Establishing expectations for pan-European diatom based ecological status assessments

Kelly, M.G.; Gómez-Rodríguez, C.; Pardo, I.; Kahlert, M.; Almeida, S.F.P.; Bennett, C.; Bottin, M.; Delmas, F.; Rosebery, J.; Descy, J.-P.; Dörflinger, G.; Kennedy, B.; Marvan, P.; Opatrilova, L.; Pfister, P.; Schneider, S.; Vilbaste, S.

Published in:
Ecological Indicators

DOI:
10.1016/j.ecolind.2012.02.020

Publication date:
2012

Document Version
Early version, also known as pre-print

Ecological Indicators 20 (2012) 177–186

Contents lists available at SciVerse ScienceDirect

Ecological Indicators

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

Establishing expectations for pan-European diatom based ecological status assessments

Martyn G. Kelly[a,*], Carola Gómez-Rodríguez[b], Maria Kahlert[c], Salomé F.P. Almeida[d], Cathy Bennett[e], Marius Bottin[f], François Delmas[f], Jean-Pierre Descy[g], Gerald Dörflinger[h], Bryan Kennedy[i], Petr Marvan[j], Libuse Opatrilova[k], Isabel Pardo[b], Peter Pfister[l], Juliette Rosebery[f], Susanne Schneider[m], Sirje Vilbaste[n]

[a] Bowburn Consultancy, 11 Monteigne Drive, Bowburn, Durham DH6 5QB, UK
[b] Departamento de Ecología y Biología Animal, Facultad de Biología, Campus Lagoas-Marcosende, 36200 Vigo, Spain
[c] Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, PO Box 7050, SE 75007 Uppsala, Sweden
[d] GeoBioSciences, GeoTechnologies and GeoEngineering (GeoBioTec) Research Unit, Department of Biology, University of Aveiro, 3810-193 Aveiro, Portugal
[e] Scottish Environment Protection Agency, Inverdee House, Baxter Street, Torry, Aberdeen AB11 9QA, UK
[f] REQUE Research Unit, Cemagref – Bordeaux Centre, 50, Avenue de Verdun, Gazinet, France
[g] Laboratoire d'Ecologie des Eaux Douces, URBO, Département de Biologie, Université de Namur, Rue de Bruxelles 61, B-5000 Namur, Belgium
[h] Division of Hydrometry, Water Development Department, Cyprus
[i] EPA, John Moore Road, Castlebar, Co. Mayo, Ireland
[j] Limni s.r.o., Kalvodova 13, 616 00 Brno, Czech Republic
[k] Vyzkumny ustav vodohospodarsky T.G. Masaryka, v.v.i.Podbabska 30, 160 00 Praha 6, Czech Republic
[l] ARGE Limnologie, Hunoldstr. 14, A-6020 Innsbruck, Austria
[m] Norwegian Institute of Water Research, Gaustadalleen 21, 0349 Oslo, Norway
[n] Estonian University of Life Sciences (former Estonian Agricultural University), Riia 181, 51014 Tartu, Estonia

ARTICLE INFO

Article history:
Received 3 November 2011
Received in revised form 31 January 2012
Accepted 7 February 2012

Keywords:
Water framework directive
Diatoms
Phytobenthos
Rivers
Monitoring

ABSTRACT

The European Union (EU) Water Framework Directive depends, for effective implementation, on Member States (MSs) agreeing to a concept of the unimpacted "reference" state, which will then provide the "expected" value in Ecological Quality Ratio (EQR) calculations. Reference assemblages of organism groups will, in turn, vary, due to geological, hydrological, climatic, physicochemical and biological factors. Member States tackle this by establishing "types" which share common characteristics. However, for the purposes of ensuring consistent application, broad transboundary types were also established within five Geographical Intercalibration Groups (GIGs, referred to here as "regions") as part of the EU's intercalibration exercise. In this paper, we evaluate these types using river diatom assemblages and also provide reference threshold values for two common metrics used in pan-European diatom assessments. A database was assembled, representing 14 EU Member States from Ireland and Portugal in the West, to Estonia and Cyprus in the East, in order to explore biogeographical patterns in assemblages unaffected by anthropogenic pressures. Multivariate analyses were used to examine this pattern and its relationship with geographic, typological and abiotic parameters. After taxonomic harmonisation, NMDS ordination of samples indicated weak differences in assemblage composition among regions. ANOSIM analyses, in turn, indicated that MS was the best factor to group similar samples whereas alkalinity, recognised as the primary environmental variable structuring diatom communities, although significant, explained less variability in the dataset. This, we believe, reflects the importance of methodological factors other than taxonomy (e.g. counting protocol, sample season) that may be constant within a MS but which vary between MSs. When two diatom metrics, the TI and IPS, were applied to the data, differences in the distribution of the metric scores between MS were generally not statistically significant even though some differences between regions were apparent. A trend of increasing values of TI (decreasing values of IPS) was observed in the sequence: Nordic < Alpine < Mediterranean < Central-Baltic < Eastern Continental regions. Additionally, some differences were observed among types within the Mediterranean and Nordic regions, though not for other regions. The data used in this exercise provides us with a region and, in some cases, a type specific benchmark dataset against which national reference data can be compared.

* Corresponding author.
E-mail address: MGKelly@bowburn-consultancy.co.uk (M.G. Kelly).

1470-160X/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ecolind.2012.02.020

1. Introduction

The principle underlying the Water Framework Directive (WFD: European Union, 2000) is that all water bodies, regardless of type or location, throughout the EU should be managed in a consistent and transparent manner, with common understanding of how much deviation from the natural state was permissible before preventative or restorative action was needed.

Although member states (MSs) were each expected to develop national methods to evaluate "ecological status", the WFD requires that methods are expressed as Ecological Quality Ratio (EQR), defined as the "observed" (O) value of a candidate metric, divided by that value "expected" (E) at reference condition (O/E). Consequently, it is important that all MSs have a common understanding of the biological state at "reference condition", as it is from this that the expected value is derived. In practice, even if all other variables were held constant, the expected value will vary depending upon the type of water body under consideration. For this reason, the EU established five Geographical Intercalibration Groups (GIGs; referred to in this paper as "regions") and, within each region, differentiated a number of water body types ("types") in order to ensure like-with-like comparisons of biological communities.

The WFD specifies the use of abiotic factors such as altitude, catchment size and geology to define types within MS and these same principles were applied to the development of intercalibration types (ECOSTAT, 2004). These types were designed to be common types for all biological quality elements (BQEs) which could be easily applied by all MSs. Consequently, the types are not optimised for any particular BQE and first attempts at intercalibration indicated significant differences in values of E among MSs within a region. Whilst this may reflect genuine differences in the natural diatom assemblages, the possibility that different approaches and criteria had been used to select reference sites could not be ruled out (Kelly et al., 2009)

In practice, this issue divides into two components: the necessity for an underlying concept of what "reference condition" actually represents and appropriate screening strategies to select samples that match this concept. The concept is part determined by the WFD, which requires MS to characterise hydromorphological and physicochemical conditions for water body types that support "high ecological status"(HES), and part by subsequent documentation from ECOSTAT (e.g. Wallin et al., 2005). This work recognised that "reference condition" was not necessarily the "pristine" state which is, across much of Europe, impossible to find due to high population densities and a long history of anthropogenic alteration of landscapes. There is, inevitably, a tension between ecological and statistical factors in deriving a viable reference concept and "best attainable condition" (*sensu* Stoddard et al., 2006) will vary from region to region. The middle way is to pursue "minimally disturbed conditions", which attempts to derive, through an iterative process, explicit criteria which establish the least amount of ambient human disturbance in a region (Stoddard et al., 2006).

The criteria for selecting sites under reference condition should be based initially and primarily on non-biological measures to avoid circularity (Bailey et al., 2004; Stoddard et al., 2006). Primary screening used data on land use in the catchment, supplemented by values of chemical variables associated with pressures. However, few sites survived this intensive screening process in MS datasets, leading to a risk that, for some countries, E would be based on limited data. As E represents the value of biological quality elements at reference conditions, the reference concept needs to yield sufficient data before robust inferences to be drawn. This can be overcome, to some extent, by pooling data from several MS to form a benchmark dataset. An alternative would be to use information from sites that fail the full screening but which may still be suitable as reference sites for certain combinations of biological quality elements (BQEs) and pressures (these are "partial" reference sites" *sensu* ECOSTAT, 2010). For example, hydromorphological changes which may impact the benthic invertebrate or macrophyte assemblages are unlikely to affect metrics used to assess phytobenthos response to nutrients. Sites where there are hydromorphological changes might, therefore, qualify as "partial" reference sites. The idea explored in this study is that the suitability of such sites can be evaluated by comparison with a benchmark dataset of "true" reference sites. This is a pragmatic solution balancing the statistical and conceptual needs of a reference site network within a MS.

One of the Biological Quality Elements (BQEs) that MSs are required to assess is "Macrophytes and Phytobenthos" and many MSs use diatoms as proxies for phytobenthos. Europe-wide comparisons are aided as methods for sampling and analysis of diatoms (CEN, 2003, 2004) are consistent and yield datasets that are amenable to pan-European comparisons. The first intercalibration exercise (Kelly et al., 2009) indicated variability in reference values, with consequences for the calculation of EQRs. However, due to the variable quality of reference screening in this earlier phase, coupled with the small size of datasets and the lack of a harmonised approach to taxonomy, it was unclear how much of this difference represented true differences in diatom assemblages at the reference state, or how much was methodological.

The aim of this paper is to examine the issue of reference conditions for phytobenthos, paying particular attention to the extent to which diatom assemblage composition and the metrics used to calculate the phytobenthos Intercalibration Common Metric (pICM: Kelly et al., 2009) vary within and between regions. In particular, we will evaluate whether the intercalibration typology is adequate for phytobenthos at a European scale, whether an alternative typology needs to be derived or whether comparisons should be made only at the region level. We use a dataset of "true" reference samples and the outputs will then provide biological criteria against which "partial" reference samples may be assessed. An example of the latter case would be a sample from an otherwise pristine river that is downstream of a physical barrier, e.g. a dam. As many MSs have only a limited number of reference sites, there are benefits (e.g. more statistically robust sample sizes) in being able to use such sites to compute E for national assessment systems.

2. Methods

2.1. Dataset and screening protocols

The data used in this paper was collected by EU Member States, either as part of research studies to develop monitoring tools for the WFD, or as part of routine monitoring programmes. Diatom sample collection and analysis followed CEN guidance (CEN, 2003, 2004). Biological data (diatom assemblage composition, as relative abundance – RA) and geographic location and environmental data were stored in Access 2000 databases. MSs assigned each site to the appropriate region – Alpine (ALP), Central-Baltic (CB), Eastern-Continental (EC), Mediterranean (MED) or Nordic (N) and, within each region, to the appropriate type (Table 1) although this was not always straightforward as the intercalibration typology rarely aligned with national typologies.

All sites were submitted to a screening process based on thresholds agreed in CB-GIG for artificial and intensive agricultural land use as well as water chemistry parameters, i.e. BOD_5, O_2, NH_4, NO_3, NO_2, PO_4 (Bennett et al., 2011). Rejection thresholds were applied for land use and the least stringent water-chemistry threshold for CB-GIG types was used for screening purposes. All parameters were evaluated simultaneously and all samples that did not comply with, at least, one of the thresholds were removed. 1096 samples from

Table 1
Brief summary of the intercalibration typology, described in more detail in ECOSTAT (2004).

Type	River characteristics
Nordic region	
R-N1	Small lowland catchments, siliceous geology, moderate alkalinity, clear
R-N2	Small-medium lowland catchments, siliceous geology, low alkalinity, clear
R-N3	Small-medium lowland catchments, siliceous geology, low alkalinity, humic
R-N4	Medium lowland catchments, siliceous geology, moderate alkalinity, clear
R-N5	Small mid-altitude catchments, siliceous geology, low alkalinity, clear
R-N7	Small highland catchments, siliceous geology, low alkalinity, clear
R-N9	Small-medium mid-altitude catchments, siliceous geology, low alkalinity, humic
Central-Baltic region	
R-C1	Small lowland catchments, siliceous geology, sandy substratum
R-C2	Small lowland catchments siliceous geology, rocky substratum
R-C3	Small mid-altitude catchments, siliceous geology
R-C4	Medium lowland catchments, mixed geology
R-C5	Large lowland catchments, mixed geology
R-C6	Small, lowland catchments, calcareous geology
Alpine region	
R-A1	Pre-alpine; small to medium–sized high altitude catchments, calcareous geology
R-A2	Alpine; small to medium high altitude catchments, siliceous geology
Mediterranean region	
R-M1	Small, mid altitude catchments, highly seasonal
R-M2	Medium, lowland catchments, highly seasonal
R-M3	Large, lowland catchments, highly seasonal
R-M4	Small/medium high altitude catchments, seasonal, high sediment transport
R-M5	Small catchments, temporary
Eastern Continental region	
R-E1	Carpathians: small to medium, mid-altitude catchments
R-E2	Plains: medium-sized, lowland catchments
R-E3	Plains: large and very large, lowland catchments
R-E4	Plains: medium-sized, mid-altitude catchments
R-E5	Balkans: medium-sized, mid-altitude catchments
R-E6	Danube River: middle and downstream stretches

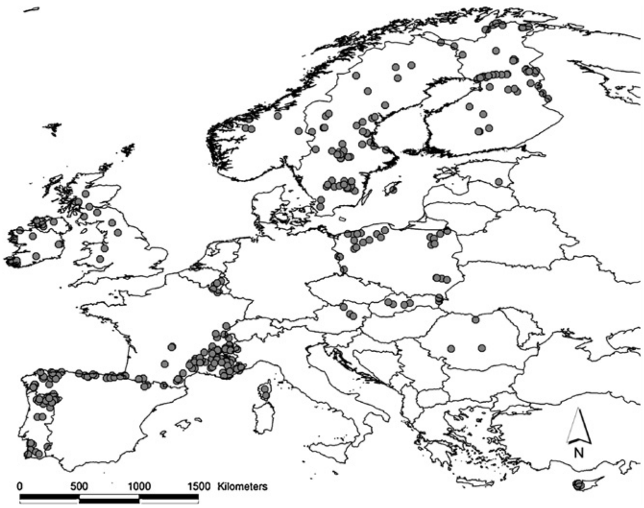

Fig. 1. Map showing location of sample sites.

533 reference sites in 14 countries distributed across all five regions passed all of these criteria (Fig. 1; Table 2).

In order to ensure that the dataset was not biased towards particular regions or stream types, a limit of 20 samples per MS per type was established; where this limit was exceeded, 20 samples were selected at random but with an override to ensure that this sampling maximised the number of sites selected (i.e. multiple samples from the same site were only allowed IF this was not at the expense of a hitherto unrepresented site). Finally, all taxa identified only to genus were removed, along with taxa with a predominately planktonic habit, in order to remove the influence of upstream impoundments. If these criteria resulted in the removal of ≥20% of the total valves, the sample was excluded from the analysis. These screening stages reduced the number of reference samples available for analyses to 409 samples.

2.2. Diatom taxonomy and harmonisation

Taxonomic harmonization took place in several stages. First, synonyms were merged, after which we applied the conventions of Kahlert et al. (in press) and Kelly and Ector (in press) and removed any taxa with a maximum RA ≤ 2% and with less than 10 records. However, even after this there were still some taxa whose differential treatment by MS may have a significant effect on the outcomes of an ordination and override the true ecological signal.

Table 2
Summary of the phytobenthos reference database by region. Values show number of sites that fulfil all screening criteria for each Member State, with the number of samples in brackets.

MS	ALP	CB	EC	MED	N	Total sites	Total samples
AT	2 (2)	2 (2)				4	4
BE_RW		5 (15)				5	15
CY				8 (13)		8	13
EE		1 (1)				1	1
ES		48 (58)				48	58
FI					55 (55)	55	55
FR	98 (189)	5 (5)		57 (118)		160	312
IE		1 (1)			15 (23)	16	24
NO					10 (10)	10	10
PL		30 (30)				30	30
PT				34 (36)		34	36
RO			11 (20)			11	20
SE		19 (19)			46 (46)	65	65
SK			47 (169)			48	169
UK		40 (199)			20 (85)	60	284
						533	1096

In order to solve this problem, a second stage of taxa harmonization used Canonical Correspondence Analysis (CCA, Ter Braak 1986, 1994, 1995; Palmer 1993; software PAST version 1.65, Hammer & Harper March 2007, Hammer et al., 2001) to analyse the dataset. Each MS was entered as a dummy categorical variable. CCA outcomes were examined for taxa that resulted in the formation of distinct clusters of samples from particular MS at the margins of the data "cloud". These taxa were then studied more closely to see if there may be reasons associated with identification conventions for their separation; if so, a pragmatic solution was applied to the dataset and the process was repeated until no obvious nomenclatural problems appeared after CCA. As a result of this, *Diadesmis gallica* and *D. pusilla* were merged, along with *Encyonopsis cesatii* var. *cesatii* and *E. cesatii* var. *geitleri, Navicula symmetrica* and *N. schroeteri, Tabellaria flocculosa* and *T. fenestrata*. Some practical identification issues remained but would have required a re-examination of the original microscope slides which was beyond the scope of this particular project.

2.3. Ordination and classification

Diatom assemblage data were converted to percentages and then square-root transformed. Non-metric Multi-Dimensional Scaling (NMDS) using Bray-Curtis dissimilarity index was used to explore the structure of the data (command: metaMDS in Vegan). Analyses of similarity ("ANOSIM": Clarke, 1993) were used to evaluate whether there were significant differences in assemblage composition among types, Member States (MSs), regions or alkalinity classes and, therefore, whether these factors could be used to group similar samples. Alkalinity measurements were divided into five approximately equally sized classes: very low (<0.1 meq/l $CaCO_3$), low (0.1–0.2 meq/l), intermediate (0.2–0.4 meq/l), high (0.4–1.0 meq/l) and very high (>1.0 meq/l). One ANOSIM analysis was conducted for each factor of interest. Additionally, independent analyses were carried out for each region, type, MSs or alkalinity classes as grouping factor. Bonferroni correction, to reduce the risk of wrongly rejecting the null hypothesis, yielded a critical probability of 0.0036 for $\alpha = 0.05$ (14 ANOSIM tests). "Type" was only used as a variable if it was represented in >1 MS. NMDS and ANOSIM analyses were performed using the R software package (R Project Core Development Team, 2005) with the Vegan package (Oksanen et al., 2007).

2.4. Calculation and analysis of metrics

A preliminary intercalibration has already been undertaken in Central-Baltic region, the largest of the regions, spanning much of lowland Europe. An intercalibration metric, pICM, was developed, based on two well-established water quality metrics, Indice de Polluosensibilité Specifique (IPS: Coste in CEMAGREF, 1982) and Trophieindex (TI: Rott et al., 1999). The pICM was then regressed against national metrics and national EQRs for High/Good and Good/Moderate status boundaries converted to equivalent values of pICM. This, then, provided a basis for comparing the position of national boundaries (Kelly et al., 2009).

The component metrics of the pICM (IPS and TI) were calculated using queries within the Access database on all screened samples (the effect of taxonomic harmonisation on the outcomes is minor: see Kelly and Ector, in press). Differences in median values of IPS and TI between regions and between types within regions were explored using the non-parametric Kruskal–Wallis test and the Mann–Whitney test. Not all sites could be reliably assigned to types by MS, so these analyses were performed on a subset of 319 samples.

2.5. Application of results to intercalibration

This exercise provides a framework for statistical comparisons of reference concepts across Europe, even if full screening information is not available. Three scenarios for how these benchmark datasets may be used to validate reference and HES data are examined. In each case, we used data supplied by MSs to:

- test the validity of "national reference" (NR) samples provided by a MS which contributed data to the benchmark dataset;
- test the validity of reference data provided by a MS which did not contribute data to the benchmark exercise; and, test the validity of high status boundary provided by a MS which does not have reference sites.

3. Results

3.1. Composition of diatom assemblages from reference samples

159 harmonised taxa were identified from the 409 reference samples. Overall, *Achnanthidium minutissimum* was the most constant and, often, most numerically abundant taxon, found in 98% of samples across all five regions, with a maximum RA of 94% (Table 3). This was followed by *Fragilaria capucina* (60%, 62.5% respectively) and *Encynonema ventricosum* (59%, 89% respectively). Overall, the differences between regions were relatively small: the most distinct of the regions was ALP, which was characterised by the absence or infrequent occurrence of acid-sensitive taxa (e.g. *Achnanthes oblongella, Eunotia* spp.) reflecting the predominately calcareous nature of these upland areas.

Achnanthidium minutissimum was also the taxon most likely to predominate in samples, forming the maximum RA of any taxon in 50% of all samples, followed by *Cocconeis placentula* (8%), *Tabellaria flocculosa* (6%) and *Achnanthidium pyrenaicum* (5%). However, there were some intra-regional differences: in ALP, *A. pyrenaicum* dominated more frequently (37%) than *A. minutisimum* (32%) whilst, in EC, *Cocconeis placentula* was the most common dominant species (23%), followed by *A. minutissimum* (15%) and *Gomphonema pumilum* (11%).

3.2. Ordination and classification

The ordination technique NMDS (stress = 20.9% with two dimensions) showed a limited amount of aggregation for some regions. ALP region is the most strongly aggregated, with low values for both axes 1 and 2 whilst EC and MED regions have overlapping distributions with low Axis 1 scores. Samples from N region tend to have high axis 1 scores with samples from CB region dispersed throughout the diagram (Fig. 2). ANOSIM analyses demonstrated significant differences in assemblage composition among regions, MSs and types (Table 4). The *R*-statistic in the regional ANOSIM was lower than that of either MS or type (Table 4). The ANOSIM testing alkalinity as a grouping factor explained less variance than region, type or MS.

3.3. Variation in metrics

A significant effect of region was also observed when values of the constituent metrics of the ICM were compared (Fig. 3; Table 5) with ALP and N regions having higher mean values of IPS and lower values for TI than the others. EC region, in particular, had some samples with low values of IPS/high values of TI.

Table 3
The thirty taxa most commonly encountered in all reference samples in the dataset along with their maximum relative abundance ("Max") and the proportion of samples in which the taxa were found ("constancy"). In most cases, taxa names refer to species as understood by Krammer and Lange-Bertalot (1986, 1997, 2000, 2004) and associated varieties; exceptions are listed in Kelly and Ector, in press or Section 2.

Taxon	All		ALP		CB		EC		MED		N	
	Max	Constancy	Max	Constancy	Max	Constancy	Max	Constancy	Max	Constancy	Max	Constancy
Achnanthidium minutissimum (Kütz.) Czarnecki	95.5	69.0	88.5	38.7	95.5	90.7	93.7	31.9	93.6	52.8	92.4	86.8
Achnanthidium pyrenaicum (Hustedt) Kobayasi	95.7	21.5	77.5	36.8	92.5	20.2	95.7	17.0	71.8	33.5	30.3	10.0
Achnanthes oblongella Oestrup	51.4	18.0			51.4	31.8			11.9	3.7	41.7	28.3
Amphora pediculus (Kützing) Grunow	60.5	21.9	23.0	18.9	22.3	33.6	21.3	21.5	60.5	24.8	1.3	4.1
Brachysira vitrea (Grunow) Ross in Hartley	31.5	14.4			8.6	14.0			0.8	2.5	31.5	39.7
Cocconeis placentula Ehrenb.	79.5	41.3	71.2	27.4	78.9	63.6	79.4	28.9	79.5	44.7	47.1	20.5
Eunotia bilunaris (Ehrenb.) F.W. Mills	32.7	16.0			32.7	17.8	1.2	0.7	1.2	5.0	19.8	38.8
Eunotia implicata Nörpel. Lange-Bertalot & Alles	36.9	15.5			10.4	15.0					36.9	44.7
Eunotia minor (Kützing) Grunow in Van Heurck	15.8	14.6			15.8	16.2	0.3	0.7	8.9	8.1	13.9	32.9
Encyonema neogracile Krammer	14.4	15.8	2.0	2.8	14.4	17.4			10.6	5.6	4.8	37.0
Encyonema ventricosum (Ag.) Grunow	89.2	41.7	89.2	34.0	36.2	56.1	29.3	20.7			9.7	43.8
Eolimna minima (Grunow) Lange-Bertalot	61.6	17.6	3.8	3.8	18.6	29.3	14.8	9.6	53.7	19.9	61.6	10.5
Fragilaria capucina Desmazieres	62.5	48.6	3.3	16.0	58.8	70.1	6.1	10.4	49.4	27.3	62.5	72.1
Fragilaria vaucheriae (Kutz.) J.B. Petersen	25.0	35.6	9.4	10.4	23.4	51.4	6.5	12.6	20.7	21.7	25.0	48.9
Gomphonema olivaceum (Hornemann) P. Dawson	27.0	14.9	9.3	26.4	19.3	19.3	18.7	16.3	4.7	6.8	27.0	7.8
Gomphonema parvulum (Kutz.) Kutz.	61.4	27.5	0.8	3.8	61.4	41.4	6.4	5.9	14.2	16.8	26.9	39.7
Gomphonema pumilum (Grun.) Reichardt & Lange-Bert	86.0	33.3	67.8	34.9	80.6	39.9	65.7	26.7	86.0	46.6	27.5	17.4
Hannaea arcus (Ehr.) Patrick	76.3	23.4	9.6	22.6	76.3	30.8	36.4	5.2	5.1	11.8	17.9	32.4
Meridion circulare (Greville) C.A. Agardh	22.6	15.5	1.5	8.5	5.4	21.8	10.9	15.6	13.9	5.6	22.6	16.9
Navicula cryptocephala Kützing	7.5	13.7	0.3	1.9	5.2	19.9	7.5	6.7	4.3	13.0	6.4	15.1
Navicula cryptotenella Lange-Bertalot	25.4	21.2	1.8	20.8	17.8	27.1	3.7	12.6	25.4	32.9	2.8	9.6
Nitzschia dissipata (Kützing) Grunow	15.0	22.6	6.5	19.8	11.4	36.1	15.0	16.3	8.7	16.8	3.8	12.3
Navicula gregaria Donkin	39.7	19.4	0.3	4.7	39.7	38.0	10.5	14.8	1.9	8.7	7.9	10.0
Navicula lanceolata (Agardh) Ehrenberg	64.3	17.2	1.5	2.8	59.0	34.0	64.3	20.7	0.9	2.5	16.3	8.2
Nitzschia palea (Kutz.) *W. Sm.*	59.5	16.5	9.2	3.8	17.9	29.0	59.5	3.0	19.5	17.4	8.5	11.9
Navicula tripunctata (O.F.Müller) Bory	40.8	14.1	3.0	11.3	18.0	20.9	40.8	22.2	5.7	13.7	4.2	0.9
Planothidium lanceolatum (Breb.) *Round et Bukh.*	42.3	16.6	2.5	11.3	9.4	21.5	42.3	21.5	10.6	14.3	3.6	10.5
Reimeria sinuata (Gregory) Kociolek & Stoermer	28.0	28.8	26.3	25.5	26.1	43.3	19.5	24.4	28.0	25.5	24.8	14.2
Synedra ulna (Nitzsch) Ehr.	70.7	25.8	2.1	9.4	30.9	35.8	70.7	16.3	2.9	11.8	18.3	35.2
Tabellaria flocculosa (Roth) Kützing	88.2	31.2			88.2	36.1	18.2	1.5	8.8	4.3	86.2	77.2
N. taxa	192		95		186			108		145		169
N. samples	942		106		321			135		161		219

Table 4
Anosim results for entire dataset of 409 reference samples. *P*-value after Bonferroni correction is 0.0036 for $\alpha = 0.05$.

Variable	*R*	Significance	Comment	Number of samples
Region	0.338	***		409
MS	0.511	***	Only MS with 5 or more samples included. AT (3 samples) and EE (1 sample) excluded (too few samples)	405
Type	0.355	***	RC1, RE4, RN4 removed (too few samples) RA2 excluded (19 samples, all in FR); RM3 excluded (7 samples, all in PT)	287
Alkalinity	0.285	***	Data missing for many samples	278

*** $p < 0.001$.

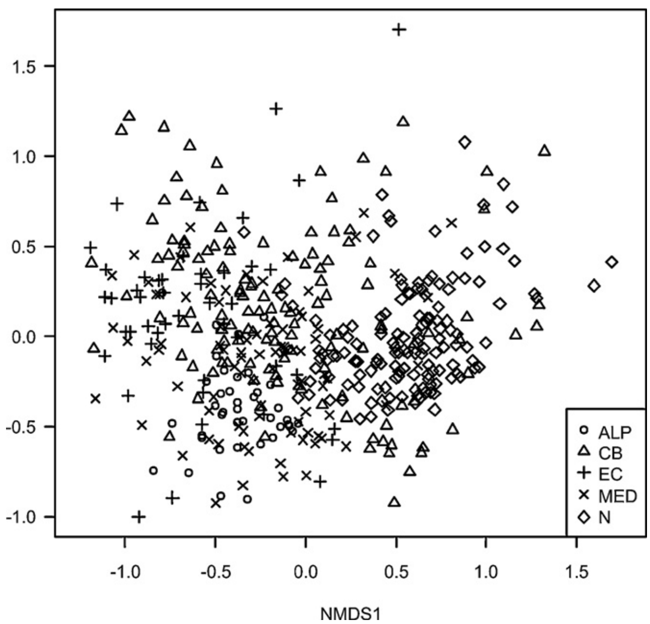

Fig. 2. NMDS plot showing the ordination of samples, classified by region, based on species relative abundance and Bray-Curtis dissimilarity.

There were no significant differences for either metric within ALP, EC and MED regions but both showed significant differences for CB and N regions. All regions, except ALP, showed a significant difference between MS for TI whilst CB and N regions also showed a significant effect for IPS. In the case of N region, lower IPS and TI values are associated with humic stream types and there may be a case for treating these separately. However, more work will be needed to ensure that this is not confounded by a MS effect.

3.4. Application of these results to intercalibration

3.4.1. Scenario 1: testing validity of NR samples provided by a MS which contributed data to the benchmark dataset

The population of reference samples available for this test was greater than that contributed to the benchmark dataset, as efforts had been made within the MS in question to expand the pool of reference samples available. These additional samples represent both extra sites and more samples from sites contributing to the benchmark dataset. MS1-GR, therefore is not wholly independent of the benchmark dataset; however it is clear from Fig. 4a and b that the two datasets have very similar values for IPS and TI. By contrast, MS1-NR is significantly different from the benchmark dataset ($p < 0.001$ for both IPS and TI). Had there been no significant difference, MS1-NR could qualify as "partial" reference sites. However, as a statistically significant difference was found, MS1 now needs to take extra steps to ensure the validity of predictions made from these samples (note that some of the MS1-NR samples represented the "best available" sites for stream types with no "true" reference sites available).

Table 5
Significance of differences in TI and IPS values between regions and between types within regions, assessed using Kruskal–Wallis test.

	Between type		Between MS		Notes
	TI	IPS	TI	IPS	
Between regions	<0.001	<0.001	<0.001	<0.001	
Within regions ALP	0.239	0.150	0.364	0.276	2 MS: mostly FR data
CB	0.001	0.002	<0.001	<0.001	6 MS
EC	0.931	0.611	0.053	0.519	2 MS
MED	0.084	0.575	0.001	0.196	3 MS
N	<0.001	<0.001	<0.001	<0.001	3 MS

3.4.2. Scenario 2: testing validity of reference data provided by a MS which did not contribute data to the benchmark exercise

Fig. 4c and d shows clear differences in IPS and TI, both of which are highly significant ($p < 0.001$) suggesting that MS2s samples do not fulfil the criteria. MS2 then reconsidered its reference data and adjusted its criteria. It could also have made a reasoned case for this discrepancy (it is possible, for example, that their national types are not represented in the benchmark dataset). Had there been no significant difference, there would have been, at the very least, reassurance that MS2s EQR values were comparable with those from other MS.

3.4.3. Scenario 3: testing validity of high status boundary provided by a MS which does not have reference sites

High ecological status (HES) samples from MS3, which set reference values by expert judgement, are also plotted in Fig. 4c and d also shows a significant deviation in IPS values ($p < 0.001$) for both metrics, suggesting that MS3s expert judgement of HES may need to be re-evaluated. The benchmark dataset provides a resource which MS3's experts may use to compare future data.

4. Discussion

The WFD embodies the EU's principle of subsidarity, leaving the details of implementation to individual MSs. The intercalibration exercise exists primarily to ensure consistent application of the ecological principles of the WFD; at the same time those involved in intercalibration need to recognise that multiple solutions to the challenges raised by the WFD are possible. By ensuring that MSs have common concepts of reference conditions and good ecological status ("GES"), intercalibration ensures consistent application of the reference concept.

The harmonised ecological status class boundaries which emerge from the intercalibration exercise are, however, legally binding on the MS. The GES boundary, in particular, has significant regulatory implications, potentially leading to higher costs for water treatment for both industry and consumers. There are, therefore, large financial consequences of this work and we need to find a balance between an ecologically defensible approach to reference condition, and one that is practical for MSs to implement.

4.1. Selection of reference sites

The approach to defining European reference conditions has evolved in the decade since the WFD was first published. The definition in the WFD itself regards the reference state as the hydromorphological and physicochemical conditions which support HES whilst, at the same time, defining HES as the biological conditions which "reflect those normally associated with that type under undisturbed conditions". Much subsequent work has involved trying to unpick this ingrained circularity, particularly as the lack of reference conditions in many areas led to subjective definitions of HES which were difficult to validate in practice. The outcome of this has been a trend, starting with Wallin et al. (2005), towards defining reference conditions using abiotic criteria alone, and then using biological assessments from sites which passed these criteria to generate "expected" values of metrics. In this paper we have identified reference sites using abiotic criteria and then define biological reference values for the intercalibration metrics following Annex III of the new IC guidance (Wallin et al., 2005; ECOSTAT, 2010). This dataset can then be used as a benchmark against which

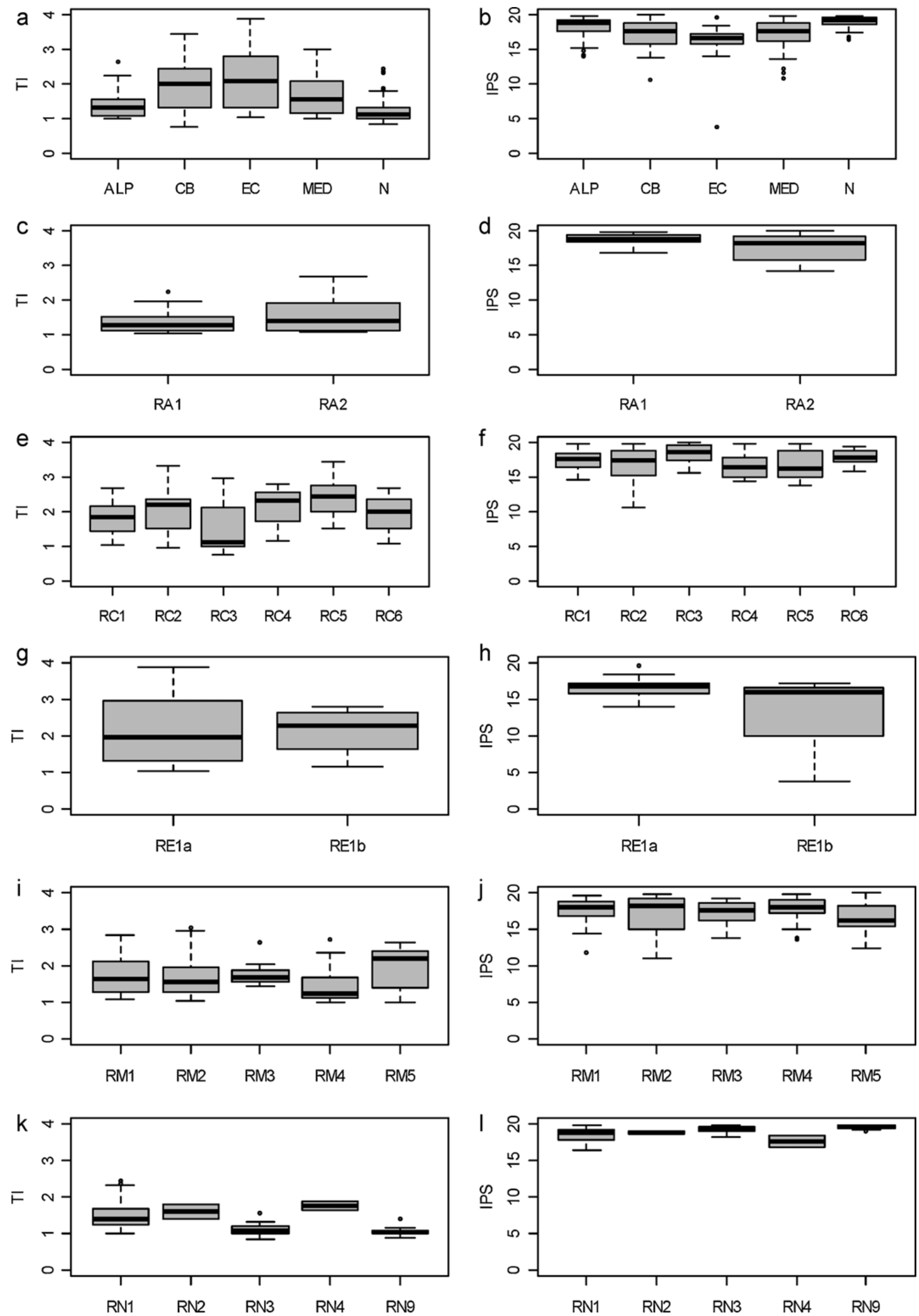

Fig. 3. Variation between values of IPS and TI of reference samples within and between the five regions. Horizontal lines show median values; boxes show 25th and 75th percentiles of the data. (a, b) Comparison between regions; (c–f) comparisons within regions: (c, d) ALP; (e, f) CB; (g, h) EC; (i, j) MED; (k, l) N.

other datasets can be evaluated, even if abiotic pressure data are not available.

However, a distinction needs to be drawn between a reference site and samples collected from that site. In this study, we noted occasional samples from apparently pristine sites which had diatom assemblages suggesting impacted conditions. However rigorous the screening, we cannot exclude all short-term incidents that may have affected the biology and we must also allow for the possibility that entirely natural events may lead to localised enrichment for a period of time. In other words, a single sample whose composition suggests impact is not, of itself, grounds for deciding that a site is not in reference condition, so long as the site fulfils the abiotic criteria. Such situations should, however, prompt a re-checking of abiotic criteria as well as additional biological sampling. We regard the abiotic screening process and the benchmark datasets which result as the starting point for debate and discussion, not as a final and unambiguous statement on the biological assemblages associated with “reference condition.”

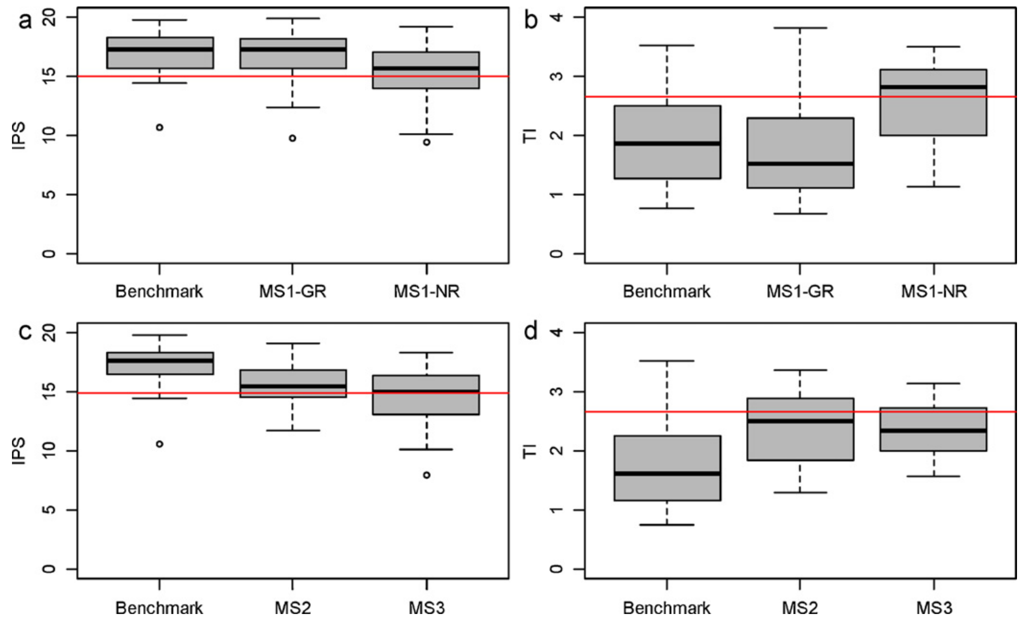

Fig. 4. Use of the benchmark dataset to validate reference and high ecological status concepts: examples from the Central-Baltic region. Benchmark; MS1-GR: subset of samples from sites which fulfil all criteria for reference conditions, MS1-NR: samples from a Member State which participated in the reference screening exercise but which also has a population of "national reference" samples; MS2: samples from a Member State which did not participate in the screening exercise; MS3: "high ecological status" samples from a Member State with no reference sites. Horizontal lines represent the 10th percentile of IPS and 90th percentile for TI. See text for more details.

4.2. Testing the intercalibration typology

The intercalibration typology was established early in the intercalibration process, based on known predictors of stream biota (ECOSTAT, 2004); however, neither this study nor earlier studies on phytobenthos (Kelly et al., 2009) indicate that it is a particularly useful approach for differentiating between different stream types. In this project it was anticipated that a combination of a larger database than was available in Kelly et al. (2009) and that a high standard of reference screening would allow a workable typology, perhaps straddling existing GIG boundaries, to emerge. This proved not to be the case: although significant differences between regions were observed using both NMDS (Table 4) and diatom metrics (Table 6), it was MS rather than type or alkalinity that emerged as the most powerful factor structuring the dataset. The lack of effect for alkalinity was particularly noteworthy, as most MSs recognise this as the key variable structuring diatom assemblages in the absence of pressure within their national assessment systems (Tison et al., 2005; Kelly et al., 2008).

It is possible that some of the compromises necessary to produce a harmonised dataset decreased the sensitivity of the multivariate analyses but alternatively, methodological differences between MS masked the expected alkalinity-driven signal. Although all MSs adhered to standard methods (CEN, 2003, 2004), these are broadly written and there is scope for interpretation to suit local circumstances. Kahlert et al. (in press) noted that harmonising taxonomy following similar conventions to those adopted in this paper had little effect on between-operator variation in a pan-European ring-test, and suggested other aspects of procedure (e.g. treatment of broken valves) may contribute systematic errors to the analytical process. In the present study, different sampling practices may also have led to differences in species composition for the samples provided by individual MS. Such practices will vary for good reasons (see Kelly et al., 1998 for a fuller discussion) but, as stream algae display distinct periodicity (e.g. Marker, 1976), adjacent MSs which adopt different sampling regimes for otherwise similar streams may have quite different diatom assemblages. The IC exercise has to work within the constraints imposed by such limitations. We did attempt to account for "season" in our analyses; however, crisp delimitation of seasons is difficult in a dataset spanning an area from 60°N to 35°S and 10°W to 37°E.

Although no case for a typology based on the diatom assemblage within regions can be made from these data, this is likely due to the nature of the intercalibration exercise rather than reflecting the true condition of phytobenthos assemblages. A key question for the present exercise, therefore, is whether or not further work would improve intercalibration of boundaries: each MS may have an internally consistent approach which allows spatial and temporal comparisons within a political unit; differences between these approaches may then have to be regarded as "noise" for the purposes of intercalibration.

Table 6
Median plus 10th (for IPS) and 90th (for TI) percentiles of metrics values recorded from reference samples for each of the regions. Values for the entire region are also given (X-GIG).

		ALP	CB	EC	MED	N	X-GIG
IPS	Median	18.8	17.7	16.7	17.7	19.3	18.3
	10th %ile	15.5	15.0	14.9	14.3	18.0	15.0
TI	Median	1.32	2.00	2.10	1.55	1.14	1.52
	90th %ile	1.96	2.66	2.71	2.52	1.69	2.56

4.3. Implications for diatom biogeography and ecology

A positive view of the results presented here is that it is possible to go to a stream almost anywhere in Europe with a fairly good idea of what diatoms we might expect to find in the absence of anthropogenic pressure. Table 3 shows how many diatom taxa are cosmopolitan across the study area and whilst some biogeographical differences are apparent (low RA of *Achnanthes oblongella* and *Eunotia* spp. coupled with relatively high RA of *Achnanthidium pyrenaicum* in ALP, for example), it is the consistency of these assemblages that is most striking. The counter view, however, is that this consistency was only obtained by merging several large species complexes and opinion amongst the authors is divided on the extent to which this represents a loss of ecological information. Many new species have been described in recent years (e.g. Krammer, 1997a,b, Lange-Bertalot, 2001; Mann et al., 2008); however, the ecological implications of this knowledge are not yet clear. There is some evidence for species within complexes having distinct ecological preferences (Poulíčková et al., 2008, for *Sellaphora pupula* ag., in lakes; Potapova and Hamilton, 2007, for *Achnanthidium minutissimum*) but two issues arise:

- Can these fine-level differences be resolved reliably by analysts (Kahlert et al., 2008, in press)? If not, additional "signal" is swamped by analytical "noise";
- Do these differences actually represent differences in ecological functioning? This decomposes "ecological status" into two components: the effect of a pressure on the phytobenthos and the consequent effect of the changes to the phytobenthos on other trophic levels. Fine-scale taxonomy might provide insights into the former but if grazers, for example, do not differentiate between closely related species of diatom, then differentiation will add nothing to our holistic view of ecological status. Development and selection of WFD tools has focussed almost entirely on the former (e.g. Hering et al., 2006) whereas greater focus on the latter would seem to offer a number of useful insights to water managers.

An interpretation of our results, bearing these points in mind is that the taxonomy of Krammer and Lange-Bertalot (1986, 1997, 2000, 2004) may be adequate to represent the main ecological gradients across Europe. Using this taxonomy, only weak biogeographic patterns emerge suggesting that there are few limits to dispersion and the main climatic drivers such as latitudinal temperature gradients are less important than methodology and local environmental factors (Bennett et al., 2010). If more refined taxonomy is used, biogeographical differences within these gradients may emerge but interpretation is complicated within and between regions by the lack of taxonomic precision. Once again, consistent taxonomy is better achievable within than between MSs and the intercalibration exercise needs to work within these constraints.

4.4. Implications for intercalibration

This project allowed a large dataset to be compiled although still only 14 of the 27 MS are represented. Although the geographical spread of these is sufficiently wide for this to be representative of the EU as a whole, we still need to interpret results with caution as the positioning of ecological status boundaries may have significant financial implications for MS. Results presented here suggest that working at regional level is most appropriate; there may be a few types with a characteristic ecological response but we believe that these should only be distinguished from the region as a whole where several MS contribute data. We have also emphasised in this paper that greater sensitivity should be achievable within MS (although, even here, data precision should not be automatically assumed).

The results also point to the advantages of using metrics such as the IPS and TI. Although potentially useful information on assemblage composition is lost or merged into broad "sensitivity" categories, such metrics help to iron-out variability due to factors such as substratum, season, position of the phytobenthos assemblage within a microsuccession, etc. The outcome, for each region, is a "benchmark dataset", integrating the diatom assemblages of all samples that survive the screening process. Nonetheless, some types of stream (e.g. very high alkalinity streams in CB region) will be under-represented as very few sites meet all screening criteria. The implication of this is that the benchmark datasets should guide the intercalibration process but not be treated as setting absolute limits.

The values in Table 6 are recommended as limits for the constituent metrics of the pICM that can be applied across a region, for the purposes of intercalibration. These values also provide a means of validating E for those MS who do not have reference values of their own (e.g. BE-FL, NL). We also recommend the median values of the benchmark datasets as interim expected values for those MS who have not yet adopted a national method, believing that the intercalibration process has provided a simple metric (pICM) that is effective across the EU plus, now, validated reference values. In combination with the sampling and analytical methods (CEN, 2003, 2004), these provide an "off-the-shelf" solution for fulfilling obligations for assessing phytobenthos. Some further research is required to explore variation within regions, to evaluate if within-region (or even cross-region) super-types are valid, and whether this improves the sensitivity of intercalibration. The principal limitation is the lack of comparable environmental data from MS. Most MSs have internally consistent environmental datasets that have enabled them to develop robust typologies or models that are more sensitive to their own conditions than values produced here and we must emphasise that this exercise validates but does not replace national reference values.

The expectation, therefore, is that metric values for a MS's reference sites should fall within the statistical limits of the benchmark dataset for the region in question. Statistical testing of the hypothesis that samples are drawn from the same population is complicated by the non-parametric distribution and interpretation is not straightforward as underlying typological factors may exert an influence which may cluster the MS's data towards one end of the benchmark dataset. Finally, although we have screened sites rigorously using abiotic criteria, there will be a few samples that suggested enrichment, either due to unknown pollution events prior to the sampling/screening exercise, or due to natural causes (death and decay of an animal carcass in the stream, for example). As we do not always have perfect knowledge of a site prior to sampling and as there may be natural processes that could potentially contribute to an episodic increase in nutrient concentration, it would be unwise to reject these samples outright. We therefore suggest that the 90th percentile for TI (10th percentile for IPS) as a pragmatic criterion for MS to use to check their own reference condition. Those NR sites that fall outside these limits will need to be re-evaluated and the expectation should be that most of the reference sites chosen by a MS not involved in this exercise will fall within the 90th (or 10th) percentile unless the MS can otherwise provide an ecological justification for the deviation. Overall, our ambition should be to minimise the role of expert judgement and "gardening" of datasets to purge sites that do not conform to a subjective "ideal".

4.5. Concluding comments

The nature of the intercalibration exercise, following a period of method development by individual MS, means that the process has little opportunity to influence collection of the data with which it has to work. Even though common standards for data collection and analysis are apparently in use, there was considerable variation due to MS, which limits the sensitivity of the analyses performed. Intercalibration is, however, the "art of the possible", recognising that the primary purpose of data collected by individual MS is to fit into ongoing water management programs. Yet, at the same time, the intercalibration exercise has presented an excellent opportunity to get a continent-wide view of the composition of diatom assemblages in near-pristine condition.

Acknowledgements

Contributors to this paper acknowledge the support of national environmental agencies for provision of data and financial support. M.G.K.: Scottish and Northern Ireland Forum for Environmental Research; M.K.: Swedish Environmental Protection Agency; S.F.P.A.: Portuguese Water Institute (INAG); L.O. and P.M.: Ministry of the Environment of the Czech Republic within the Institutional research plan MZP0002071101 – Research and Protection of the Hydrosphere; S.S.: Norwegian Directorate for Nature Management. In addition, we are grateful to the Swedish Environmental Protection Agency for support for the workshop at the Erken Field Station in Sweden and to Richard Johnson. (University of Agricultural Sciences, Uppsala, Sweden) and Eckhard Coring (EcoRing, Hardegsen, Germany) for helpful comments on a draft manuscript.

References

Bailey, R.C., Norris, R.H., Reynoldson, T.B., 2004. Bioassessment of Freshwater Ecosystems: Using the Reference Condition Approach. Kluwer Academic Publishers, New York.

Bennett, C., Owen, R., Birk, S., Buffagni, A., Erba, S., Mengin, N., Murray-Bligh, J., Ofenböck, G., Pardo, I., van de Bund, W., Wagner, F., Wasson, J.-G., 2011. Bringing European river quality into line: an exercise to intercalibrate macro-invertebrate classification methods. Hydrobiologia 667, 31–48.

Bennett, J.R., Cumming, B.F., Ginn, B.K., Smol, J.P., 2010. Broad-scale environmental response and niche conservatism in lacustrine diatom communities. Global Ecology and Biogeography 19, 724–732.

CEN, 2003. Water quality - Guidance standard for the routine sampling and pretreatment of benthic diatoms from rivers. EN 13946: 2003. Comité European de Normalisation, Geneva.

CEN, 2004. Water quality – Guidance standard for the identification, enumeration and interpretation of benthic diatom samples from running waters. EN 14407:2004. Comité European de Normalisation, Geneva.

Clarke, K.R., 1993. Non-parametric multivariate analysis of changes in community structure. Australian Journal of Ecology 18, 117–143.

CEMAGREF, 1982. Etude de Méthodes Biologiques Quantitatives d'Appreciation de la Qualité des Eaux. Rapport Q.E. Lyon-A.F.B. Rhône-Mediterrannée-Corse.

ECOSTAT, 2004. Overview of Common Intercalibration Types. Working Group 2. A Ecological Status, Final Version 5.1. Available from: http://ec.europa.eu/environment/water/water-framework/objectives/index_en.htm.

ECOSTAT, 2010. Guidance Document on the Intercalibration Process 2008–2011. 102 pp. Common Implementation Strategy for the Water Framework Directive.

Hammer, O., Harper, D.A.T., Ryan, P.D., 2001. PAST: palaeontological statistics software package for education and data analysis. Palaeontologica Electronica 4 (1), 9 pp.

Hering, D., Johnson, R.K., Kramm, S., Schmutz, S., Szoszkiewicz, K., Verdonschot, P.F.M., 2006. Assessment of European streams with diatoms, macrophytes, macroinvertebrates and fish: a comparative metric-based analysis of organism response due to stress. Freshwater Biology 51, 1757–1785.

Kahlert, M., Albert, R.-L., Anttila, E.-L., Bengtsson, R., Bigler, C., Eskola, T., Gälman, V., Gottschalk, S., Herlitz, E., Jarlman, A., Kasperoviciene, J., Kokociński, M., Luup, H., Miettinen, J., Paunksnyte, I., Piirsoo, K., Quintana, I., Raunio, J., Sandell, B., Simola, H., Sundberg, I., Vilbaste, S., Weckström, J., 2008. Harmonization is more important than experience–results of the first Nordic-Baltic diatom intercalibration exercise 2007 (stream monitoring). Journal of Applied Phycology 21, 471–482.

Kahlert, M., Kelly, M.G., Albert, R-L., Almeida, S.F.P., Bešta, T., Blanco, S., Coste, M., Denys, L., Ector, L., Fránková, M., Hlúbiková, D.,·Ivanov, P., Kennedy, B., Marvan, P., Mertens, A.,·Miettinen, J.,·Picinska-Fałtynowicz, J. Rosebery, J., Tornés, E.,·Vilbaste, S., Vogel, A. Identification is a minor source of uncertainty in diatom-based ecological status assessments on a continent-wide scale: results of a European ring-test. Hydrobiologia, in press.

Kelly, M.G., Ector, L. Effect of streamlining taxa lists on diatom-based indices: implications for intercalibrating ecological status. Hydrobiologia, in press.

Kelly, M.G., Cazaubon, A., Coring, E., Dell'Uomo, A., Ector, L., Goldsmith, B., Guasch, H., Hürlimann, J., Jarlman, A., Kawecka, B., Kwandrans, J., Laugaste, R., Lindstrøm, E.-A., Leitao, M., Marvan, P., Padisák, J., Pipp, E., Prygiel, J., Rott, E., Sabater, S., van Dam, H., Vizinet, J., 1998. Recommendations for the routine sampling of diatoms for water quality assessments in Europe. Journal of Applied Phycology 10, 215–224.

Kelly, M., Bennett, C., Coste, M., Delgado, C., Delmas, F., Denys, L., Ector, L., Fauville, C., Ferréol, M., Golub, M., Jarlman, A., Kahlert, M., Lucey, J., Ní Chatháin, B., Pardo, I., Pfister, P., Picinska-Faltynowicz, J., Rosebery, J., Schranz, C., Schaumburg, J., van Dam, H., Vilbaste, S., 2009. A comparison of national approaches to setting ecological status boundaries in phytobenthos assessment for the European Water Framework Directive: results of an intercalibration exercise. Hydrobiologia 621, 169–182.

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

Krammer, K., 1997a. Die Cymbelloiden Diatomeen Eine Monographie der weltweit bekannten Taxa. Teil 1: Allgemeines und Encyonema Part. Bibliotheca Diatomologica 36, 1–382.

Krammer, K., 1997b. Die Cymbelloiden Diatomeen Eine Monographie der weltweit bekannten Taxa. Teil 2: Encyonema part, Encyonopsis and Cymbellopsis. Bibliotheca Diatomologica 37, 1–469.

Krammer, K., Lange-Bertalot, H., 1986. Die Süßwasserflora von Mitteleuropa 2: Bacillariophyceae. Teil 1: Naviculaceae. Gustav Fischer Verlag, Stuttgart.

Krammer, K., Lange-Bertalot, H., 1997. Die Süßwasserflora von Mitteleuropa 2: Bacillariophyceae. Teil 2: Bacillariaceae, Epithemiaceae, Surirellaceae. 2te Auflage. Gustav Fischer Verlag, Stuttgart.

Krammer, K., Lange-Bertalot, H., 2000. Die Süßwasserflora von Mitteleuropa 2: Bacillariophyceae. Teil 3: Centrales, Fragilariaceae, Eunotiaceae. 2te Auflage. Gustav Fischer Verlag, Stuttgart.

Krammer, K., Lange-Bertalot, H., 2004. Süßwasserflora von Mitteleuropa 2: Bacillariophyceae. Teil 4: Achnanthaceae. Kritische Ergänzungen zu Achnanthes s.l., Navicula s. str., Gomphonema. Spektrum Akademischer Verlag/Gustav Fischer, Heidelberg.

Lange-Bertalot, H., 2001. Diatoms of Europe, vol. 2. Navicula sensu stricto, 10 Genera Separated from Navicula sensu lato, Frustulia A.R.G. Gantner Verlag K.G., Ruggell.

Mann, D.G., Thomas, S.J., Evans, K.M., 2008. Revision of the diatom genus *Sellaphora*: a first account of the larger species in the British Isles. Fottea, Olomouc 8, 15–78.

Marker, A.F.H., 1976. The benthic algae of some streams in southern England I Biomass of the epilithon in some small streams. Journal of Ecology 64, 343–358.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, R.B., 2007. Vegan: Community Ecology Package version 1. 8-6. http://cran.r-project.org/.

Potapova, M., Hamilton, P.B., 2007. Morphological and ecological variation within the *Achnanthidium minutissimum* (Bacillariophyceae) species complex. Journal of Phycology 45, 561–575.

Poulíčková, A., Špačková, J., Kelly, M.G., Duchoslav, M., Mann, D.G., 2008. Ecological variation within *Sellaphora* species complexes (Bacillariophyceae): specialists or generalists? Hydrobiologia 614, 373–386.

R Project Core Development Team, 2005. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Rott, E., Pipp, E., Pfister, P., van Dam, H., Ortler, K., Binder, N., Pall, K., 1999. Indikationslisten für Aufwuchsalgen in österreichischen Fließgewassern. Teil 2: Trophieindikation (sowie geochemische Präferenzen, taxonomische und toxikologische Anmerkungen). Wasserwirtschaftskataster herausgegeben vom Bundesministerium f. Land- u. Forstwirtschaft, Wien.

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. Ecology Applied 16, 1267–1276.

Tison, J., Park, Y.-S., Coste, M., Wasson, J.G., Ector, L., Rimet, F., Delmas, F., 2005. Typology of diatom communities and the influence of hydro-ecoregions: a study on the French hydrosystem scale. Water Research 39, 3177–3188.

Wallin, M., Wiederholm, T., Johnson, R.K., 2005. Guidance on establishing reference conditions and ecological status class boundaries for inland surface waters. 93 pp. CIS Working Group 2.3 – REFCOND. http://www-nrciws.slu.se/REFCOND/7th_REFCOND_final.pdf.