judgement at an early stage of the exercise, then often modified afterwards. As a result, a vague link between national types and common types hampered the translation of the intercalibration results into the national typology system. Open issues include: how many national types and water bodies are actually covered by the current results? What is the proportion of national types and water bodies not included in intercalibration so far? These aspects certainly need to be addressed in the future to establish sound links between national and common types and to identify the remaining gaps of the intercalibration exercise.

9. Methods harmonization

In cases where intercalibration revealed that some national methods had produced boundaries that were out of step with the boundaries of other national methods from the same intercalibration group, the group could apply a provision to adjust for this 'boundary bias'. In many cases more profound changes were also made by the groups, such as at the level of data acquisition or numerical evaluation. It was often unclear what these adjustments meant in terms of the ecological characteristics of the communities – in other words, ''the ecology behind the number'' was not always tangible (Willby, 2011). Moreover, because the assessment uncertainty of most national methods was not quantified, there was no way to evaluate if adjustments were justified with regard to the inherent error of the methods. Progress has been made in recent years however, regarding the quantification of uncertainty, e.g., for phytoplankton (Carvalho et al., 2013), phytobenthos (Kelly et al., 2009), macroalgae (Mascaró et al., 2013), and fishes (Borja et al., 2013b).

10. Lessons learned and the way forward

In Europe, impressive progress was achieved in developing and harmonizing ecological assessment methods (Birk et al., 2012a, 2013). However, there are still many gaps which need to be closed (Table 2). While most of the gaps are for coastal and transitional waters, gaps remain for large river assessment methods (except for phytobenthos), and river macrophyte assessment methods in the Northern region. For lakes, main gaps concern the Central Baltic fish fauna assessment methods, and all methods in the Eastern Continental region.

Despite the huge efforts invested in the intercalibration exercise in terms of time and persons involved, not all assessment methods have been harmonized yet. Several different reasons were responsible for this lack of results, as outlined below and summarized in Table 3:

(i) *Absence of adequate and comparable datasets across countries*: Interestingly, in some cases, a huge amount of data were collected but with different methods (e.g. in macroinvertebrates of transitional waters) thus complicating the identification of suitable datasets for comparison. Hence, in those cases, a selection of smaller datasets, collated under the same sampling methodology, with adequate pressure gradients, and accompanying environmental data, is needed for successful intercalibration in the future.

(ii) *Inadequate characterization of pressure gradients within the GIGs*: Sometimes, Member States had no quantitative measures of pressures and/or the methodologies used to determine pressures were not comparable. This prevented the comparison of assessment methods across those countries. Again, a selection of smaller datasets with adequate and measured pressure gradients will allow for the completion of the intercalibration.

(iii) *Inadequacy of benchmark sites or a very complicated approach for continuous benchmarking, which prevented the application in some BQEs*: Sometimes, the effort was put more on the statistical approach of the benchmarking, rather than in understanding the ecology behind the datasets, and the response of BQEs to pressures.

(iv) Common types, i.e., biogeographic types across Europe, were insufficiently standardized to determine consistent, comparable type-specific reference conditions. Each country defined national types using widely divergent criteria (e.g. for lakes see Gassner et al., 2005; Kolada et al., 2006; Zenker and Baier, 2009; Kagalou and Leonardos, 2009; Borics et al., 2014). In consequence, merging and comparing these types within the intercalibration exercise was difficult. This was a prominent issue especially for coastal and transitional waters (Ramos et al., 2012).

(v) *Difficulty in comparing methods with completely different metrics and/or assessment concepts*: For Europe the adoption of common methods could have avoided most of the obstacles and considerably limited the need for an elaborate numerical comparison, but it was rarely the case. Instead, Member States fostered a proliferation of aquatic ecological assessment methods that has also been questioned elsewhere (e.g., Borja et al., 2009). Even within Member States, several regional methods were occasionally applied for the same BQE and water body type. Although there may be reasonable justifications, e.g. different biogeographical conditions and human pressures (Birk et al., 2012a), also other reasons came into play, e.g., unwillingness to change the practices developed prior to the WFD (Moss, 2008; Kelly, 2013). In the future, we recommend that Member States lacking a national classification scheme should be encouraged to adopt already established, well-tested and agreed assessment methods (Kelly et al., 2014). Methods already intercalibrated, and especially those defining their boundaries based on an ecological rationale, should be given primary consideration. Also Intercalibration Common Metrics representing basic, robust ''off-the-shelf'' solutions with agreed boundaries may be adopted in countries lacking methods or intending to improve their existing methods (Lyche Solheim et al., 2013). Taking into account the massive monitoring and assessment efforts required by the WFD, the related environmental legislation such as the Marine Strategy Framework Directive (EC, 2008b), the Data Collection Framework for the Common Fisheries Policy (EC, 2008c), and the important budget restrictions in several countries, finding pragmatic and cost effective approaches is paramount (Zampoukas et al., 2013).

(vi) Insufficient attention has been paid to the ecological meaning of assessment methods. We would like to emphasize the need to reclaim a holistic vision of healthy ecosystems (Tett et al., 2013). In the intercalibration exercise the focus was on statistical efforts to harmonize classification outcomes between Member States. In many cases the focus on statistics has obscured and overshadowed the possibility of establishing an ecologically meaningful guiding image (Willby, 2011; Davies and Jackson, 2006) and inhibited the possibility of meaningful and persuasive communication with non-technical stakeholders and end users (Kelly, 2012). It is essential to return to communicating what is meant by healthy aquatic ecosystems, and why they are important to ecologists, stakeholders and the whole society. This will require re-visiting the results with the intent to translate them back to a higher level of ecological understanding, e.g., as has been done for the Marine Strategy Framework Directive (Borja et al., 2013a).

Table 3 – Major challenges of intercalibration of ecological assessment methods and strategies for their overcoming.

Issue	Achievements	Challenges	Strategies for overcoming
Development of national assessment method in each Member State	Ecological assessment methods tuned to local conditions, pressures and datasets	Highly diverse/incomparable methods using different metrics and assessment concepts	Application of the better-performing methods and common metrics
		Several methods based on small and restricted datasets with no pressure-response relationships	Intercalibration process to ensure methods' quality and comparability
Biogeographical/methodological differences between datasets	New approaches for benchmark standardization developed, based on reference sites, alternative benchmark sites or pressure-response relationships	Lack of reference sites or alternative benchmark sites in many countries and regions	Use of more homogenous datasets (data collected with similar methods)
		Statistically intricate process with little transparency	Continuous benchmarking using pressure-response relationships (where reference or benchmark sites not available)
Comparison and harmonisation of national assessment methods	240 methods from 28 countries intercalibrated and included in the EC Decision (2013)	Complex and demanding process	Simplification of the process, including small but representative biological datasets
	Unprecented collaboration between countries, scientists and policy makers	Difficult to retrieve the ecological meaning	Prioritizing the understanding of the ecology behind the datasets and the assessment methods
		Difficult to communicate to a wider audience	

11. Conclusions

In order to take advantage of the efforts invested in the intercalibration, a simplification of the process should be promoted. This can include: (i) small and comparable biological datasets can be used. These data do not necessarily need to cover, but need to be representative of all countries involved in intercalibration; (ii) good quantitative pressure gradients are used, again although not all countries are represented; (iii) alternative intercalibration approaches are encouraged, prioritizing the understanding of the ecology behind the datasets and the assessment methods, rather than the perfect application of statistical guidelines; and (iv) a consensus among the intercalibration countries is reached on the quality classes boundaries. However, care should be taken not to take the simplification too far, and to ensure coherence of the intercalibration exercise as a whole.

The experience of last 10 years indicates that intercalibration is a valid approach for comparing and harmonizing national assessment systems. We recommend that this approach should be more widely used, e.g., for countries forming transboundary river basins and federations, as the setting of common management objectives is a key for an effective long-term protection of water resources.

REFERENCES

Adler, R.W., 2003. The two lost books in the water quality trilogy: The elusive objectives of physical and biological integrity. Environ. Law 33, 29–78.

ANZECC, 2000. National Water Quality Management Strategy: Australian and New Zealand Guidelines for Fresh and Marine Water Quality. Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand, Canberra, Australia.

Aubry, A., Elliott, M., 2006. The use of environmental integrative indicators to assess seabed disturbance in estuaries and coasts: Application to the Humber Estuary, UK. Mar. Pollut. Bull. 53, 175–185.

Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., Solimini, A.G., van de Bund, W., Zampoukas, N., Hering, D., 2012a. Three hundred ways to assess Europe's surface waters: an almost complete overview of biological methods to implement the Water Framework Directive. Ecol. Indic. 18, 31–41.

Birk, S., van Kouwen, L., Willby, N., 2012b. Harmonising the bioassessment of large rivers in the absence of near-natural reference conditions – a case study of the Danube River. Freshw. Biol. 57, 1716–1732.

Birk, S., Willby, N., Kelly, M.G., 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. Sci. Total Environ. 454–455, 490–499.

Böhmer, J., Arbaciauskas, K., Benstead, R., Gabriels, W., Porst, G., Reeze, B., Timm, H., 2014. Water Framework Directive Intercalibration Technical Report: Central Baltic Lake Benthic Invertebrate Ecological Assessment Methods. Publications Office of the European Union, Luxembourg.

Borics, G., Lukács, B.A., Grigorszky, I., László-Nagy, Z., Tóth, G.L., Bolgovics, Á., Szabó, S., Görgényi, J., Várbíró, G., 2014. Phytoplankton-based shallow lake types in the Carpathian basin: steps towards a bottom-up typology. Fundam. Appl. Limnol. 184, 23–34.

Borja, A., Josefson, A.B., Miles, A., Muxika, I., Olsgard, F., Phillips, G., Rodríguez, J.G., Rygg, B., 2007. An approach to the intercalibration of benthic ecological status assessment in the North Atlantic ecoregion, according to the European Water Framework Directive. Mar. Pollut. Bull. 55, 42–52.

Borja, A., Miles, A., Occhipinti-Ambrogi, A., Berg, T., 2009. Current status of macroinvertebrate methods used for assessing the quality of European marine waters: implementing the Water Framework Directive. Hydrobiologia 633, 181–196.

Borja, A., Elliott, M., Andersen, J.H., Cardoso, A.C., Carstensen, J., Ferreira, J.G., Heiskanen, A.-S., Marques, J.C., Neto, J.M., Teixeira, H., Uusitalo, L., Uyarra, M.C., Zampoukas, N., 2013a. Good Environmental Status of marine ecosystems: What is it and how do we know when we have attained it? Mar. Pollut. Bull. 76, 16–27.

Borja, A., Elliott, M., Henriksen, P., Marbà, N., 2013b. Transitional and coastal waters ecological status assessment: advances and challenges resulting from implementing the European Water Framework Directive. Hydrobiologia 704, 213–229.

Brucet, S., Poikane, S., Lyche Solheim, A., Birk, S., 2013. Biological assessment of European lakes: ecological rationale and human impacts. Freshw. Biol. 58, 1106–1115.

Buffagni, A., Erba, S., Furse, M.T., 2007. A simple procedure to harmonize class boundaries of assessment systems at the pan-European scale. Environ. Sci. Policy 10, 709–724.

Carvalho, L., Poikane, S., Lyche Solheim, A., Phillips, G., Borics, G., Catalan, J., De Hoyos, C., Drakare, S., Dudley, B.J., Järvinen, M., Laplace-Treyture, C., Maileht, K., McDonald, C., Mischke, U., Moe, J., Morabito, G., Nõges, P., Nõges, T., Ott, I., Pasztaleniec, A., Skjelbred, B., Thackeray, S.J., 2013. Strength and uncertainty of lake phytoplankton metrics for assessing eutrophication impacts in lakes. Hydrobiologia 704, 127–140.

CWA, 2006. Clean Water Act of 1972 (Codified as Amended at 33 U.S.C. §§ 1251–1376) .

Davidson, T.A., Bennion, H., Jeppesen, E., Clarke, G.H., Sayer, C.D., Morley, D., Odgaard, B.V., Rasmussen, P., Rawcliffe, R., Salgado, J., Simpson, G.L., Amsinck, S.L., 2011. The role of cladocerans in tracking long-term change in shallow lake trophic status. Hydrobiologia 676, 299–315.

Daly, H.E., 1999. Ecological Economics and the Ecology of Economics. Elgar Publications, UK.

Davies, S.P., Jackson, S.K., 2006. The biological condition gradient: A descriptive model for interpreting change in aquatic ecosystems. Ecol. Appl. 16 (4) 1251–1266.

EC, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23rd October 2000 Establishing a Framework for Community Action in the Field of Water Policy. Official Journal of the European Communities, European Commission, Brussels (22 December, L 327/1).

EC, 2001. Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Strategic Document as Agreed by the Water Directors. Available at: http://ec.europa.eu/environment/water/water-framework/objectives/pdf/strategy.pdf (accessed 13.03.13).

EC, 2003. Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Guidance Document 6. Towards a Guidance on Establishment of the Intercalibration Network and on the Process of the Intercalibration Exercise. Office for Official Publications of the European Communities, Luxembourg.

EC, 2005a. Commission Decision of 17 August of 2005 on the Establishment of a Register of Sites to form the Intercalibration Network in Accordance with Directive 2000/60/EC of the European Parliament and of the Council (2005/646/EC). Official Journal of the European Communities, European Commission, Brussels (19.09.2005, L 243/1).

EC, 2005b. Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Guidance Document 14. Guidance on the Intercalibration Process 2004–2006. Office for Official publications of the European Communities, Luxembourg.

EC, 2008a. Commission Decision of 30 October 2008 Establishing, Pursuant to Directive 2000/60/EC of the European Parliament and the Council, the Values of the Member State Monitoring System Classifications as a Result of the Intercalibration Exercise 2008/915/EC. Official Journal of the European Communities L332/20. European Commission, Brussels.

EC, 2008b. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 Establishing a Framework for Community Action in the Field of Marine Environmental Policy.

EC, 2008c. Council Regulation No 199/2008 of 25 February 2008 CONCERNING the Establishment of a Community Framework for the Collection, Management and Use of Data in the Fisheries Sector and Support for Scientific Advice Regarding the Common Fisheries Policy. .

EC, 2011. Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Guidance Document 14. Guidance on the Intercalibration Process 2008–2011. Office for Official Publications of the European Communities, Luxembourg.

EC, 2013. Commission Decision of 20 September 2013 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 and Repealing Decision 2008/915/EC. Official Journal of the European Union (L 266/1-47).

Gassner, H., Tischler, G., Wanzenböck, J., 2003. Ecological integrity assessment of lakes using fish communities – suggestions of new metrics developed in two Austrian prealpine lakes. Int. Rev. Hydrobiol. 88, 635–652.

Gassner, H., Wanzenböck, J., Zick, D., Tischler, G., Pamminger-Lahnsteiner, B., 2005. Development of a fish based lake typology for natural Austrian lakes >50 ha based on the reconstructed historical fish communities. Int. Rev. Hydrobiol. 90, 422–432.

Heiskanen, A.S., van de Bund, W., Cardoso, A.C., Nõges, P., 2004. Towards good ecological status of surface waters in Europe – interpretation and harmonisation of the concept. Water Sci. Technol. 49, 169–177.

HELCOM, 2010. HOLAS Workshop on Baltic Sea Pressure Index. Available at: http://meeting.helcom.fi/c/document_library/get_file?p_l_id=16324&folderId=963302&name=DLFE-40883.pdf (accessed 25.03.13).

Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C.K., Heiskanen, A.S., Johnson, R.K., Moe, J., Pont, D., Solheim, A.L., van de Bund, W., 2010. The European Water Framework Directive at the age of 10: a critical review of the achievements with recommendations for the future. Sci. Total Environ. 408, 4007–4019.

Jeppesen, E., Nõges, P., Davidson, T.H., Haberman, J., Nõges, T., Blank, K., Lauridsen, T.L., Søndergaard, M., Sayer, C., Laugaste, R., Johansson, L.S., Bjerring, R., Amsinck, S.L., 2011. Zooplankton as indicators in lakes: a scientific-based plea for including zooplankton in the ecological quality assessment of lakes according to the European Water Framework Directive (WFD). Hydrobiologia 676, 279–297.

Kagalou, I., Leonardos, I., 2009. Typology, classification and management issues of Greek lakes: implication of the Water Framework Directive (2000/60/EC). Environ. Monit. Assess. 150, 469–484.

Kelly, M., 2012. The semiotics of slime: visual representation of phytobenthos as an aid to understanding ecological status. Freshw. Rev. 5 (2) 105–119.

Kelly, M., 2013. Data rich, information poor? Phytobenthos assessment and the Water Framework Directive. Eur. J. Phycol. 48 (4) 437–450.

Kelly, M., Juggins, S., Guthrie, R., Jamieson, J., Rippey, B., Hirst, H., Yallop, M., 2008. Assessment of ecological status in UK rivers using diatoms. Freshw. Biol. 53, 403–422.

Kelly, M., Bennion, H., Burgess, A., Ellis, J., Juggins, S., Guthrie, R., 2009. Uncertainty in ecological status assessments of lakes and rivers using diatoms. Hydrobiologia 633 (1) 5–15.

Kelly, M., Urbanič, G., Acs, E., Bertrin, V., Burgess, A., Denys, L., Kalhert, M., Karjalainen, S.M., Kennedy, B., Marchetto, A., Picinska-Faltynowicz, J., Poikane, S., Scheonfelder, J., Schoenfelder, I., Varbiro, G., 2014. Comparing aspirations: intercalibration of ecological status concepts across European lakes using littoral diatoms. Hydrobiologia 34, 125–141.

Kolada, A., Soszka, H., Cydzik, D., Gołub, M., 2006. Abiotic typology of Polish lakes. Limnologica 35, 145–150.

Mascaró, O., Alcoverro, T., Dencheva, K., Díez, I., Gorostiaga, J.M., Krause-Jensen, D., Balsby, T.J.S., Marbà, N., Muxika, I., Neto, J.M., Nikolić, V., Orfanidis, S., Pedersen, A., Pérez, M., Romero, J., 2013. Exploring the robustness of macrophyte-based classification methods to assess the ecological status of coastal and transitional ecosystems under the Water Framework Directive. Hydrobiologia 704, 279–291.

Moss, B., 2008. The Water Framework Directive: Total environment or political compromise? Sci. Total Environ. 400, 32–41.

Nixon, S., Bewes, V., Mills, D., 2012. Study of Pressures and Measures in the Major River Basin Management Plans: Comparison of Typologies. Project Report. WRc. .

Omernik, J.M., 1987. Ecoregions of the conterminous United States. Ann. Assoc. Am. Geogr. 77, 18–125.

Olin, M., Holmgren, K., Rask, M., Allen, M., Connor, L., Duguid, A., Duncan, W., Harrison, A., Hesthagen, T., Kelly, F., Kinnerbäck, A., Rosell, R., Saksgård, R., 2014. Water Framework Directive Intercalibration Technical Report: Northern Lake Fish Fauna Ecological Assessment Methods. Publications Office of the European Union, Luxembourg.

Pardo, I., Gómez-Rodríguez, C., Wasson, J.-G., Owen, R., van de Bund, W., Kelly, M.G., Bennett, C., Birk, S., Buffagni, A., Erba, S., Mengin, N., Murray-Bligh, J., Ofenböck, G., 2012. The European reference condition concept: a scientific and technical approach to identify minimally-impacted river ecosystems. Sci. Total Environ. 420, 33–42.

Penning, E., Dudley, B., Mjelde, M., Hellsten, S., Hanganu, J., Kolada, A., van den Berg, M., Maemets, H., Poikane, S., Phillips, G., 2008. Classifying aquatic macrophytes as indicators of eutrophication in European lakes. Aquat. Ecol. 42 (2) 237–251.

Poikane, S., Alves, M., Argillier, C., van den Berg, M., Buzzi, F., Hoehn, E., de Hoyos, C., Karottki, I., Laplace-Treyture, C., Solheim, A., Ortiz-Casas, J., Ott, I., Phillips, G., Pilke, A., Pádua, J., Remec-Rekar, S., Riedmüller, U., Schaumburg, J., Serrano, M., Soszka, H., Tierney, D., Urbanič, G., Wolfram, G., 2010. Defining chlorophyll-a reference conditions in European lakes. Environ. Manage. 45 (6) 1286–1298.

Poikane, S., Portielje, R., van den Berg, M., Phillips, G., Brucet, S., Carvalho, L., Ott, I., Soszka, H., Van Wichelen, J., 2014. Defining ecologically relevant water quality targets for lakes in Central Europe. J. Appl. Ecol. 51 (3) 592–602.

Ptacnik, R., Lepistö, L., Willén, E., Brettum, P., Andersen, T., Rekolainen, S., Solheim, A.L., Carvalho, L., 2008. Quantitative responses of lake phytoplankton to eutrophication in Northern Europe. Aquat. Ecol. 42 (2) 227–236.

Quintino, V., Elliott, M., Rodrigues, A.M., 2006. The derivation, performance and role of univariate and multivariate indicators of benthic change: Case studies at differing spatial scales. J. Exp. Mar. Biol. Ecol. 330 (1) 368–382.

Ramos, E., Juanes, J.A., Galván, C., Neto, J.M., Melo, R., Pedersen, A., Scanlan, C., Wilkes, R., van den Bergh, E., Blomqvist, M., Karup, H.P., Heiber, W., Reitsma, J.M., Ximenes, M.C., Silió, A., Méndez, F., González, B., 2012. Coastal waters classification based on physical attributes along the NE Atlantic region. An approach for rocky macroalgae potential distribution. Estuar. Coast. Mar. Sci. 112, 105–114.

Sandin, L., Johnson, R.K., 2000. Ecoregions and benthic macroinvertebrate assemblages of Swedish streams. J. N. Am. Benthol. Soc. 19, 462–474.

Scheffer, M., 1998. Ecology of Shallow Lakes. Chapman & Hall, London.

Stoddard, J.L., Larsen, D.P., Hawkins, C.P., Johnson, R.K., Norris, R.H., 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. Ecol. Appl. 16 (4) 1267–1276.

Tett, P., Gowen, R.J., Painting, S.J., Elliott, M., Forster, R., Mills, D.K., Bresnan, E., Capuzzo, E., Fernandes, T.F., Foden, J., Geider, R.J., Gilpin, L.C., Huxham, M., McQuatters-Gollop, A.L., Malcolm, S.J., Saux-Picart, S., Platt, T., Racault, M.F., Sathyendranath, S., van der Molen, J., Wilkinson, M., 2013. Framework for understanding marine ecosystem health. Mar. Ecol. Prog. Ser. 94, 1–27.

Uriarte, A., Borja, A., 2009. Assessing fish quality status in transitional waters, within the European Water Framework Directive: Setting boundary classes and responding to anthropogenic pressures. Estuar. Coast. Mar. Sci. 82, 214–224.

USEPA (US Environmental Protection Agency), 2011. A Primer on Using Biological Assessments to Support Water Quality Management. Office of Science and Technology, Office of Water, Washington, DC.

Yoder, C.O., Barbour, M.T., 2009. Critical technical elements of state bioassessment programs: a process to evaluate program rigor and comparability. Environ. Monit. Assess. 150, 31–42.

Van de Bund, W., Poikane, S., Romero, J.R., 2008. Comparability of the Results of the Intercalibration Exercise – Summary of Responses and Way Forward. European Commission, Document ENV-COM240108-5, Brussels, 14 pp.

Willby, N.J., 2011. From metrics to Monet: the need for an ecologically meaningful guiding image. Aquat. Conserv. 21, 601–603.

Zampoukas, N., Piha, H., Bigagli, E., Hoepffner, N., Hanke, G., Cardoso, A.C., 2013. Marine monitoring in the European Union: How to fulfill the requirements for the Marine Strategy Framework Directive in an efficient and integrated way. Mar. Policy 39, 349–351.

Zenker, A., Baier, B., 2009. Relevance of abiotic criteria used in German lake typology for macroinvertebrate fauna. Hydrobiologia 636, 379–392.