Intercalibration of national methods to assess the ecological quality of rivers in Europe using benthic invertebrates and aquatic flora

Inaugural-Dissertation

zur Erlangung des Doktorgrades

Dr. rer. nat.

des Fachbereichs

Biologie und Geografie

an der

Universität Duisburg-Essen

vorgelegt von

Sebastian Birk

geboren in Bottrop

April, 2009

Die der vorliegenden Arbeit zugrunde liegenden Experimente wurden in der Abteilung für Angewandte Zoologie/Hydrobiologie der Universität Duisburg-Essen durchgeführt.

1. Gutachter: Prof. Dr. Daniel Hering

2. Gutachter: Prof. Dr. Bernd Sures

3. Gutachter: Prof. Dr. Jürgen Böhmer

Vorsitzender des Prüfungsausschusses: Prof. Dr. Perihan Nalbant

Tag der mündlichen Prüfung: 15. Oktober 2009

“With a thousand eyes, the river looked at him, with green ones, with white ones, with crystal ones, with sky-blue ones. How did he love this water, how did it delight him, how grateful was he to it! In his heart he heard the voice talking, which was newly awaking, and it told him: Love this water! Stay near it! Learn from it! Oh yes, he wanted to learn from it, he wanted to listen to it. He who would understand this water and its secrets, so it seemed to him, would also understand many other things, many secrets, all secrets.”

Hermann Hesse: Siddhartha – An Indian Tale

Table of contents

List of tables

List of figures

List of abbreviations

AQEM	The Development and Testing of an Integrated Assessment System for the Ecological Quality of Streams and Rivers throughout Europe using Benthic Macroinvertebrates (European Research Project)
ASPT	Average Score Per Taxon
AT	Austria
BG	Bulgaria
BMWP	Biological Monitoring Working Party score
BOD	Biological Oxygen Demand
CLC	Corine Land Cover
CZ	Czech Republic
DE	Germany
DK	Denmark
DMS	Dutch Macrophyte Score
DSFI	Danish Stream Fauna Index
EQR	Ecological Quality Ratio
EU	European Union
FR	France
GD	General Degradation Index
GIG	Geographical Intercalibration Group
HU	Hungary
IBMR	Indice Biologique Macrophytique en Rivière
ICM	Intercalibration Common Metric
ICMi	Intercalibration Common Multimetric index (EE ICMi=Eastern European ICMi, dICMi=diatom ICMi)
IMI-IC	Integrative Multimetric Index for Intercalibration
IPS	Indice de Polluosensibilité Spécifique
MTR	Mean Trophic Rank
NL	Netherlands
PCA	Principal Components Analysis
PL	Poland
R	Correlation coefficient
R^2	Coefficient of determination
R-C…	Common intercalibration stream type of the Central-Baltic GIG
R-E…	Common intercalibration stream type of the Eastern Continental GIG
RI	Reference Index
RO	Romania
SE	Sweden
SI	Saprobic Index
SK	Slovak Republic
STAR	Standardisation of River Classifications: Framework method for calibrating different biological survey results against ecological quality classifications to be developed for the Water Framework Directive (European Research Project)
TI	Trophic Index
#fam	Number of invertebrate families
%EPT	Relative abundance of Ephemeroptera, Plecoptera and Trichoptera taxa
95CI	95 percent confidence interval of the regression line/curve

Acknowledgements

First sincere thanks go to Daniel Hering, my Ph.D. advisor and mentor in the various facets of scientific life. His keen intellect was my role model for tackling the complex questions of my work, and his pragmatism saved me valuable hours of theoretical idling. Nigel Willby and Christian Chauvin introduced me to the mysterious worlds of river macrophyte assessment. Our long discussions about *Callitriche hamulata* and sandy lowland brooks also made even my dead-end efforts worthwhile. Moreover, I am grateful for the enlightening experiences I made as a member of the Rivers Intercalibration Steering Group. Especially Isabel Pardo, John Murray-Bligh, Roger Owen, Jean-Gabriel Wasson, Martyn Kelly and Wouter van de Bund broadened my horizon regarding the concepts of river ecology and their practical implementation. The support of Jürgen Böhmer in all issues of intercalibration is also gratefully acknowledged.

Further thanks are due to Ursula Schmedtje, who invited me to work with the International Commission for the Protection of the Danube River. My challenging cooperation with the colleagues from Eastern Europe were a pleasure and enriched my work in various ways. In this regard I am especially grateful for the close collaboration with Birgit Vogel.

Furthermore, I always appreciated the sound working atmosphere at the Department of Applied Zoology/Hydrobiology. I like to express my sincere thanks to the colleagues that supported me in many different ways: Thomas Korte, Christian Feld, Armin Lorenz, Jelka Lorenz, Jörg Strackbein, Carolin Meier, Nadine Haus, Marta Wenikajtys and Peter Rolauffs. My work significantly benefited from the discussions held among the experts of the working group for river macrophyte intercalibration.

Finally, I like to thank my wife Inga, who always backed my ventures. I am also grateful to my parents and grandparents for all their guidance and advice.

This work was financially supported by the 5th Framework Programme of the European Commission, the German Working Group on Water Issues (LAWA), the German Federal Environment Agency (UBA), and the UNDP Danube Regional Project.

Preface

The question “What should I do?“ is posed by Kant (1800) as one of the four principal questions of philosophy. It addresses the broad field of ethics, encompassing right conduct and good life. Its relevance was mainly recognized for the interpersonal relations in social life. But with increasing awareness of the severe environmental effects of human behaviour the relationship between man and the natural world was brought into focus (Hardin, 1968, White, 1968). In this context Kant’s “What should I do?” can thus be specified as “How do I have to behave towards the natural environment?”. Ecology cannot answer this question as it implies normative statements beyond the descriptive character of science (Hume, 1978, Valsangiacomo, 1998). Our notion of the right conduct towards the environment forms part of the social discourse and, as such, becomes manifest, for instance, in environmental policy. Here, it shapes the moral background embedding the application of science.

The doctoral thesis at hand comprises applied science serving the implementation of the European Water Framework Directive (WFD) (European Commission, 2000). This comprehensive legislation establishes a framework for common action in the field of water policy among the 27 Member States of the European Union. The WFD obliges Member States to classify the ecological quality of their rivers, lakes, coastal waters and estuaries. Countries are applying assessment methods to evaluate the status of biological quality elements, i.e. selected groups of plants or animals inhabiting the aquatic environment. These methods distinguish between different types of surface waters, for instance small sandy lowland brooks or alpine streams with gravely substrates, and classify water bodies within these types in either high, good, moderate, poor or bad quality status. The WFD requires that all surface water bodies must achieve good ecological quality status, determining this status through normative definitions (European Commission, 2000, p. 38):

“The values of the biological quality elements for the surface water body type show low levels of distortion resulting from human activity, but deviate only slightly from those normally associated with the surface water body type under undisturbed conditions.”

This definition of good ecological status represents a key element of European water policy. The Union commits its Member States to the right conduct towards the aquatic environment and imposes restoration action if water bodies fail to achieve this objective. The concept of good ecological quality is therefore of crucial importance in the implementation of the WFD. However, the Directive leaves it to the Member States to put this rather vague definition into practice: Thus, the individual countries are in charge of developing national assessment methods and classifying the ecological status of their water bodies. To compare and to harmonize the national interpretations of good ecological status, the WFD stipulates an intercalibration exercise (Heiskanen et al., 2004).

The purpose of intercalibration is to set a common level of ambition among Member States in achieving the WFD's objectives. Ideally, intercalibration must ensure that, for instance, a German water body in good status according to the German assessment method would be classified as "good" by the Dutch or Danish method, if the same water body was located on a Dutch or Danish river. However, the biological communities of surface waters differ between countries even within the same water body type, under conditions not influenced by man. Furthermore, the national status classifications are characterized by differing assessment concepts and traditions (Birk, 2003, Birk & Schmedtje, 2005). Regarding only the classification of rivers and lakes, both undertaken using four biological quality elements (phytoplankton, phytobenthos and macrophytes, benthic invertebrates, fish), more than 200 national assessment methods have to be intercalibrated between the 27 Member States of the European Union. This gives an idea of the difficult and complex character of intercalibration.

The scientific work presented in this thesis establishes the methodological basis for the technical implementation of intercalibration. The fundamental question guiding the entire research is: How can the definitions of good ecological status be best compared between national assessment methods? Since all assessment methods employ biological indices to classify the ecological status, investigating the correlations of these indices is a primary task of intercalibration. According to the Directive, good status shall "deviate only slightly from […] undisturbed conditions". This statement highlights two important aspects relevant for the comparison of national classifications: First, undisturbed conditions form the reference point of ecological status assessment. And second, good status is defined as a slight deviation from this reference. This thesis looks into the role of reference conditions in the intercalibration exercise. In particular, different approaches aiming at harmonized reference setting are tested. In this regard the question is raised, whether good status can be defined without reference to undisturbed conditions.

The four chapters of this dissertation cover a total of 26 national methods for the ecological quality assessment of rivers using benthic invertebrates (15 methods), macrophytes (9 methods) and benthic diatoms (2 methods). In the various analyses more than 1,900 biological samples or surveys taken at rivers in 17 European countries are processed. The work includes data of three stream types common to Member States in Central and Western Europe, and four common types located in Eastern Europe.

Each chapter comprises an individual case study focussing on specific quality elements or distinct geographical regions. The basic approach throughout the thesis is to compare the assessment methods using international datasets that cover river sites impacted by different levels of anthropogenic pressure. This allows discrepancies to be identified in the national quality class boundary settings that define good status, i.e. the high-good and good-moderate boundaries. Following ECOSTAT (2004a) two options of intercalibration are examined in this thesis: direct comparison of assessment methods and indirect comparison of assessment methods using common metrics (Buffagni et al., 2005).

The case studies provide a broad and coherent picture of the questions of intercalibration. The contents of the four chapters are interdependent; in the first two studies elementary intercalibration approaches are investigated on which the latter two chapters are based. In Chapter 1 the direct comparison of invertebrate-based methods is explored. By means of correlation analyses various biological indices are matched for eight countries sharing two common stream types. The outcomes reveal strong relationships between methods, but deviating definitions of good ecological quality. Supportive environmental data is used to illustrate the level of anthropogenic pressure associated with the respective good-moderate boundary of each national method.

The following two chapters deliver fundamental insights into the intercalibration of assessment methods for river macrophytes. In search of the most suitable way for comparing national classifications both intercalibration options are studied in Chapter 2. The results show that national macrophyte methods are conceptually different, making intercalibration even more challenging. In particular, divergences in the detection of pressures (nutrient enrichment versus unspecific stresses) and the definition of the natural reference state become evident. In view of these difficulties Chapter 3 identifies the similarities of national methods to establish common grounds in macrophyte intercalibration. Sites classified in either high or bad status by the majority of national methods allow for a generic description of macrophyte communities under undisturbed and degraded conditions. Furthermore, method comparison is enabled by delineating indicator taxa that are used in a common metric for macrophytes.

The work of Chapter 4 includes the comparison of ecological classifications for five Eastern European countries. Common metrics are applied in the intercalibration of national methods using benthic diatoms and invertebrates. The availability of data from undisturbed reference sites, indispensable for the intercalibration approach described by Kelly et al. (2008) and Owen et al. (2010), is generally scarce for most of the stream types dealt with in this chapter. Therefore, an alternative approach based on sites impacted by similar levels of disturbance is employed. The biological benchmarks derived from these sites set transnational reference points for the harmonization of national quality classifications. For Austria and the Slovak Republic the outcomes of this study have led to legally binding requirements that are stipulated in a Commission Decision on quality class boundaries (European Commission, 2008).

The contents of this thesis contribute to the early outcomes of the ongoing intercalibration process, that now involves an increasing number of scientists all over Europe. The work at hand represents an essential contribution to the process of successfully completing intercalibration. Moreover, this dissertation can be seen in support of implementing a moral standard by scientific means: the definition of the right conduct towards the environment.

1 Direct comparison of assessment methods using benthic macroinvertebrates: a contribution to the EU Water Framework Directive intercalibration exercise

1.1 Introduction

In the individual European countries the practice of evaluating ecological river quality is very different (Metcalfe-Smith, 1994; Knoben et al., 1995; Birk & Hering, 2002). Although river monitoring programmes in most countries are based on the benthic macroinvertebrate community, design and performance of individual methods to assess rivers with this organism group vary significantly. On the one hand this is due to different traditions in stream assessment. While in many Central and Eastern European countries modifications of the Saprobic System have been applied for decades as standard methods (Birk & Schmedtje, 2005, see also Chapter 4), other countries rely on the Biological Monitoring Working Party score (BMWP, 1978), which has been adjusted for the use in various countries (Armitage et al., 1983; Just et al., 1998; Alba-Tercedor & Pujante, 2000; Kownacki et al., 2004). On the other hand the EU Water Framework Directive had a great effect on European freshwater management, since it outlines an innovative concept of bioassessment: Not the impact of single pressures on individual biotic groups but the deviation of the community from undisturbed conditions is decisive for ecological status classification. In many EU Member States efforts are being made to adapt the national programmes to these new requirements; however, different approaches are being used, since in some countries a single stressor (e.g. organic pollution) is overwhelming, while in other regions different stressors are of equal importance and simultaneously affect river inhabiting communities.

To overcome the difficulties in comparing the various national assessment methods the Directive outlines an intercalibration procedure of the methods' outputs. Member States are enabled to establish or to maintain their own methods; a definition of high, good or moderate biological quality is provided centrally through the intercalibration exercise. The aim of the intercalibration exercise is to identify and to resolve significant inconsistencies between the quality class boundaries established by Member States and indicated by the normative definitions of the Directive (ECOSTAT, 2004a).

The first efforts to compare different national assessment methods in Europe go back to 1975. Three intercalibration campaigns organised by the Commission of the

European Communities included comparisons of field sampling, sample treatment and quality assessment applied in Germany, Italy and United Kingdom (Tittizer, 1976; Woodiwiss, 1978; Ghetti & Bonazzi, 1980). These early studies established strong correlations between the individual assessment methods and compared the methods directly. This approach towards intercalibration was then followed by various authors both to demonstrate the relationship of methods and to point out discrepancies between national quality classifications (Ghetti & Bonazzi, 1977; Rico et al., 1992; Friedrich et al., 1995; Biggs et al., 1996; Morpurgo, 1996; Stubauer & Moog, 2000). In their preparatory study for the Water Framework Directive Nixon et al. (1996) explicitly recommended direct comparison to be used for the intercalibration of assessment methods.

However, the official intercalibration exercise for the Water Framework Directive has adopted an alternative approach due to the lack of comparable base data: indirect comparison via Intercalibration Common Metrics, thus, generating a “common” multimetric assessment procedure, which is more or less applicable in most of Europe, and comparing national assessment methods against this common method (Buffagni et al., 2006).

In this chapter I

(1) evaluate the principal suitability of directly comparing assessment methods for intercalibration procedures;

(2) test a variety of different regression techniques to refine the practical application of direct comparison for intercalibration purposes;

(3) directly compare assessment methods frequently applied for two broadly defined European river types and suggest steps for harmonizing class boundaries.

1.2 Methods

1.2.1 Overview

This study was based on a two-step analysis: First, different assessment methods, which are presently being used in national water management, were calculated with the same taxa lists. The results of the individual assessment methods were then directly compared by regression analysis.

All data used in this study resulted from the AQEM project (Hering et al., 2004) and the STAR project (Furse et al., 2006). Only data on invertebrate samples restricted to two broadly defined stream types were used. With the data from each stream type up to 10 national assessment systems were calculated, which were first normalized by calculating “Ecological Quality Ratios” (i.e. transferring the results into a common scale ranging from 0 to 1 where 1 equals the reference condition). These normalized assessment results were fed into a regression analysis, to translate the index results of country A into the index results of country B. Comparison of more than two methods was enabled by including the index of country C and translating these results into the index results of country B (“common scale”). In addition, the assessment results were correlated to environmental gradients. In a second step, the class boundaries between the individual quality classes, as applied by the national assessment systems, were compared.

To test the impact of different regression techniques on the results, linear and nonlinear techniques were compared.

1.2.2 Samples and sites

This study was based on benthic invertebrate data sampled in the EU projects AQEM and STAR with standardised field and laboratory protocols (Furse et al., 2006). The data were limited to two broadly defined stream type groups: small, siliceous mountain streams and medium-sized lowland streams in Central and Western Europe. In the official intercalibration exercise for the Water Framework Directive, these stream types were named “small-sized, mid-altitude brooks of siliceous geology” (R-C3) and “medium-sized, lowland streams of mixed geology” (R-C4) in Central Europe (Table 1).

294 samples taken at 125 sites located in four different countries in spring and summer were analysed for the small mountain streams. The lowland stream type embraced a total of 217 samples taken at 71 sites in four different countries in spring, summer and autumn.

The ecological quality of each sampling site was pre-classified based on expert judgement of the field researchers having sampled the streams and, if available, additional knowledge derived from previous studies. Each site was assigned to one of five quality classes (“high”, “good”, “moderate”, “poor”, “bad”) referring to the estimated main stressor's degree of impairment. For the AQEM sites, the pre-classification of most sites was replaced by the post-classification after sampling due to additional environmental parameters gained during the field work (physical-chemical and hydromorphological variables).

Table 1: Overview of samples included in the analysis

Stream type	Country	Stream type	Ecoregion no.	Number of samples
Small siliceous mountain streams	Austria	Small-sized shallow mountain streams	9	36
	Czech Republic	Small-sized shallow mountain streams	9, 10	40
		Small-sized streams in the Central Sub-alpine Mountains	9	32
		Small-sized streams in the Carpathians	10	28
	Germany	Small streams in lower mountainous areas of Central Europe	9	86
		Small-sized Buntsandstein-streams	9	24
	Slovak Republic	Small-sizes siliceous mountains streams in the West Carpathians	10	48
Medium-sized lowland streams	Denmark	Medium-sized deeper lowland streams	14	46
	Germany	Mid-sized sand bottom streams in the German lowlands	14	86
	Sweden	Medium-sized deeper lowland streams	14	14
		Medium-sized streams on calcareous soils	14	35
	United Kingdom	Medium-sized deeper lowland streams	18	36

1.2.3 National assessment methods and quality classifications

Altogether ten biological assessment indices were compared in this analysis (Table 2), all of which are either in current usage in certain European countries or are about being implemented into water management as standard techniques. Most represented biotic index or score methods (Saprobic Index, Biological Monitoring Working Party Score, Average Score Per Taxon, Danish Stream Fauna Index). All indices were part of the respective national method planned for biological monitoring in the context of the Water Framework Directive. With the exception of DSFI and ASPT, applied in Sweden, calculation of index values was based on a nationally adjusted indicator species list.

For the indices applied in Austria, the Czech Republic, Germany and Denmark, stream type specific reference values existed; these described the value of an index to be expected under "undisturbed conditions". The system used in the United Kingdom predicted site specific reference values, Sweden defined reference conditions for broad-scale natural geographical regions but in Poland and the Slovak Republic reference values have not yet been established. All indices distinguished between five classes of biological quality. The British and Swedish methods and the German multimetric index defined class boundary values as Ecological Quality Ratios. The Polish BMWP and the Saprobic Systems used quality classes given as absolute index values. The Austrian, Czech and German quality bands were stream type specific. An overview of nationally defined reference conditions and class boundaries is given in Table 3.

Table 2: Overview of national assessment methods (BI - Biotic Index, MI – Multimetric Index)

Stream type	Country	Assessment index	Category	Abundance	Reference
Small siliceous mountain streams	Austria	SI (AT) – Austrian Saprobic Index	BI	Y	Moog et al. (1999)
	Czech Republic	SI (CZ) – Czech Saprobic Index	BI	Y	CSN 757716 (1998)
	Germany	SI (DE) – German Saprobic Index	BI	Y	Friedrich & Herbst (2004)
	Poland	BMWP (PL) – Polish Biological Monitoring Working Party score	BI	N	Kownacki et al. (2004)
	Slovak Republic	SI (SK) – Slovak Saprobic Index	BI	Y	STN 83 0532-1 to 8 (1978/79)
	United Kingdom	ASPT (UK) - Average Score Per Taxon	BI	N	Armitage et al. (1983)
Medium-sized lowland streams	Denmark	DSFI (DK) – Danish Stream Fauna Index	BI	N	Skriver et al. (2000)
	Germany	GD (DE) – Module “General Degradation” of the German Assessment System Macrozoobenthos	MI[1]	Y	Böhmer et al. (2004)
	Sweden	ASPT (SE)- Average Score Per Taxon applied in Sweden	BI	N	Swedish Environmental Protection Agency (2000)
		DSFI (SE) – Danish Stream Fauna Index applied in Sweden	BI	N	
	United Kingdom	ASPT (UK) - Average Score Per Taxon	BI	N	Armitage et al. (1983)

1.2.4 Data preparation

National assessment methods were calculated to the taxa lists of each sample. Absolute index values were converted into Ecological Quality Ratios (EQR) by dividing

[1] Includes the following single metrics: “relative abundance of ETP taxa”, “German Fauna Index Type 15”, “number of Trichoptera taxa”, “Shannon-Wiener diversity”, “share of rheobiontic taxa”, “share of shredders [%]”

the calculated (observed) value by the index specific reference value. Since, for the Saprobic Indices, biological quality decreased with increasing index values these were converted by the following equation:

$$EQR_{SI} = 1 - \frac{\text{observed SI value} - \text{reference SI value}}{\text{maximum SI value} - \text{reference SI value}}$$

To validate the national reference values, an index specific reference value was calculated as the 75th percentile of all samples taken at sites pre- or post-classified as high quality status (excluding outliers). For the small mountain streams, sampling sites located in Austria (6 samples), Czech Republic (14 samples), Germany (13 samples) and Slovak Republic (1 sample) were used. For the lowland type sites from Denmark (13 samples), Germany (26 samples), Sweden (2 samples) and United Kingdom (9 samples) were the basis of this calculation.

Table 3: Original reference and class boundary values of the national assessment methods (abs – absolute value).

Small siliceous mountain streams						
Index	SI (AT)	SI (CZ)	SI (DE)	BMWP (PL)	SI (SK)	ASPT (UK)
Reference (abs)	≤ 1.50	≤ 1.20	≤ 1.25	n.a.	n.a.	≥ 6.62[2]
High|good	1.50	1.20	1.40	100	1.79	1.00
Good|moderate	2.10	1.50	1.95	70	2.30	0.89
Moderate|poor	2.60	2.00	2.65	40	2.70	0.77
Poor|bad	3.10	2.70	3.35	10	3.20	0.66
Lit. source	-	Brabec et al. (2004)	Rolauffs et al. (2003)	Kownacki et al. (2004)	STN 83 0532-1 to 8 (1978/79)	National Rivers Authority (1994)
Medium-sized lowland streams						
Index	DSFI (DK)	GD (DE)	BMWP (PL)	ASPT (SE)	DSFI (SE)	ASPT (UK)
Reference (abs)	7	1	n.a.	≥ 4.7	≥ 5	≥ 6.38[2]
High|good	7	0.80	100	0.90	0.90	1.00
Good|moderate	5	0.60	70	0.80	0.80	0.89
Moderate|poor	4	0.40	40	0.60	0.60	0.77
Poor|bad	3	0.20	10	0.30	0.30	0.66
Lit. source	-	Böhmer et al. (2004)	Kownacki et al. (2004)	Swedish Environmental Protection Agency (2000)	Swedish Environmental Protection Agency (2000)	National Rivers Authority (1994)

Conversion into the EQR scale resulted in values ranging from 0 to >1 since several samples revealed biological index values representing higher quality than the respective reference value. These values were not transformed into the value "1" in

[2] Values were derived by RIVPACS predictions for the corresponding stream type group based on averaged environmental parameter values and combined season information for the analysed samples.

order to improve the correlation and regression analysis by enlarging the quality gradient.

1.2.5 Correlation and regression analysis

The magnitude of the relation between two assessment methods was specified by the “coefficient of determination”. Beside linear regression, I applied nonlinear modelling via automatic curve-fitting using the software TableCurve 2D (SYSTAT Software Inc., 2002).

1.2.6 Comparison of quality class boundaries

In order to compare the national quality classes the boundary values of the different assessment methods were transformed into a “common scale”. In this study two common scales were used: (1) The national method showing the highest mean correlation of all indices. (2) The “Integrative Multimetric Index for Intercalibration” (IMI-IC), an artificial index designed here for the purpose of intercalibration. This index was defined as the mean of all index values calculated for a sample. The transformation was done based on the results of linear regression analyses, in which the predictor variables were represented by the national indices and the response variables by the “common scale”. Each boundary value transformed by regression was given including its 95 percent confidence interval. Class boundaries showing overlapping ranges (translated class boundary +/- confidence interval) were considered as being equal.

Based on environmental variables, abiotic gradients were generated for each stream type and the pressure gradients best correlating to the methods analysed in this intercalibration approach were identified. Indirect gradient analysis was aimed at the identification and quantification of physical-chemical and hydromorphological gradients that can be assigned to human impairment. Therefore, Principle Component Analysis (PCA) was run separately on correlation matrices of physical-chemical, catchment land use, hydromorphological and microhabitat variables of the mountain and lowland dataset. A dimensionless value of abiotic pressure, including the 95 percent confidence interval, was assigned to each national class boundary via regression analysis. These pressure data were used to support class boundary comparisons.

1.3 Results

1.3.1 Definition of reference values

The 75th percentiles of reference values were specified in Table 4. Each reference was based on a slightly different number of samples due to the elimination of outliers. Except for the German indices and the assessment methods for which no reference was nationally defined (Polish BMWP and Slovak SI), the 75th percentile, as calculated in this study, generally represented higher biological quality than the minimum values of the national reference.

Table 4: Reference values of national assessment methods derived by using the 75th percentile of index values calculated from samples taken at high status sites. For small mountain streams the number of high status sites' samples is individually specified in brackets. Values of lowland streams are based on 50 samples.

Small siliceous mountain streams						
Index	SI (AT)	SI (CZ)	SI (DE)	BMWP (PL)	SI (SK)	ASPT (UK)
75th percentile	1.46 (32)	0.91 (34)	1.44 (33)	187 (33)	1.21 (30)	7.26 (33)
Medium-sized lowland streams						
Index	DSFI (DK)	GD (DE)	BMWP (PL)	ASPT (SE)	DSFI (SE)	ASPT (UK)
75th percentile	7	0.67	150	6.57	7	6.57

1.3.2 Descriptive statistics of national indices calculated from the AQEM-STAR datasets

The overall mean of normalized index values (0 to 1) for the small mountain streams amounted to 0.87, while the same statistic for medium-sized lowland streams was 0.77 (Table 5). The maximum values of all indices except DSFI exceeded 1.0. This was due to the selection of the 75th percentile of AQEM-STAR high status sites as the reference value. The values of the Polish BMWP and the German GD covered ranges of more than 1.0, while the Austrian and German SI, and the British and Swedish ASPT showed value ranges of less than 0.65.

1.3.3 Correlation and regression of national assessment methods

The correlation analysis revealed differences between assessment methods (Table 6). The linear equations of the regression analysis of national methods against methods representing a common scale (best correlating national index, IMI-IC) are displayed in

Table 7. Nonlinear equations are listed additionally if they provide higher coefficients of determination.

Table 5: Descriptive statistics of national indices calculated from the AQEM-STAR datasets (normalized index values).

Small siliceous mountain streams (n = 294)							
	Mean	Minimum	Maximum	25th percentile	75th percentile	Range	Quartile range
SI (AT)	0.902	0.526	1.112	0.833	0.972	0.585	0.138
SI (CZ)	0.853	0.374	1.112	0.761	0.963	0.739	0.202
SI (DE)	0.920	0.444	1.055	0.895	0.984	0.611	0.088
BMWP (PL)	0.768	0.102	1.273	0.636	0.936	1.171	0.299
SI (SK)	0.890	0.444	1.281	0.798	0.984	0.837	0.186
ASPT (UK)	0.908	0.448	1.077	0.869	0.988	0.629	0.119
Medium-sized lowland streams (n = 217)							
	Mean	Minimum	Maximum	25th percentile	75th percentile	Range	Quartile range
DSFI (DK) and DSFI (SE)	0.767	0.286	1.000	0.571	1.000	0.714	0.429
GD (DE)	0.709	0.090	1.149	0.552	0.896	1.060	0.343
BMWP (PL)	0.741	0.173	1.480	0.580	0.900	1.307	0.320
ASPT (SE) and ASPT (UK)	0.869	0.457	1.091	0.797	0.956	0.634	0.159

For small mountain streams coefficients of determination ranged from 0.20 (Slovak SI and Polish BMWP) to 0.77 (Austrian SI and Slovak SI). Nonlinear regression gained higher R^2 values in 23 out of 36 relations. The mean difference in R^2 values between linear and nonlinear regressions was 0.04. The maximum difference in R^2 values of 0.12 was between linear and nonlinear equations for the relationship between SI (SK) and ASPT (UK). German SI had the highest average correlation to the other assessment methods ($R^2 = 0.67$). The IMI-IC for this stream type was characterised by coefficients of determination ranging from 0.62 (Slovak SI) to 0.87 (German SI). In Figure 1 regression lines of BMWP (PL) against SI (DE) were exemplarily plotted for linear and nonlinear regression. R^2 values for regressions of methods for the lowland streams varied between 0.41 (German GD and Polish BMWP) and 0.67 (British and Swedish ASPT, and Danish and Swedish DSFI). In 6 out of 16 correlations, nonlinear regression provided a higher proportion of the variance explained. Mean difference of the linear and nonlinear coefficients of determination was $R^2 = 0.02$ and the maximum difference was $R^2 = 0.06$ (Polish BMWP and British ASPT). DSFI showed the highest mean correlation for the lowland samples ($R^2 = 0.60$). The IMI-IC had coefficients of determination ranging from 0.73 (Polish BMWP) to 0.90 (Danish and Swedish DSFI). All correlations were significant at $p < 0.05$. Since none of the differences between the linear and nonlinear coefficients of determination were significant, I assumed linear relationships between indices in the following analyses.

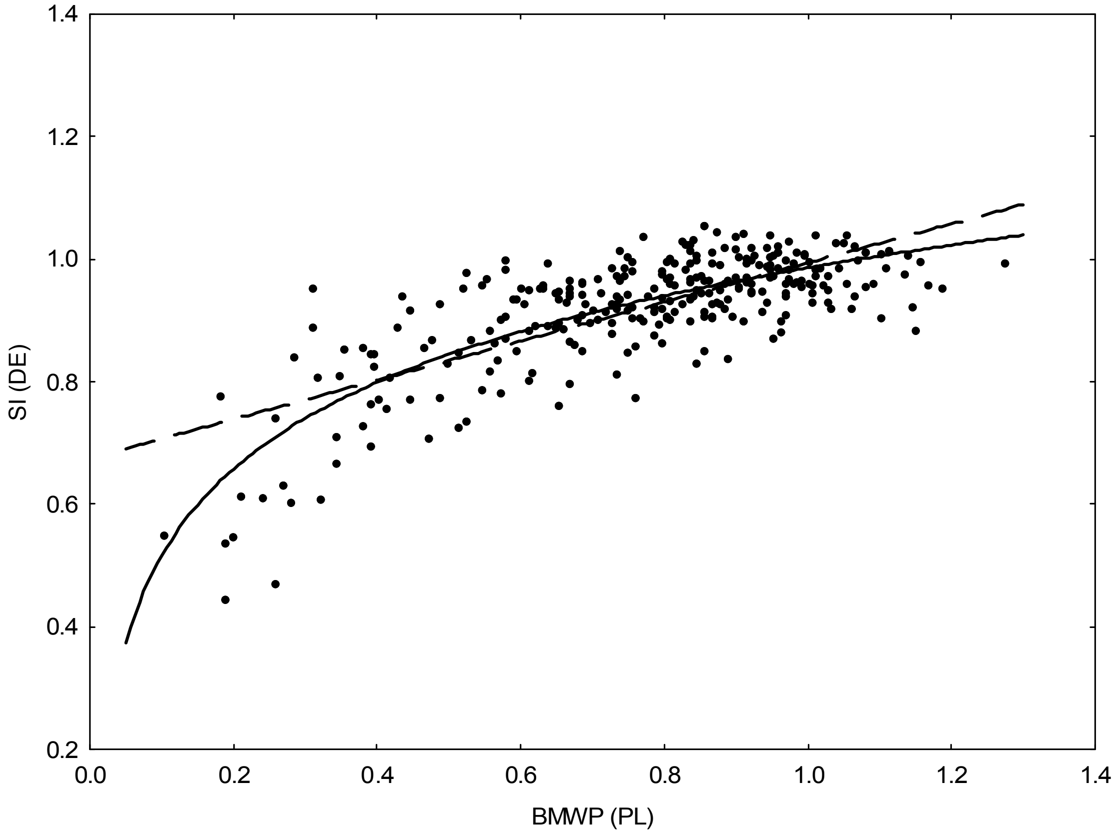

Figure 1: Regression of BMWP (PL) against SI (DE). Both linear (R^2 = 0.53, dashed) and nonlinear (R^2 = 0.63) regression lines are plotted.

1.3.4 Correlation to environmental gradients (PCA)

Index values of the small mountain streams showed the strongest relationship with the PCA gradient reflecting nutrient enrichment and organic pollution. Determination coefficients of this gradient and the assessment methods varied from 0.19 (Slovak SI) to 0.53 (British ASPT). Index values of the lowland streams showed highest correlations with the main hydromorphological gradient that comprised physical features of the river channel, its banks and immediate vicinity, including information on the degree of impairment. The coefficients of determination ranged between 0.12 (Polish BMWP) and 0.35 (German GD).

1.3.5 Comparison of national quality classes

The comparison of biological quality classes was based on the transformation of boundary values of the assessment methods into a common scale. This allowed for a direct juxtaposition of class boundaries in Table 8.

Small-sized siliceous mountain streams

The common scales used in the comparison procedure for the mountain streams were SI (DE) and IMI-$IC_{R\text{-}C3}$ (multimetric index composed of all national assessment methods). In SI (DE) scale, the high-good boundaries of SI (AT) and ASPT (UK) were similar considering the 95 percent confidence interval. ASPT (UK) and SI (CZ) showed overlapping good-moderate boundary intervals and thus shared equal class boundaries. The same applied for the group of indices SI (AT), SI (DE), BMWP (PL) and SI (SK). Based on IMI-$IC_{R\text{-}C3}$ the high-good boundaries of SI (AT) and ASPT (UK) shared common intervals. For the good-moderate boundary the comparison showed similar values for SI (AT), BMWP (PL) and SI (SK).

The pollution/eutrophication gradient showed similar pressure between high-good boundaries of SI (AT), SI (CZ), SI (DE), ASPT (UK), and BMWP (PL) and SI (SK). For the good-moderate boundary corresponding levels of chemical impairment were between SI (AT) and SI (DE), SI (SK) and BMWP (PL), and SI (CZ) and ASPT (UK). The average confidence interval amounted to 0.025 units.

Medium-sized, lowland, mixed geology

The DSFI and IMI-$IC_{R\text{-}C4}$ (multimetric index composed of all national assessment methods) were used as common scales for the boundary comparisons of the lowland stream type. Using DSFI as the common scale, none of the national indices showed similar high-good class boundaries but the good-moderate boundaries of DSFI (SE) and ASPT (UK) were corresponding. The average confidence interval amounted to 0.017 DSFI units.

Table 6: Coefficients of determination based on linear and nonlinear regression ($p < 0.05$) – (IMI-IC: Integrative Multimetric Index for Intercalibration (see text for explanation); PE1: pollution/eutrophication gradient; HY1: hydromorphological gradient)

Small siliceous mountain streams (n = 294)												
Index	SI (AT)		SI (CZ)		SI (DE)		BMWP (PL)		SI (SK)		ASPT (UK)	
	linear	nonl.	linear	nonl.	linear	Nonl.	linear	nonl.	linear	nonl.	linear	nonl.
SI (AT)	1.00	-	0.62	-	0.70	0.74	0.36	0.39	0.73	0.77	0.45	0.46
SI (CZ)	0.62	-	1.00	-	0.62	0.64	0.31	0.35	0.55	-	0.38	-
SI (DE)	0.70	0.73	0.62	0.70	1.00	-	0.53	0.63	0.48	0.56	0.69	0.73
BMWP (PL)	0.36	0.37	0.31	0.34	0.53	-	1.00	-	0.20	0.23	0.69	0.70
SI (SK)	0.73	-	0.55	-	0.48	0.51	0.20	0.21	1.00	-	0.24	0.26
ASPT (UK)	0.45	0.50	0.37	0.45	0.69	0.70	0.69	0.75	0.24	0.36	1.00	-
IMI-IC$_{R-C3}$	0.79	0.80	0.72	0.74	0.86	0.87	0.72	0.75	0.62	0.66	0.75	-
PE1	0.31	0.33	0.23	0.27	0.46	-	0.37	0.38	0.19	0.23	0.53	-

Medium-sized lowland streams (n = 217)								
Index	DSFI (DK) and DSFI (SE)		GD (DE)		BMWP (PL)		ASPT (SE) and ASPT (UK)	
	linear	nonl.	linear	nonl.	linear	nonl.	linear	nonl.
DSFI (DK) and DSFI (SE)	1.00	-	0.61	-	0.53	0.54	0.65	-
GD (DE)	0.61	-	1.00	-	0.41	0.46	0.49	-
BMWP (PL)	0.53	0.54	0.41	-	1.00	-	0.51	-
ASPT (SE) and ASPT (UK)	0.65	0.67	0.49	0.50	0.51	0.57	1.00	-
IMI-IC$_{R-C4}$	0.90	-	0.76	-	0.73	0.75	0.80	-
HY1	0.23	-	0.35	-	0.12	0.13	0.24	0.26

Table 7: Coefficients of linear regression equations (a - slope, b - intercept) for the common scales and the abiotic gradients (IMI-IC: Integrative Multimetric Index for Intercalibration (see text for explanation); PE1: pollution/eutrophication gradient; HY1: hydromorphological gradient)

Small siliceous mountain streams												
Index	SI (AT)		SI (CZ)		SI (DE)		BMWP (PL)		SI (SK)		ASPT (UK)	
Parameter	a	b	a	b	a	b	a	b	a	b	a	b
SI (DE)	0.784	0.212	0.562	0.440	1.000	0	0.319	0.675	0.511	0.465	0.687	0.296
$IMI\text{-}IC_{R\text{-}C3}$	0.992	-0.021	0.717	0.261	1.100	-0.138	0.441	0.535	0.688	0.261	0.850	0.102
PE1	-0.845	1.000	-0.567	0.720	-1.089	1.236	-0.450	0.577	-0.542	0.721	-0.976	1.120

Medium-sized lowland streams								
Index	DSFI (DK) and DSFI (SE)		GD (DE)		BMWP (PL)		ASPT (SE) and ASPT (UK)	
Parameter	a	b	a	b	a	b	a	b
DSFI	1.000	0.000	0.579	0.356	0.344	0.570	1.349	-0.405
$IMI\text{-}IC_{R\text{-}C4}$	0.825	0.154	0.566	0.386	0.357	0.580	1.301	-0.343
HY1	-0.627	0.934	-0.583	0.857	-0.360	0.720	-1.078	1.396

Table 8: EQR values of the high-good (H|G) and good-moderate (G|M) quality class boundaries transferred into “common scale”. In addition, the values of the abiotic gradients (PE1, HY1) corresponding to the national class boundaries are displayed. For each value derived by regression the 95 percent confidence interval is specified (IMI-IC: Integrative Multimetric Index for Intercalibration (see text for explanation); PE1: pollution/eutrophication gradient; HY1: hydromorphological gradient)

Small siliceous mountain streams													
Class boundary	Common scale	SI (AT)		SI (CZ)		SI (DE)		BMWP (PL)		SI (SK)		ASPT (UK)	
		Boundary value	*95% confid.*	Boundary value	*95% confid.*	Boundary value	*95% confid.*	Boundary value	*95% confid.*	Boundary value	*95% confid.*	Boundary value	*95% confid.*
H|G	SI (DE)	0.984	0.008	0.949	0.008	1.016	-	0.846	0.011	0.870	0.010	0.983	0.008
	$IMI\text{-}IC_{R\text{-}C3}$	0.955	0.008	0.911	0.008	0.979	0.008	0.771	0.010	0.806	0.011	0.952	0.009
	PE1	0.169	0.023	0.206	0.019	0.130	0.023	0.336	0.022	0.291	0.023	0.144	0.019
G|M	SI (DE)	0.799	0.012	0.895	0.007	0.801	-	0.794	0.016	0.776	0.020	0.907	0.006
	$IMI\text{-}IC_{R\text{-}C3}$	0.721	0.012	0.842	0.008	0.743	0.009	0.700	0.015	0.680	0.021	0.858	0.007
	PE1	0.368	0.032	0.262	0.019	0.364	0.025	0.409	0.032	0.391	0.045	0.251	0.014
Medium-sized lowland streams													
Class boundary	Common scale	DSFI (DK)		GD (DE)		BMWP (PL)		ASPT (SE)		DSFI (SE)		ASPT (UK)	
		Boundary value	*95% confid.*	Boundary value	*95% confid.*	Boundary value	*95% confid.*	Boundary value	*95% confid.*	Boundary value	*95% confid.*	Boundary value	*95% confid.*
H|G	DSFI	1.000	-	1.048	0.012	0.724	0.018	0.809	0.016	0.900	-	0.944	0.025
	$IMI\text{-}IC_{R\text{-}C4}$	0.979	0.012	1.061	0.008	0.744	0.012	0.827	0.011	0.897	0.009	0.958	0.017
	HY1	0.307	0.054	0.162	0.021	0.480	0.036	0.426	0.035	0.370	0.042	0.318	0.054
G|M	DSFI	0.714	-	0.875	0.016	0.610	0.016	0.674	0.019	0.800	-	0.795	0.016
	$IMI\text{-}IC_{R\text{-}C4}$	0.744	0.008	0.892	0.011	0.628	0.011	0.697	0.012	0.814	0.007	0.814	0.010
	HY1	0.486	0.035	0.335	0.030	0.552	0.034	0.534	0.041	0.432	0.035	0.437	0.034

Table 9: Comparison of the saprobic indicator taxa lists of Austria, Czech Republic, Germany and Slovak Republic: Share of common taxa and coefficients of determination derived from correlation analysis of indicator values and indicator weights.

	SI (AT)			SI (CZ)			SI (DE)			SI (SK)		
	Share of common taxa	Indicator value	Indicator weight	Share of common taxa	Indicator value	Indicator weight	Share of common taxa	Indicator value	Indicator weight	Share of common taxa	Indicator value	Indicator weight
SI (AT)	-	1.00	1.00	56 %	0.64	0.14	72 %	0.74	0.04	77 %	0.88	0.53
SI (CZ)	36 %	0.64	0.14	-	1.00	1.00	54 %	0.74	0.14	53 %	0.73	0.31
SI (DE)	35 %	0.74	0.04	41 %	0.74	0.14	-	1.00	1.00	41 %	0.73	0.04
SI (SK)	45 %	0.88	0.53	48 %	0.73	0.31	49 %	0.73	0.04	-	1.00	1.00

In the IMI-IC$_{R-C4}$ scale, the high-good boundaries of DSFI (DK) and ASPT (UK) had similar values and the good-moderate boundaries of DSFI (SE) and ASPT (UK) corresponded closely. Confidence intervals showed an average value of 0.011 units.

Boundary comparisons using the hydromorphological gradient were difficult because the large confidence intervals (0.038 units in average) resulted in overlapping boundary ranges. Both good quality boundaries of GD (DE) showed the lowest level of pressure. For the good-moderate boundary, levels of pressure were similar between DSFI (DK), DSFI (SE) and ASPT (UK), and between BMWP (PL) and ASPT (SE).

1.4 Discussion

1.4.1 Role of reference conditions in the intercalibration exercise

Within the intercalibration exercise, class boundaries of national assessment methods need to be defined as Ecological Quality Ratios. The position of each boundary on this relative scale is dependent on (1) the definition of reference conditions and (2) the procedure of setting class boundaries. If the former is not properly dealt with in the intercalibration process, the different nationally defined reference values may strongly impact upon comparability.

In this chapter I have defined a common reference, which is based on sites in several countries. As a result of this common reference, it was possible to include several methods in the comparison, even if countries have not yet defined reference values for a specific method. A further advantage of common references is that differences in national approaches to define references are avoided. On the other hand, common references are in danger of not adequately accounting for the differences between the more specific national streams types.

More importantly, countries have applied different procedures to define reference values and quality classification schemes. While this study is restricted to the analysis of national class boundary settings, it must be an objective of the official intercalibration exercise to overcome differences in the references too.

1.4.2 Relations between assessment methods

In this study, the calculation of national assessment metric values is based on taxa lists derived by application of the standardised STAR-AQEM field and laboratory protocol.

Thus, the correlation analyses of index values mainly reveal the numerical relation between these indices and is less biased by differences in field and laboratory procedures. The character of these relations depends on the architecture of the individual indices, e.g. number and indicative value of taxa included in the evaluation, type of abundance information used and the assessment formula. The effect of different national sampling methods on the comparability of taxa lists and metric results as a major constraint of intercalibration is investigated by Friberg et al. (2006). Buffagni et al. (2006) present a practical approach enabling the use, in intercalibration, of datasets derived by the national monitoring programmes.

An additional factor, impacting on the relationships, is the dataset itself, in particular the number of samples, the biogeographical gradient, the types of pressures influencing sampling sites and the range of degradation covered. The different ranges of index values (cf. Table 5) indicate a larger degradation gradient being covered by the lowland dataset. This is, in particular, obvious from the Polish BMWP and British ASPT values, which have been calculated for both datasets.

For the mountain stream data, relationships are strongest between the values of the different Saprobic Indices of Austria, Czech Republic, Germany and Slovak Republic and between the score methods applied in Poland and the United Kingdom. In general, the strength of correlations between the different Saprobic Indices results from similarities in indicator taxa and their indication values (Table 9). For instance, the Austrian and Slovak Saprobic Indices ($R^2 > 0.73$) share the largest number of indicator taxa and are most closely related concerning indicator taxa value and weight. Schmidt-Kloiber et al. (2006) provide a comprehensive analysis of saprobic indicator taxa applied in Europe.

For the lowland stream dataset, BMWP (PL) and ASPT (UK) correlate less strongly ($R^2 < 0.60$), which can be explained by the different taxonomic composition of the lowland dataset compared to that of the mountain streams. The two indices have 66 indicator taxa in common, amounting to a share of 73 percent (Polish BMWP) and 80 percent (British ASPT), respectively. BMWP indicator values of the common taxa in the Polish and British systems are correlated with $R^2 = 0.73$.

Method comparisons of earlier studies show similar results. Based on 232 samples from various lowland and mountain stream types in Germany, Friedrich et al. (1995) found correlations of R^2 = 0.71 between ASPT (UK) and a previous version of the German Saprobic Index. The weak relation of ASPT and the Austrian Saprobic Index has already been demonstrated by Stubauer & Moog (2000), who used a large dataset covering all Austrian stream types (n = 588; R^2 = 0.52). Analyses of Birk & Rolauffs (2004) revealed strong correlations between the Austrian and German Saprobic Indices (n = 262; R^2 = 0.75).

Several indices revealed higher coefficients of determination when applying a nonlinear fit, in particular if BMWP (PL) was involved. This index combines the parameters taxon richness and sensitivity into a single value which may cause the observed relationship. Also, due to the large range of values covered by the method, the nonlinearity of the relationships became evident (cf. Figure 1). Nevertheless, these difference of the coefficients of determination are not significant. Therefore, the simple model of linear relationship between indices is most appropriate in this example of direct comparison.

1.4.3 Comparison of class boundary values

While earlier intercalibration studies focussed on the comparison of quality class bands (Ghetti & Bonazzi, 1977; Friedrich et al., 1995; Morpurgo, 1996), the Water Framework Directive specifically requires the comparability of the high-good and good-moderate quality class boundaries. Thus, the intercalibration exercise is focussed on the range medium to high biological quality. The original procedure outlined in the Directive is restricted to the use of just a few intercalibration sites, selected because they represent the boundary status between quality classes. However, this approach seems not to be feasible, since sites known to be on class boundaries cannot be selected prior to the intercalibration is completed and those boundaries are defined. Furthermore, the uncertainty of intercalibration results is high if the analysis is based on insufficient data.

Therefore, the primary step, in comparing national class boundary values and best identifying the type and magnitude of the relationship between national assessment methods, should be based on a large number of samples covering the entire quality gradient. In a further step, regression analysis should be used to transform boundary values into other assessment scales. By applying an acceptable level of uncertainty (e.g. confidence interval of 95 percent derived from regression analysis), ranges of index values can be compared.

The comparison of assessment methods has revealed discrepancies between national classification schemes of more than 25 percent in particular cases (e.g. high-good boundary of German SI and Polish BMWP translated in German SI scale). The extent of differences between class boundaries is largely dependent on the common scale used for comparison. While class boundaries clearly differ if compared through the German Saprobic Index scale, no differences occur between the same boundaries if compared through a multimetric index. Each method used as a common scale is somewhat related to other assessment methods as expressed by the correlation coefficient and the regression equation.

Based on these findings I recommend using the intercalibration approach described in this chapter only for comparison of methods addressing similar components of the biocoenosis, e.g. for methods that are closely related such as ASPT, BMWP and the Saprobic Indices, or methods that are fully compliant with the requirements of the Water Framework Directive (i.e. methods evaluating taxonomic composition and abundance, ratio of disturbance sensitive to insensitive taxa and diversity of the macroinvertebrate community). This principle makes sure that “like with like” comparisons are applied in intercalibration and minimises errors in the comparison analysis due to the selection of inappropriate common scales. Furthermore, the relation between assessment methods needs to be carefully evaluated. Nonlinear correlations yielding significantly better fit and smaller confidence intervals are to be favoured over weaker linear relations.

1.4.4 When shall boundaries be considered as different?

Intercalibration encompasses two steps: Firstly, national quality boundaries are compared. If this analysis discovers major differences in classification schemes, they need to be harmonized in a second step. For the first step, I have described a possible procedure to translate boundary values into a common scale, which determines whether or not boundary values are corresponding. According to my results only a few class boundaries are similar, which thus requires the remaining boundaries to be harmonized.

The use of abiotic pressure data in intercalibration allows for additional interpretation of results. Sandin & Hering (2004) applied organic pollution gradients to set intercalibration class boundaries defining a standard level of pollution. I particularly propose to use pressure information for the process of boundary comparison. Figure 2 displays the relative position of the national good-moderate boundaries, including confidence intervals translated into a common biotic scale and an abiotic pressure scale (pollution/eutrophication gradient). Comparisons based on the interpretation of biotic data indicate that four out of six class boundaries are deviating (cf. Table 8), while the consideration of pressure data (Figure 2) reveals only two groups of boundaries with overlapping pressure intervals. Thus, harmonization is only needed between the two groups of boundaries.

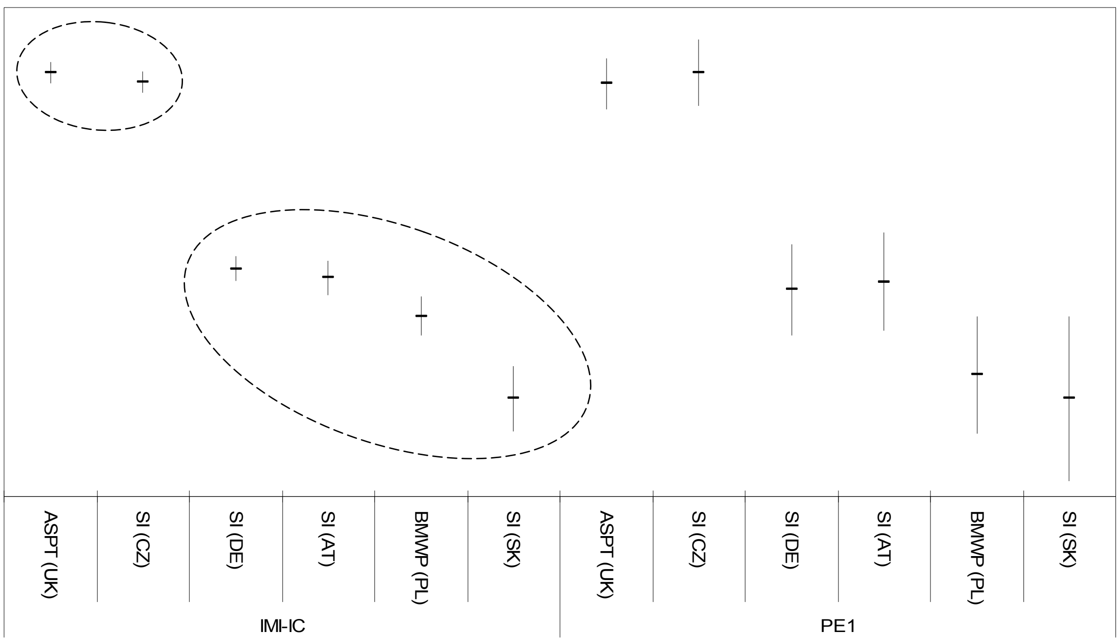

Figure 2: Relative comparison of good-moderate class boundary values (incl. 95 percent confidence intervals) using IMI-IC$_{R-C3}$ and corresponding chemical pressure values of the small siliceous mountain streams. Based on the results of the pressure data analysis two groups of similar boundaries are highlighted by dashed circles.

1.5 Conclusions

Intercalibration represents a crucial step towards the implementation of a pan-European water quality standard. Besides scientific issues, which I partly addressed in this chapter, it holds a major social challenge. Although assessment methods are in general scientifically sound instruments, the element of quality classification is a concession to the practical requirements of decision making in water policy. According to the Water Framework Directive the quality assigned to a site can decide on restoration efforts to be spent or saved. Therefore, intercalibration is of political interest since the definition of quality boundaries sets the environmental standard to be achieved. Furthermore, intercalibration holds an ethical component: By selecting certain quality criteria we agree on a level of anthropogenic degradation acceptable for our freshwater systems. Although beyond its scope science needs to consider all these aspects in the preparation of reasonable and tenable results.

2 Intercalibration of assessment methods for macrophytes in lowland streams: direct comparison and analysis of common metrics

2.1 Introduction

According to the EU Water Framework Directive (WFD; European Commission, 2000) European surface waters must achieve good ecological quality by the year 2015. Responsibility for the quality assessment lies with the individual Member States, which have developed or modified assessment methods at the national level. To ensure the comparability of the national methods, an intercalibration exercise is stipulated by the Directive, in which quality class boundaries are checked for comparability and consistency with normative requirements.

Although benthic macroinvertebrates are presently most commonly applied for the quality assessment of rivers (Birk & Hering, 2002), macrophytes are also surveyed in some countries to monitor the effects of anthropogenic pressures, especially eutrophication (Kelly & Whitton, 1998; Birk & Schmedtje, 2005). Macrophytes were first used in water quality assessment in relation to various modifications of the saprobic system. Several indicator catalogues (e.g. Sládeček, 1973) included single macrophyte species to evaluate the degree of organic pollution. More generally, the monitoring of macrophyte communities was confined to the description of the vegetation without inferring water quality (e.g. Holmes & Whitton, 1975; Janauer et al., 2003). With the increasing awareness of the effects of nutrient enrichment the community assessment of phototrophs gained in importance. The Mean Trophic Rank (MTR, Holmes et al., 1999), for instance, focuses on the impact of nutrient enrichment only, since it was elaborated and tested specifically for the application of the EU Urban Waste Water Treatment Directive (Council of the European Communities, 1991). The Water Framework Directive recently led to the development of national methods aimed at assessment of ecological quality of the aquatic flora (e.g. Molen et al., 2004; Leyssen et al., 2005; Meilinger et al., 2005). These methods differ in design and performance from macroinvertebrate based systems and, thus, require a separate intercalibration process.

The intercalibration procedure as outlined in the Directive comprises the comparison of intercalibration sites whose individual biological quality, in the opinion of the Member

States, represents the boundary between quality classes. Recent studies on intercalibration of macroinvertebrate methods are based on data representing a broad quality gradient, and class boundaries are compared via correlation and regression analysis. While in Chapter 1 of this thesis the national methods are directly compared by the help of a "common scale" (method best correlating with all other methods), Buffagni et al. (2006) use "common metrics" as a general scale. Common metrics are defined as biological metrics widely applicable within a geographical region, which can be used to derive comparable information among different countries and stream types (Buffagni et al., 2005).

In this chapter I apply the two above outlined approaches of boundary comparison to macrophyte data from lowland rivers covering a broad spectrum of anthropogenic disturbance from reference to heavily impacted sites. Furthermore, I test both techniques for their applicability in the intercalibration of four assessment methods for macrophytes.

2.2 Methods

2.2.1 Samples and sites

This study is based on river macrophyte survey data collected at medium-sized lowland streams in six countries in the framework of the EU project STAR (Furse et al., 2006). In the official intercalibration exercise for the Water Framework Directive, this stream type is named "medium-sized, lowland streams of mixed geology" (R-C4) in Central Europe (ECOSTAT, 2004b). Data used here were limited to 108 sites at which macrophytes covered at least 1 percent of the total channel area investigated (Table 10).

Table 10: Overview of the sites surveyed at medium-sized lowland streams.

Country	Number of surveys
Denmark	11
Germany	11
Latvia	36
Poland	24
Sweden	20
United Kingdom	6

Macrophytes were sampled using a single survey in late summer or early autumn. A 100 m stream length was surveyed in each stream by wading in a zigzag manner across the channel. Macrophytes of non-wadable sites were observed by boat or by walking along the banks. All macrophytes species were recorded as well as the percent cover of the overall macrophyte growth. Species were normally identified in the field, but if identification was uncertain a representative sample was collected for later identification. In addition, physico-chemical data were sampled. Table 11 lists statistical descriptors for the sampling site's trophic status.

Table 11: Range of trophic status covered by the dataset (n=108): descriptive statistics of the chemical parameters nitrate and total phosphorus

	Min	25th	Median	75th	Max
Nitrate (mg l^{-1})	0.03	0.25	1.50	2.00	12.10
Total phosphorus (mg l^{-1})	0.01	0.09	0.21	0.28	15.40

2.2.2 National assessment methods and quality classifications

Four methods to assess the quality of streams, which are being used in France, Germany, the Netherlands and the United Kingdom were compared (Table 12). All methods are based on species-level data and integrate specific indicator values and abundance information. Except for the German Reference Index, abundance is specified in classes of relative plant coverage. Abundance data used by the German index is an estimation of the three-dimensional structure of the instream vegetation (Kohler, 1978). Table 13 compares the different macrophyte abundance schemes. While the French and British methods were used alone, the German and Dutch indices are part of generic methods to assess the "aquatic flora", which is defined as including macrophytes and phytobenthos.

Table 12: Overview of macrophyte assessment methods

Country	Assessment method	Reference
France	IBMR (FR) – Indice Biologique Macrophytique en Rivière	NF T90-395 (2003)
Germany	RI (DE) – Reference Index	Schaumburg et al. (2004)
The Netherlands	DMS (NL) – Dutch Macrophyte Score ("Soortensamenstelling macrofyten")	Molen et al. (2004)
United Kingdom	MTR (UK) – Mean Trophic Rank	Holmes et al. (1999)

The Dutch and German methods aim at assessing the degree of deviation from the reference state and are, thus, based on stream type specific reference conditions. It is therefore necessary to classify the streams sampled here into specific stream types: For the German method sampling sites were assigned to the stream type "medium sized lowland rivers of northern Germany” (Schaumburg et al., 2004). Since the stream typology of the Netherlands is more complex sites have been allocated to eight different national types for the Dutch index (Elbersen et al., 2003).

Table 13: Comparison of macrophyte abundance schemes

<table>
<tr><th colspan="2">IBMR (FR)</th><th colspan="2">RI (DE)</th><th colspan="2">DMS (NL)</th><th colspan="2">MTR (UK)</th></tr>
<tr><th>Abundance class</th><th>Cover [%]</th><th>Abundance class (Kohler, 1978)</th><th>Plant quantity (Schneider, 2000)</th><th>Abundance class</th><th>Cover [%]</th><th>Abundance class</th><th>Cover [%]</th></tr>
<tr><td>1</td><td>< 0.1</td><td>1</td><td>1</td><td rowspan="4">1</td><td rowspan="4">< 5</td><td>1</td><td>< 0.1</td></tr>
<tr><td>2</td><td>0.1 – 1</td><td>2</td><td>8</td><td>2</td><td>0.1 - 1</td></tr>
<tr><td rowspan="3">3</td><td rowspan="3">1 – 10</td><td rowspan="3">3</td><td rowspan="3">27</td><td>3</td><td>1 - 2.5</td></tr>
<tr><td>4</td><td>2.5 - 5</td></tr>
<tr><td rowspan="3">2</td><td rowspan="3">5 - 50</td><td>5</td><td>5 - 10</td></tr>
<tr><td rowspan="2">4</td><td rowspan="2">10 – 50</td><td rowspan="3">4</td><td rowspan="3">64</td><td>6</td><td>10 - 25</td></tr>
<tr><td>7</td><td>25 - 50</td></tr>
<tr><td rowspan="2">5</td><td rowspan="2">> 50</td><td rowspan="2">3</td><td rowspan="2">> 50</td><td>8</td><td>50 - 75</td></tr>
<tr><td>5</td><td>125</td><td>9</td><td>> 75</td></tr>
</table>

The French, German and Dutch methods distinguish between five classes of ecological quality (Table 14). Since the British MTR was developed to illustrate responses to urban discharges by surveying two physically similar sites upstream and downstream the method is not designed for classifying the ecological quality of rivers. For interpretation purposes only, Holmes et al. (1999) suggest MTR boundary values to determine if the investigated site is (1) ‘unlikely to be eutrophic’, (2) ‘likely to be either eutrophic or at risk of becoming eutrophic’ or (3) ‘badly damaged by either eutrophication, organic pollution, toxicity or physically damaged’. Here, the MTR value discriminating between (1) and (2) was exemplarily used as the good-moderate ecological status boundary.

2.2.3 Description of biotic metrics analysed to provide “common macrophyte metrics”

70 macrophyte metrics were analysed to detect “common metrics” enabling intercalibration of national assessment methods (Table 15). These metrics cover the categories “richness and diversity”, “composition and abundance”, “sensitivity and tolerance”, and “ecosystem function”. The basic criterion for the selection of common

metrics was a correlation ($R^2 > 0.5$; $p < 0.05$) of the metric with all assessment methods evaluated in this study. As an additional criterion, redundant metrics were excluded from further analysis. Of metric pairs with a coefficient of determination of > 0.65, the metric showing the lesser correlation with the assessment methods was omitted.

Table 14: Class boundaries of the national assessment methods and derived reference values using the 95th percentile value of all survey sites (n.a. – not applicable).

Index	IBMR (FR)	RI (DE)[3]	DMS (NL)	MTR (UK)
High – good	15	0.5	0.8	n.a.
Good – moderate	12	0.25	0.6	66[4]
Moderate - poor	9	0.15	0.4	n.a.
Poor - bad	7	0	0.2	n.a.
Literature source	NF T90-395 (2003)	Schaumburg et al. (2005)	van den Berg et al. (2004)	Holmes et al. (1999)
Reference (95th percentile)	13.2	0.86	0.42	60.4

2.2.4 Data preparation

The national assessment methods were manually calculated for each macrophyte sample, with the exception of DMS (NL), which was calculated by the software QBWat (Pot, 2005). Due to the minimum criteria for confidence specified by the German and Dutch indices, they could not be determined for 15 and 9 sites, respectively. The index values were converted into Ecological Quality Ratios, i.e. dividing the observed score of each site by a reference value to normalise the output. The 95th percentile value of all samples was chosen as index reference assuming that approximately five percent of surveyed sites hold macrophyte communities in reference state.

2.2.5 Correlation and regression analysis: macrophyte assessment methods, potential common metrics and pressure gradients

The relationships between the four assessment indices were analysed and the strength of correlation was specified by the “coefficient of determination” (R^2). This measure was also used to determine common macrophyte metrics suitable for intercalibration. Both linear and nonlinear regression was tested using the software TableCurve 2D (SYSTAT Software Inc., 2002).

[3] Classification scheme relates to sites where only the Reference Index provides validated results within the assessment method for aquatic flora.
[4] Boundary based on recommendations for the interpretation of MTR scores to evaluate the trophic state (Holmes et al., 1999; see text for details).

Table 15: Metrics tested with the macrophyte dataset. For taxa assignment to growth forms refer to Table 18 (# taxa - number of taxa, % - relative abundance, ca - composition/ abundance, f - functional, rd - richness/diversity, st - sensitivity/tolerance).

Name of metric	Metric type	Reference
Proportion of community with preference for certain amount of water supply		
Typical macrophytes (# taxa and %)	f / rd / ca	Holmes et al. (1999)
Species submerged (# taxa and %)	f / rd / ca	
Species amphibious (# taxa and %)	f / rd / ca	Szoszkiewicz et al. (2006a)
Mosses and liverworts (# taxa and %)	f / rd / ca	
Species terrestrial (# taxa and %)	f / rd / ca	
Diversity indices		
Shannon diversity	rd	Shannon & Weaver (1949)
Simpson diversity	rd	Simpson (1949)
Evenness	rd	Pielou (1966)
Shannon diversity (growth forms)	rd	following Wiegleb (1991), van de Weyer (2003)
Evenness (growth forms)	rd	
Morphological groups according to growth forms		
Species anchored but with floating leaves or heterophyllus (# taxa and %)	f / rd / ca	Szoszkiewicz et al. (2006a)
Species floating free (# taxa and %)	f / rd / ca	
Growth forms (# taxa and %)	f / rd / ca	
Growth form Myriophyllids (# taxa and %)	f / rd / ca	Wiegleb (1991), van de Weyer (2003)
Growth form Parvopotamids (# taxa and %)	f / rd / ca	
Growth form Peplids (# taxa and %)	f / rd / ca	
Growth form Vallisnerids (# taxa and %)	f / rd / ca	
Reference and disturbance indicating taxa and growth forms of lowland streams		
Disturbance indicating taxa (# taxa and %)	st	van de Weyer (2003)
Reference taxa (# taxa and %)	st	
Reference growth forms (# taxa and %)	st	
Disturbance indicating growth forms (# taxa and %)	st	
Ratio: reference taxa to disturbance indicating taxa (# taxa and %)	st	
Disturbance indicating growth form: Elodeids (# taxa and %)	st	
Disturbance indicating growth form: Lemnids (# taxa and %)	st	
Disturbance indicating growth form: Myriophyllids (# taxa and %)	st	
Disturbance indicating growth form: Parvopotamids (# taxa and %)	st	
Disturbance indicating growth form: Peplids (# taxa and %)	st	following van de Weyer (2003)
Reference growth form: Batrachids (# taxa and %)	st	
Reference growth form: Ceratophyllids (# taxa and %)	st	
Reference growth form: Magnonymphaeids (# taxa and %)	st	
Reference growth form: Magnopotamids (# taxa and %)	st	
Reference growth form: Myriophyllids (# taxa and %)	st	
Reference growth form: Parvopotamids (# taxa and %)	st	
Reference growth form: Peplids (# taxa and %)	st	
Selected reference taxa (*Potamogeton natans*, *P. polygonifolius*, *Nuphar lutea*, *Sagittaria sagittifolia*, *Sparganium emersum*, *Berula erecta*) (# taxa and %)	st	
Ratio: reference growth forms to disturbance indicating growth forms (# taxa and %)	st	
Nitrogen indicating metric		
Ellenberg_N	st	Ellenberg et al. (1992)

Physical-chemical, hydromorphological and land use/type data were used to construct complex stressor gradients by means of principle components analysis (PCA). General degradation gradients were derived from physical-chemical, hydromorphological and land use data. In addition, separate degradation gradients were constructed via PCA, using water chemistry, hydromorphological and microhabitat data. (see Hering et al., 2006a). The results were used to test the response of the macrophyte methods to individual pressure groups. Gradients best correlating to the macrophyte assessment methods were determined.

2.2.6 Comparison of quality class boundaries

Two intercalibration approaches were applied in this study: (1) National quality classes of the macrophyte methods were compared directly following the procedure described in Chapter 1 of this thesis. The assessment method showing the highest correlation to all other indices was used as a “common scale”. (2) The approach of indirect boundary comparison (Buffagni et al., 2006) employed common metrics as response variables in the regression analysis.

2.3 Results

2.3.1 Comparison of classification schemes

The classification results of the four methods applied differed noticeably. According to the German method more than 50 percent of sites were in high and good status. The French, British and Dutch methods assessed nearly all sites as of moderate or worse quality (Figure 3).

Due to the different range of quality covered by the individual methods, the 95th percentile value chosen as the reference value was allocated to different quality classes for each of the four national classification schemes (Table 14): The reference value was allocated as high quality in the German RI system, good quality in the French IBMR system, and moderate quality in the Dutch DMS and British MTR systems. Nevertheless, the reference obtained in the analysis for the British MTR corresponded to the mean of top 10 percent MTR values for similar British lowland river types given by Holmes et al. (1999).

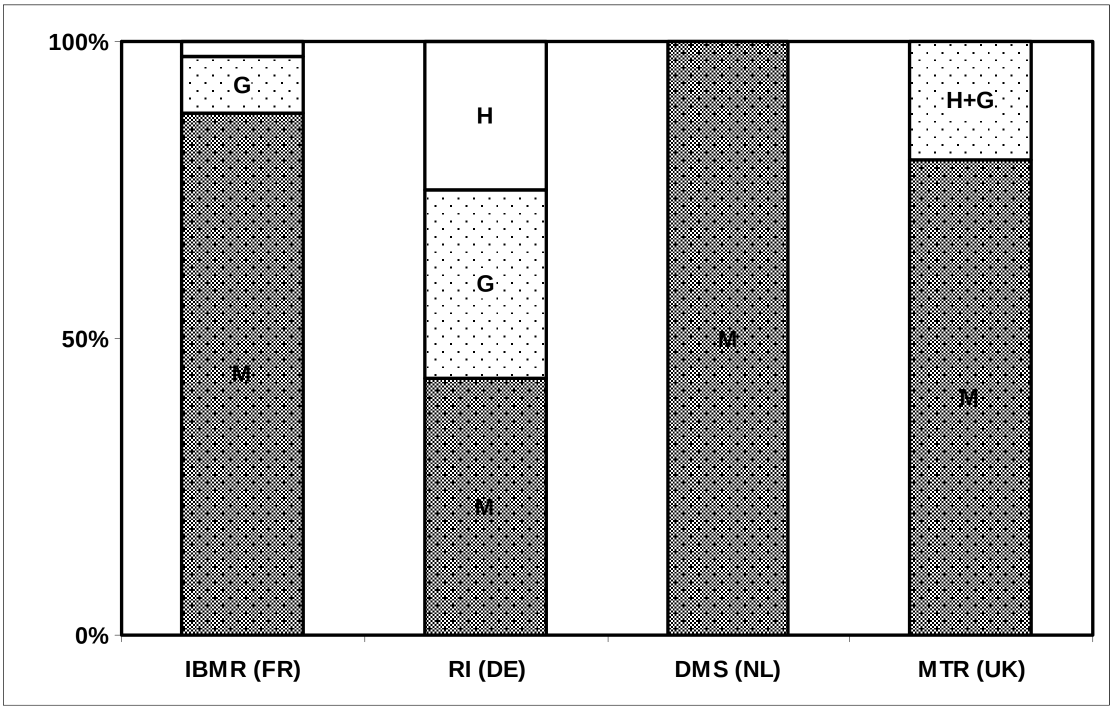

Figure 3: Distribution of quality classes in the dataset resulting from four macrophyte assessment methods (H – high; G – good; M - moderate and worse). Quality classes of RI (DE) are based on the analysis of the Reference Index and additional criteria (Schaumburg et al., 2004). The class boundary between high/good (H+G) and moderate quality of MTR (UK) is based on recommendations for the interpretation of MTR scores to evaluate the trophic state (Holmes et al., 1999; see text for details).

2.3.2 Correlation and regression analysis

Macrophyte assessment methods

The coefficients of determination given in Table 16 revealed the differences between the four assessment methods. The French and British methods were most closely related (R^2 > 0.75). The German RI showed lower correlations with these methods, especially with the French IBMR, while DMS (NL) was negatively correlated to all other methods. Nonlinear regression generally resulted in higher coefficients of determination. Between RI (DE) and MTR (UK) the difference between the two regression models was R^2 = 0.12.

Potential common macrophyte metrics

Of the 70 biotic macrophyte metrics tested, only Ellenberg_N correlated significantly to all four assessment methods. For all four assessment methods nonlinear regression yielded higher coefficients of determination to Ellenberg_N than linear regression.

While IBMR (FR), RI (DE) and MTR (UK) were negatively correlated to this metric, the Dutch index values were positively related to Ellenberg_N.

None of the other biotic metrics showed strong correlations with all four macrophyte assessment methods. For example, the richness measure “number of species” was strongly related to the German and Dutch methods (Table 16). However, due to the type of relation to the German RI (Figure 4) it could not be considered as a common macrophyte metric, since the regression function was non-monotonic. Thus, for each normalised value for “number of species”, two values of the German RI were possible.

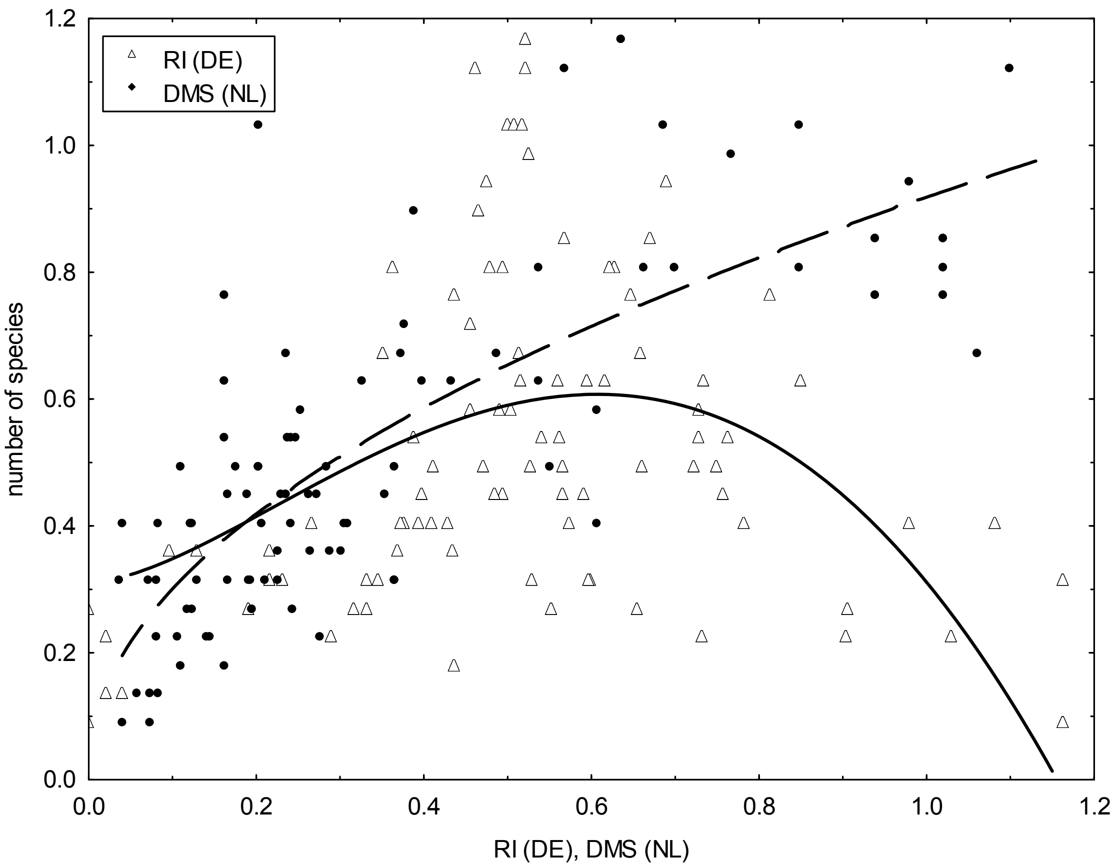

Figure 4: Nonlinear regression of German RI (solid line; R^2 = 0.28) and Dutch DMS (dashed line; R^2 = 0.59) against the number of species.

Table 16: Correlation and regression analysis of macrophyte assessment methods, selected macrophyte metrics and environmental gradients: Type of correlation (pos. - positive, neg. - negative) and coefficients of determination (R^2) based on linear and nonlinear regression. Nonlinear R^2 is only given if providing higher coefficients of determination ($p < 0.05$; n.s. – not significant).

	IBMR (FR)			RI (DE)			DMS (NL)			MTR (UK)		
	Type	Linear	Nonlinear	Type	Linear	Nonlinear	Type	Linear	Nonlinear	Type	Linear	Nonlinear
Macrophyte assessment methods												
IBMR (FR)	pos.	1.00	1.00	pos.	0.22	0.31	neg.	0.06	-	pos.	0.76	0.79
RI (DE)	pos.	0.22	0.26	pos.	1.00	1.00	-	n.s.	n.s.	pos.	0.41	-
DMS (NL)	neg.	0.06	0.10	-	n.s.	0.15	pos.	1.00	-	neg.	0.05	0.07
MTR (UK)	pos.	0.76	0.77	pos.	0.41	0.53	neg.	0.05	-	pos.	1.00	-
Selected macrophyte metrics												
Ellenberg_N	neg.	0.46	0.56	neg.	0.46	0.58	pos.	0.05	0.11	neg.	0.69	0.70
Number of species	neg.	0.04	0.06	-	n.s.	0.28	pos.	0.56	0.59	-	n.s.	0.06
Disturbance indicating growth forms (%)	neg.	0.07	-	-	n.s.	n.s.	pos.	0.56	-	neg.	0.05	n.s.
Environmental gradients												
Pollution/ eutrophication	neg.	0.46	-	neg.	0.14	0.22	pos.	0.09	0.10	neg.	0.51	0.52
General degradation	-	n.s.	n.s.	-	n.s.	n.s.	neg.	0.41	0.42	-	n.s.	n.s.

The DMS (NL) showed coefficients of determination of $R^2 > 0.5$ with several functional metrics (e.g. “relative abundance of disturbance indicating growth forms”, “relative abundance of disturbance indicating growth form: Lemnids” and “number of selected reference taxa”).

Environmental gradients (PCA)

The French, German and British methods related most strongly to the PCA gradient reflecting water chemistry (“pollution/eutrophication”, PCA axis 1, Eigenvalue: 0.527; Table 16). The Dutch method was correlated with “general degradation” including chemical, hydromorphological and land use parameters (PCA axis 1, Eigenvalue: 0.287). Coefficients of determination of the regression analysis are listed in Table 16 (see Hering et al., 2006a for details of the gradients).

Table 17: EQR values of the high-good (H|G) and good-moderate (G|M) quality class boundaries transferred into MTR and Ellenberg_N scales via nonlinear regression analysis. For each value derived by regression the 95 percent confidence interval is specified (n.a. – not applicable). (1) $f(x) = a + b \cdot x^{1.5}$; (2) $f(x) = a + b \cdot x^3$

Class boundary	Common scale	IBMR (FR)			RI (DE)			MTR (UK)		
		Equation	Boundary value	*95% confid.*	Equation	Boundary value	*95% confid.*	Equation	Boundary value	*95% confid.*
H|G	MTR	(1)	1.497	0.150	(2)	0.638	0.056	n.a.		
	Ellenberg_N	(2)	1.185	0.287	(2)	0.394	0.079			
G|M	MTR	(1)	0.820	0.044	(2)	0.565	0.067	-	1.094	-
	Ellenberg_N	(2)	0.638	0.103	(2)	0.294	0.094	(1)	0.911	0.143

2.3.3 Direct comparison of quality class boundaries

The British MTR correlated best with all other methods and was therefore used as the “common scale” according to Chapter 1 of this thesis. Due to its weak relationship with any of the other macrophyte methods, the Dutch DMS was not included in direct class boundary comparison. Considering the 95 percent confidence intervals, direct comparison revealed large differences in national definitions of the high-good quality boundary (> 0.6 MTR units on average, Table 17). The differences between the good-moderate boundaries were smaller (< 0.3 MTR units on average). The mean value of confidence intervals amounted to 0.079 MTR units.

The nonlinear regression graph (Figure 5) shows decreasing slope values with increasing deviation of IBMR (FR) and RI (DE) from the reference state. Especially in the lower range of the RI (DE), the British MTR was not responding to changes of the German method. Therefore, the high-good and good-moderate class boundary intervals of RI (DE) transferred into MTR scale were overlapping (cf. Table 17).

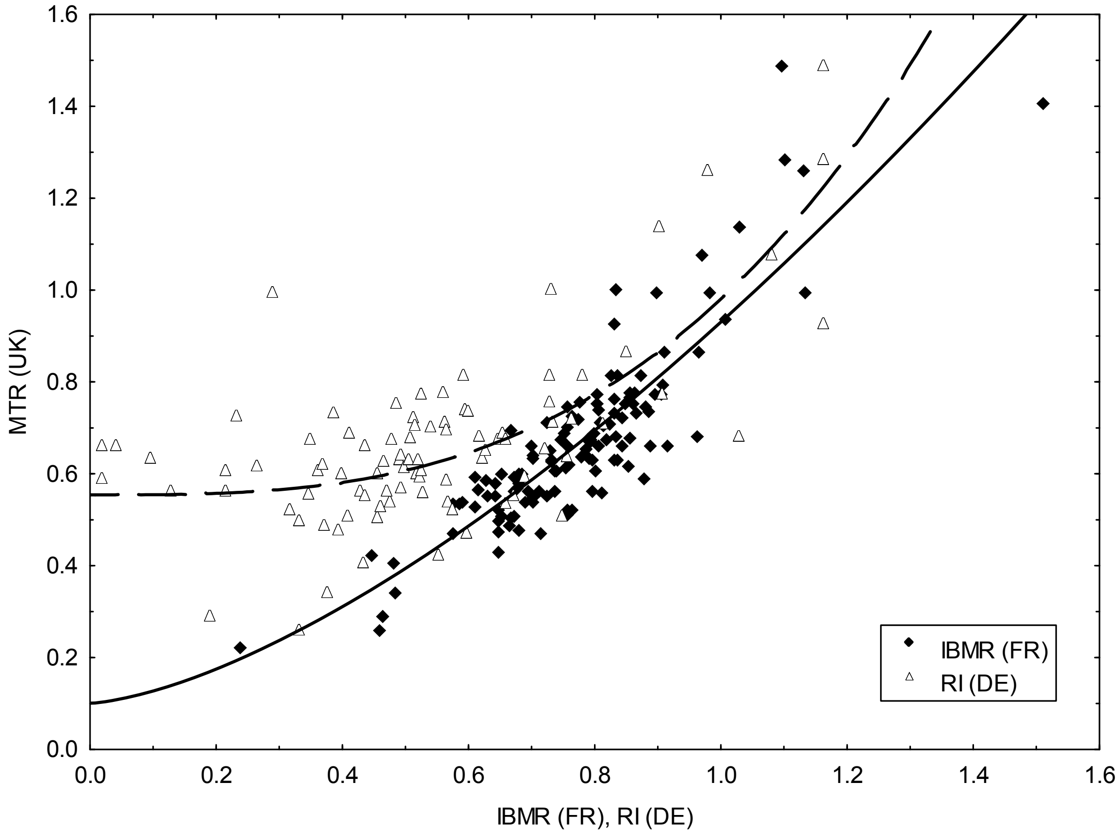

Figure 5: Nonlinear regression of French IBMR (solid line; R^2 = 0.77) and German RI (dashed line; R^2 = 0.53) against British MTR.

2.3.4 Indirect comparison of quality class boundaries using Ellenberg_N as common macrophyte metric

The high-good boundary comparison of the French and German method using Ellenberg_N resulted in a difference of > 0.4 units. For the German and British method, confidence intervals of the good-moderate class boundaries shared similar ranges when compared via Ellenberg_N. The average confidence interval amounted to 0.141 units.

As observed in the “direct comparison approach” the quality class boundaries of RI (DE) showed overlapping confidence ranges using Ellenberg_N (Table 17). Regression analysis disclosed a similar type of relation between the German method and each of MTR and Ellenberg_N (Figure 6).

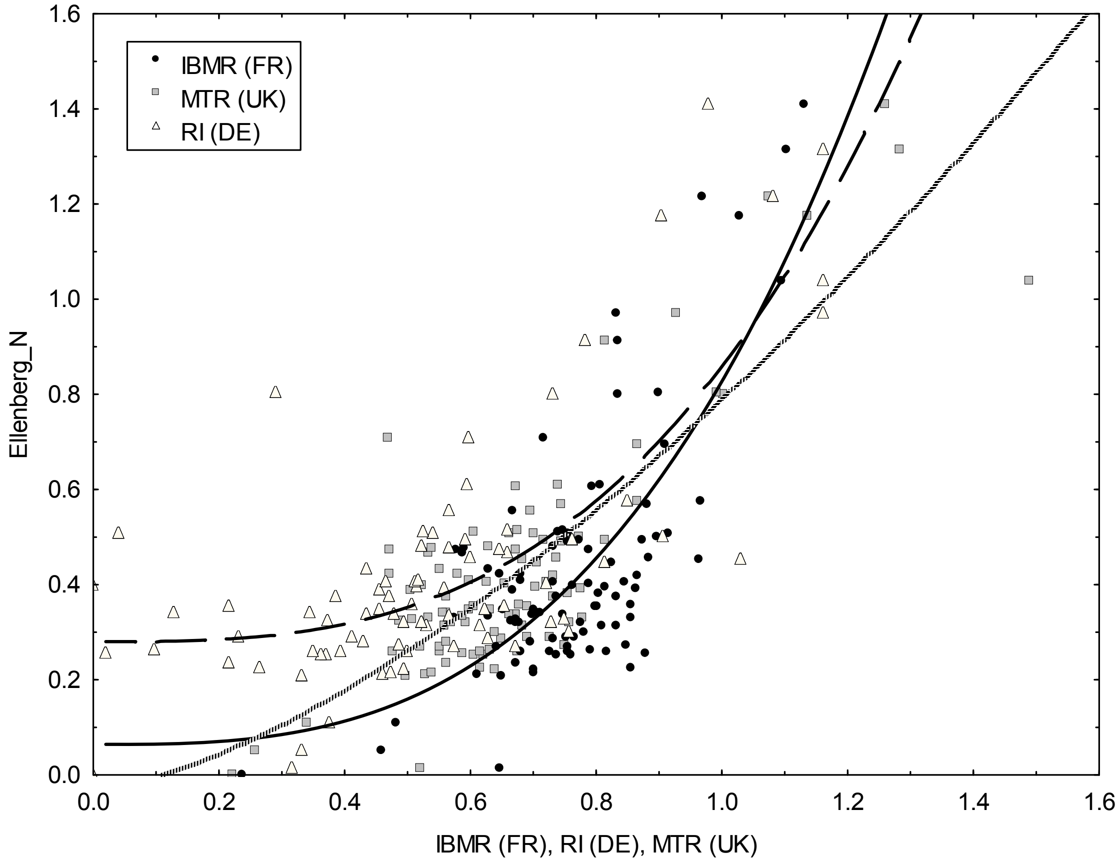

Figure 6: Nonlinear regression of French IBMR (solid line; R^2 = 0.56), German RI (dashed line; R^2 = 0.58) and British MTR (dotted line; R^2 = 0.70) against Ellenberg_N.

2.4 Discussion

The obviously different quality classes of the sites assessed with the four methods (Figure 3) reveal that intercalibration efforts for macrophyte methods are indispensable. Starting from this conclusion I applied analytical methods currently used in intercalibration of benthic invertebrate systems (Buffagni et al., 2006; see also Chapter 1 of this thesis) to compare quality class boundaries of macrophyte assessment methods.

2.4.1 Testing of intercalibration approaches

This study discloses difficulties in adopting commonly used intercalibration approaches to macrophyte based assessment methods. Direct comparison of class boundaries only yielded sound results between the closely related methods IBMR (FR) and MTR (UK). These two indices share many common indicator species (Szoszkiewicz et al., 2006a) whose indicator values correlate strongly (R^2 = 0.61). The sound correlation of the German RI with MTR (UK) seems to allow for direct boundary comparison. However, the specific nonlinear character of this relationship impedes significant resolution between the good quality boundaries, thus making direct comparison of these methods impossible. The low correlations of DMS (NL) with all other national assessment methods exclude this index from further intercalibration analysis.

Against this background I tested whether intercalibration could be accomplished using common metrics (Buffagni et al., 2006). My analysis showed that none of the tested macrophyte metrics met the common metric criteria for all assessment methods. Only Ellenberg_N showed strong relationships with at least three methods. The metric is based on the response of higher plants to nitrogen compounds (nitrate and/or ammonium) and, thus, corresponds to trophic categories and to general nutritional conditions in the rivers that are represented with a broad gradient in the dataset (see Table 11). The French, German and British methods relating to this potential common metric also respond significantly to the abiotic PCA gradient reflecting organic pollution and eutrophication. These findings underline the general ability of macrophyte methods to assess the trophic status of rivers. While Holmes et al. (1999) designed the British MTR for this specific purpose, the German method in particular is aimed at detecting “general degradation”, i.e. the level of deviation from a reference community (Schaumburg et al., 2004).

Like in direct comparison, DMS (NL) cannot be included in the intercalibration analysis using common metrics. Although it shares objectives with the German method (unspecific pressure assessment based on type specific macrophyte communities), I found no biotic metric suitable for intercalibration. Either different types of relation (cf. Figure 6) or no common relationship at all, limits the applicability of the common metric approach. DMS (NL) is characterised by strong relations to richness and diversity measures and, most remarkably, by positive correlations with metrics indicating disturbance in lowland streams of North-Rhine Westphalia (Western Germany; Table 18; van de Weyer, 2003). In this respect, the broad spectrum of environmental factors influencing the occurrence of macrophytes in streams on various spatial scales (Wiegleb, 1988) may confine the validity of indicator species to narrow geographic regions. Furthermore, Korte & van de Weyer (2005) observed that, in two separate methods for the assessment of German lowland streams, indicative characteristics of macrophyte species are evaluated differently. Nevertheless, the weak but significant positive correlation of the Dutch method with Ellenberg_N points at basic differences in the conception of the reference state. This is also indicated by the negative correlation of DMS (NL) to the "general degradation" gradient. However, since the dataset analysed covers only sites of moderate or worse status according to the Dutch classification system the validity of my findings is limited to a restricted range of quality.

Further incomparability results from the different calculation methods. The French, German and British indices are calculated by weighted average equations, yielding values less influenced by the species richness of the site. Abundance scores are accounted by multiplication by the indicator values. Results of DMS (NL) are obtained by summation of taxa scores, whose values depend on the relative abundance of the species. For certain species in specific river types this score value decreases with increasing abundance and vice versa.

National assessment methods for all biological quality elements will need to assess ecological quality in a general way; therefore, the intercalibration exercise of macrophyte-based methods has to simultaneously target the effects of different types of degradation. The selection of common intercalibration metrics should thus respond to general degradation (see also Buffagni et al., 2005 using common metrics for intercalibration of invertebrate-based methods). I tested a broad range of general and specific macrophyte metrics covering biotic parameters like taxonomic composition and abundance, richness and diversity, and functional groups (Table 15). Since none of the metrics analysed qualified for intercalibration purposes, further research to produce suitable common macrophyte assessment metrics is indispensable.

Table 18: Reference taxa and disturbance indicating taxa of lowland streams and their growth forms (following van de Weyer 2003).

Reference taxa	Growth form
Chara fragilis Desvaux	Charids
Chara sp. L. ex Vaillant	
Nitella flexilis C. A. Ag.	
Nitella sp. C. A. Ag.	
Ceratophyllum submersum L.	Ceratophyllids
Berula erecta (Huds.) Coville	Herbids
Juncus bulbosus L.	
Nuphar lutea (L.) Sibth. & Sm.	Magnonymphaeids
Nymphaea alba L.	
Persicaria amphibia (L.) Gray	
Ranunculus flammula L.	
Potamogeton alpinus Balbis	Magnopotamids
Potamogeton gramineus L.	
Potamogeton natans L.	
Potamogeton polygonifolius Pourret	
Potamogeton lucens L.	
Potamogeton obtusifolius Mert. & Koch	
Potamogeton perfoliatus L.	
Potamogeton praelongus Wulfen	
Myriophyllum alterniflorum DC.	Myriophyllids
Myriophyllum verticillatum L.	
Utricularia intermedia Hayne	
Utricularia vulgaris L.	
Potamogeton berchtoldii Fieber	Parvopotamids
Potamogeton compressus L.	
Potamogeton filiformis Pers.	
Callitriche cophocarpa Sendtn.	Peplids
Callitriche hamulata Kutz ex W.D.J. Koch	
Callitriche platycarpa Kütz.	
Alisma plantago-aquatica L.	Vallisnerids
Sagittaria sagittifolia L.	
Sparganium emersum Rehmann	
Sparganium erectum L.	
Sparganium sp. L.	
Disturbance indicating taxa	
Ceratophyllum demersum L.	Ceratophyllids
Ceratophyllum demersum var. apiculatum Cham.	
Elodea canadensis Michx.	Elodeids
Lemna gibba L.	Lemnids
Lemna minor L.	
Spirodela polyrhiza (L.) Schleid	
Myriophyllum spicatum L.[5]	Myriophyllids
Ranunculus fluitans Lamk.[5]	
Potamogeton crispus L.	Parvopotamids
Potamogeton pectinatus L.	
Potamogeton pusillus L.	
Potamogeton trichoides Cham. & Schltdl.	
Zannichellia palustris L.	
Callitriche obtusangula Le Gall	Peplids

5 According to van de Weyer (2003) these species indicate increased current velocity (e.g. due to channel straightening).

2.4.2 Implications for the macrophyte intercalibration exercise

This chapter presents preliminary results which may become relevant in the further discussion of macrophyte intercalibration. Nevertheless, several procedural requirements of the official intercalibration exercise are not met: (1) The international STAR dataset covers different biogeographical regions. Therefore, the applicability of national methods may be affected because the assessment is adjusted to the regional flora and the indicative characteristics of its macrophyte species. (2) Since macrophytes were surveyed according to a standard procedure (Furse et al., 2006), national survey techniques and their effect on the taxa list are neglected. (3) I based my analyses on macrophytes only, whereas two of the methods examined, RI (DE) and DMS (NL) are designed to assess the broader “aquatic flora”, including phytobenthos. Considering these items, the following implications for the macrophyte intercalibration exercise can, however, be stated.

The comparison of quality classes for European river assessment methods using benthic invertebrates was successfully accomplished (Owen et al., 2010). This can substantially be attributed to the strong relationships of the methods (see Chapter 1 of this thesis), their focus on similar pressures and their common tradition. In the view of the present study, the practicability of the analytical approaches applied to the intercalibration of macrophyte methods (direct comparison, use of common metrics) is questionable. Two main factors that complicate comparisons between methods are (1) differently defined reference conditions and (2) gaps in knowledge about pressure-impact relationships. The delineation of reference communities, particularly for the medium-sized lowland rivers of Central Europe, is difficult due to the lack of existing reference sites. Therefore, expert opinion is used to estimate natural conditions in the lowlands. Furthermore, the Dutch and German methods both define reference states via index scores, but include diverse macrophyte species and apply different formulae. The lack of knowledge about pressure-impact relationships may generally impede the intercalibration of macrophyte methods. While higher plants are well known for their response to nutrient pollution, the effect of other impairments on the community is has little been studied (Kelly & Whitton, 1998; Janauer, 2001). This also delimits the availability of appropriate common assessment metrics. This study demonstrates on the one hand that intercalibration of methods specifically addressing eutrophication is possible but, on the other hand, it also highlights deficiencies for the coming macrophyte intercalibration exercise.

Since the intercalibration of national methods has to be accomplished by 2011, scientific activities at the European level are currently being carried out to fulfil these legal requirements. Thus, the intercalibration task has initiated a process of Europe-wide discussion on ecological quality and the harmonization of its assessment. Tailor-made approaches for each biological element are required relying on national expertise and international coordination. As a first step towards intercalibration of macrophyte methods in Central Europe, I propose to compile an international database including national data on macrophytes and abiotic pressures taken from sites at common intercalibration types. Since field procedures of the countries involved are very similar (visual survey of 100 m stream sections), this will enable more extensive analyses of the relation between the assessment indices and the definition of the reference state. The outcome may necessitate detailed bilateral discussion on the assessment results at individual sites. This time consuming approach has already yielded results in a preliminary intercalibration study between macrophyte methods of Austria and Germany (Pall et al., 2005).

With regard to the multitude of issues to be addressed in the near future, intercalibration represents a major chance for the implementation of harmonized quality standards at the European level beyond the short timeframe given by the Directive. For macrophyte based ecological quality assessment in particular, which is still in its early stages in Europe, communality can be gained by maintaining and extending international collaboration to enhance scientific exchange and trigger common outputs.

3 Towards harmonization of ecological quality classification: establishing common grounds in European macrophyte assessment for rivers

3.1 Introduction

River macrophyte communities are determined by the characteristics of the local habitat in which they occur, namely light availability, current velocity, sediment patterns and nutrient supply. Biogeographical zone, catchment geology and stream hydrology establish the large-scale framework influencing occurrence and abundance of macrophytes (Lacoul & Freedman, 2006). Since most of these factors can be subject to anthropogenic alteration, macrophytes are effective bioindicators that respond to various human pressures by a change in cover, richness or taxonomical composition (e.g. Baattrup-Pedersen & Riis, 1999, Ferreira et al., 2005, Szoszkiewicz et al., 2006a).

Combined with benthic microalgae, macrophytes thus form an obligatory element in the monitoring of ecological river quality as stipulated by the EU Water Framework Directive 2000/60/EC (WFD). For discrete stream types the taxonomical composition and abundance of macrophytes are appraised by biological assessment methods. The status observed at the monitored river stretch is compared to the status expected under near-natural conditions. The resulting Ecological Quality Ratio (EQR) evaluates the river quality in a score ranging from 0 (worst status) to 1 (reference status). This range is divided into five classes of ecological quality: high, good, moderate, poor and bad.

The WFD requires that all water bodies should attain good ecological status within the near future. However, countries obliged to fulfil these requirements are applying different assessment methods. To set a common level of ambition in reaching the WFD's objective good ecological status is harmonized through the so-called "intercalibration exercise" (Heiskanen et al., 2004). The specific challenge of this exercise is to calibrate the national interpretations of good ecological status. Although the WFD provides general guidelines for the high, good and moderate quality status, the practical implementation of these normative definitions has to be compared between the various countries.

For benthic diatoms (Kelly et al., 2008) and benthic invertebrates (Owen et al., 2010) the national assessment methods were intercalibrated by the use of common metrics (Buffagni et al., 2007). These metrics allow the national definitions of good ecological status to be compared across different countries and stream types. Common metrics take advantage of similar assessment principles that all national methods have in common. For instance, anthropogenic pressure generally causes a decrease of taxa richness in invertebrate communities, making the total number of taxa a suitable common metric (Buffagni et al., 2005). However, in Chapter 2 I reported on difficulties in finding common metrics for the intercalibration of river macrophyte methods. Differing national assessment concepts were identified: Some countries focus on the appraisal of specific pressures, especially nutrient enrichment, while others emphasize assessment of general degradation.

This fundamental difference raises the question, if any conceptual similarities exist between the national macrophyte classifications. Expert discussions confirmed a common notion of type-specific macrophyte communities at high ecological status (Birk et al., 2007a). Motivated by this finding, the present study investigates whether this common notion can be empirically defined. In particular, my work is based on the following hypotheses:

- Certain macrophyte communities that occur in a common stream type are classified in high quality status by the majority of national assessment methods.
- These communities feature species that are regarded as indicators of near-natural conditions across national methods.
- Following this concept general disturbance indicators can also be identified based on data of communities commonly classified in poor or bad quality status.

These hypotheses were tested by applying seven macrophyte assessment methods to an international dataset that covered three European stream types. I correlated the abundance of individual macrophyte taxa to the average national EQR per survey and thus gained indicators of a common high or poor quality status. Based on these outcomes I

(1) describe the macrophyte communities of each stream type under near-natural and degraded conditions,

(2) develop a common macrophyte metric and relate it to the national methods,

(3) propose amendments to certain national methods in order to improve their relationship with the common metric and

(4) identify type-specific reference values to convert the common metric scores into EQRs.

3.2 Methods

3.2.1 Data basis

In this study data on taxonomic composition and abundance of river macrophytes was used. Sampling sites were located on rivers belonging to three common stream types (ECOSTAT, 2004b) that were shared by twelve countries in Central and Western Europe and the Baltic region (Table 19). The common stream types were delineated by their altitude, catchment size, geology, substrate composition and alkalinity. The types covered small to medium-sized streams in the lowlands and small streams in the mountains.

Table 19: Characterisation of the common stream types

Stream type abbreviation	Common stream type	Catchment area [km^2]	Altitude [m]	Geology	Channel substrate	Alkalinity [meq/l]
R-C1x2	Small lowland sandy streams	10 – 100	< 200	Siliceous	Sand	> 1
R-C3	Small mid-altitude siliceous streams	10 – 100	200 - 800	Siliceous	Boulders, cobbles and gravel	< 0.4
R-C4x2	Medium-sized lowland steams	100 - 1000	< 200	Mixed	Gravel and sand	> 2

In total, 609 macrophyte surveys were provided by the countries listed in Table 20. The data originated from national monitoring programmes or scientific projects (e.g. Furse et al., 2006). Countries applied national macrophyte survey protocols that were in line with the requirements of the European Standard EN 14184:2003. Representative river stretches were visually inspected by wading, diving or boating, using rake, grapnel or aqua-scope where necessary. Representative sites spanned about 100 metres of river length.

The macrophyte abundance was recorded in different scales (Table 21). Most countries specified the abundance as relative coverage of the surveyed area. Percent values were graded into five, seven or nine classes. The Austrian method combined the number of single plant records per surveyed section and the plant quantity per habitat following Kohler (1978) in a five-class scheme. Germany estimated the plant quantity (Melzer et al., 1986) in one of five different classes. The Dutch abundance data were given in various scales (Braun-Blanquet, 1928, Tansley, 1946).

Table 20: Number of macrophyte surveys used in the analysis, listed per country and common stream type

Stream type abbreviation	Country	Number of surveys
R-C1x2	Belgium (Flanders)	105
	Belgium (Wallonia)	1
	Denmark	15
	Germany	38
	Latvia	15
	Lithuania	1
	Netherlands	14
	Poland	11
	Total number	*200*
R-C3	Austria	31
	Belgium (Wallonia)	43
	Czech Republic	13
	France	78
	Germany	81
	Great Britain	33
	Total number	*279*
R-C4x2	Belgium (Flanders)	15
	Denmark	4
	Germany	32
	Great Britain	3
	Latvia	29
	Lithuania	9
	Luxemburg	3
	Netherlands	8
	Poland	27
	Total number	*130*

3.2.2 National assessment methods

Seven countries participated in this exercise with their national assessment methods (Table 22). Most methods focused on the assessment of specific human pressure (Austria, France, Great Britain, Poland, Wallonia). The principal component of this approach was formed by a list of indicator taxa graded by their sensitivity, mainly to nutrient enrichment. Numerical assessment results were obtained by computing a sensitivity metric, i.e. the average score of indicative species weighted by their abundance. In case of the Austrian, French and Wallonian metrics this also included a factor considering the taxon's ecological amplitude.

Table 21: Conversion table of national macrophyte abundance classes into the international abundance scale

<table>
<tr><th rowspan="2">Country</th><th colspan="12">International abundance scale</th></tr>
<tr><th colspan="2">1st class (rare)</th><th colspan="2">2nd class (occasional)</th><th colspan="2">3rd class (frequent)</th><th colspan="2">4th class (abundant)</th><th colspan="4">5th class (very abundant)</th></tr>
<tr><td>Austria
Belgium (Wallonia)
Czech Republic
Denmark
France
Germany
Latvia
Lithuania
Luxembourg</td><td colspan="2">1</td><td colspan="2">2</td><td colspan="2">3</td><td colspan="2">4</td><td colspan="4">5</td></tr>
<tr><td>Belgium (Flanders)</td><td colspan="2">1</td><td colspan="2">2</td><td colspan="2">3</td><td>4</td><td>5</td><td colspan="2">6</td><td colspan="2">7</td></tr>
<tr><td>Great Britain
Poland</td><td colspan="2">1</td><td colspan="2">2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td colspan="2">8</td><td>9</td></tr>
<tr><td>The Netherlands (Braun-Blanquet, 1928)</td><td>1</td><td>2</td><td>3</td><td>4</td><td colspan="2">5</td><td>6</td><td>7</td><td colspan="2">8</td><td colspan="2">9</td></tr>
<tr><td>The Netherlands (Tansley, 1946)</td><td colspan="2">1</td><td colspan="2">2</td><td>3</td><td>4</td><td>5</td><td>6</td><td colspan="2">7</td><td colspan="2">8</td></tr>
</table>

The Flemish and German methods were oriented towards the indication of non-specific anthropogenic disturbance. Besides sensitivity measures these methods considered additional metrics, such as richness of macrophyte growth forms, or taxa richness and dominance. The basic element of the German Reference Index (RI) was the type-specific definition of reference and non-specific disturbance indicating taxa. The RI was a numerical expression of the relation of both response groups at a river site. The supplementing assessment criteria directly contributed to the score of the RI. The Flemish method integrated three metrics in the appraisal of ecological status by the "one out – all out" principle: The type-specific index for water vegetation, the perturbation index for water vegetation and the richness of various growth forms. Based on the experiences gained in earlier intercalibration studies (Birk et al., 2007a) the assessment of macrophyte growth form was not considered in the main analysis.

However, I additionally tested the performance of the Flemish method including the growth form metric.

Table 22: National assessment methods using macrophytes in rivers

Country	Name of method	Intercalibrated assessment metric(s)	Relevant stream type(s)	Literature reference
Austria	Austrian Index for Macrophytes in Rivers (AIM Rivers)	Single metric combining ecological preference and abundance	R-C3	BMLFUW (2006a)
Belgium (Flanders)	MAFWAT (Makrophyten Waterlopen)	(1) Type specific index for water vegetation (TSw) (2) Perturbation index (organic pollution, eutrophication) for water vegetation (Vw)	R-C1x2, R-C4x2	Leyssen et al. (2005)
Belgium (Wallonia) France	Indice Biologique Macrophytique en Rivière (IBMR)	Single metric combining occurrence (indicator value per taxon), ecological amplitude and abundance	R-C3, R-C4x2 (only France)	NF T90-395:2003
Germany	Deutsches Bewertungsverfahren für Makrophyten und Phytobenthos (PHYLIB)	Index relating Species Response Groups (Reference, Disturbance, Indifferent) plus additional criteria for R-C3: acidification module R-C4x2: evenness, number of submerged taxa, ratio of *Myriophyllum spicatum* and *Ranunculus* sp.	R-C1x2, R-C3, RC4x2	Schaumburg et al. (2006)
Great Britain	River Nutrient Macrophyte Index (RNMI)	Single metric combining occurrence (indicator value per taxon) and abundance	R-C3, R-C4x2	Willby et al. (2006)
Poland	Macrophyte Index for Rivers (MIR)	Single metric combining occurrence (indicator value per taxon) and abundance	R-C1x2, R-C4x2	Szoszkiewicz et al. (2006b)

3.2.3 Intercalibration analysis

Preparatory steps of the intercalibration analysis comprised the harmonization of the macrophyte taxonomy, especially the identification of synonymous taxon nominations due to different reference literature used by the countries. Furthermore, the abundance data were converted from the national into an international abundance scale (Table 21). A level of aquaticity was assigned to each macrophyte taxon that characterized the taxon's affinity to water (C. Chauvin, pers. comm.). Table 23 provides an overview of the different aquaticity levels used in this study.

The national metrics were applied to the macrophyte survey data. Using the national stream-type specific reference values all metric results were transformed into

Ecological Quality Ratios (EQR). The ecological quality of each survey was classified according to the national methods. Those surveys were identified that the majority of methods classified in high status and none of the methods in moderate or worse status. In the following these surveys are named “common high status sites”.

Table 23: Level of aquaticity characterizing the affinity of the macrophyte taxon to water according to C. Chauvin (pers. comm.)

Level of aquaticity	Description
1	Exclusively aquatic species (or mainly aquatic in regular conditions).
2	Aquatic taxon with common terrestrial forms or truly amphibious (common aquatic forms as well as terrestrial forms).
3	Supra-aquatic bryophytes and lichens. Commonly submerged a part of the hydrological cycle.
4	Helophytes or Amphiphytes. Erected forms with basis commonly inside water.
5	Hygrophilous taxa. Possibly submerged (at least the basis) a part of the year.
6	Bank, wood, grasslands or ruderal herbaceous species. May be found in water accidentally or in conditions of high flow.
7	Woody riparian species. May be flooded temporarily.
8	Brackish water or salty marshes species.

For each common stream type the values of the national EQRs were normalized to a scale ranging from 0 to 1. These normalized values were then averaged for each survey. In case of the two Flemish sensitivity metrics the lowest value per survey was taken (“worst case”) according to the national protocol (Leyssen et al., 2005). As a result, a mean index score was assigned to each survey that was composed of the average of normalized national metric values. Each national metric therefore had an equal contribution to the mean index score.

In a next step this mean index was correlated with the abundance of macrophyte taxa recorded in the surveys using the international abundance scale and including zero abundance. The relation of taxa abundance to the mean index was quantified by Spearman’s coefficient of correlation. The analysis yielded a coefficient for each taxon and comprised a spectrum of values identifying taxa correlated – either positively or negatively – or not correlated to the average national assessment results. Positive correlation meant: the higher the mean index, the higher this taxon’s abundance. Negative correlation meant: the lower the mean index, the higher this taxon’s abundance.

I used the correlation coefficients to define taxon-specific indicator scores. These scores were assigned only to taxa records at species level except for selected algae and mosses. Taxa with only one record in the database or taxa with an aquaticity level > 5 were given no indicator score. I rescaled the coefficients of scoring taxa based on the maximum or minimum correlation separately for each common stream type. For instance, if the range of correlation coefficients was from -0.3 to +0.5 I rescaled from -0.5 to +0.5 which produced an actual range running from -0.6 to +1.0 with a zero score coinciding with a zero correlation. Based on the indicator scores I could describe the common type specific macrophyte community occurring at reference and degraded conditions.

The indicator scores were used in a common type-specific, weighted average metric, the so-called "macrophyte Intercalibration Common Metric" (mICM) following the terminology of Buffagni et al. (2005):

$$mICM_x = \frac{\sum (s_i * abd_i)}{\sum abd_i},$$

where $mICM_x$ was the macrophyte Intercalibration Common Metric value of a survey at the common stream type x,

s_i was the taxon specific correlation value of the i-th taxon and

abd_i was the international abundance class of the i-th taxon.

The mICM was plotted against each national metric per common stream type. Linear regression models were applied and the resulting coefficients of determination (R^2) were checked. In case of R^2 values < 0.5 I compared the mICM scoring taxa list and the national indicator list. Obvious discrepancies between both lists were adjusted by proposing small amendments to the national list. However, I focussed only on those amendments that allowed for an increase of the R^2 value ≥ 0.5 in the regression analysis. Furthermore, to demonstrate the performance of the Flemish method including the growth form metric the mICM was also correlated with the worst case of the three Flemish metrics.

For each stream type I determined the median mICM value from the pool of common high status sites. This value served as the common stream type-specific reference by which the mICM was transformed into an EQR. To characterize the distribution of mICM EQR values among the common high status sites I calculated the 5th and 10th percentile values. Lower percentiles, such as these have been widely used in invertebrate classification as a statistical basis for the high-good boundary (e.g. Clarke et al., 1996).

3.3 Results

For each common stream type Table 24 displays the range of coefficients resulting from correlating the mean index and taxa abundances. In addition, those taxa best correlated to the mean index (either positively or negatively) are listed. In total, 102 (R-C1x2), 140 (R-C3) and 110 (R-C4x2) indicator taxa were defined. All indicator taxa, their level of aquaticity and rescaled indicator scores are shown in the Appendix.

Table 24: Range of Spearman's correlation coefficients (CorrCoef) and taxa showing highest positive (+) and negative (-) correlation of abundance to the mean index gradient

Stream type	CorrCoef range	Taxa best correlated to the mean index	
		+	-
R-C1x2	0.46 to -0.34	*Callitriche hamulata* Kuetz. ex W.D.J. Koch, *Caltha palustris* L., *Cardamine amara* L.	*Lemna minor* L., *Potamogeton pectinatus* L., *Potamogeton perfoliatus* L.
R-C3	0.61 to -0.53	*Pellia epiphylla* L. Corda, *Racomitrium aciculare* (Hedw.) Brid., *Scapania undulata* (L.) Dum	*Amblystegium riparium* (Hedw.) B.S.G., *Cladophora* sp. Kuetz., *Phalaris arundinacea* L.
R-C4x2	0.55 to -0.51	*Fontinalis antipyretica* Hedw., *Hildenbrandia* sp. Nardo, *Potamogeton alpinus* Balbis	*Lemna minor* L., *Potamogeton pectinatus* L., *Sagittaria sagittifolia* L.

Coefficients of determination (R^2) obtained in the regression analyses of the national metrics against the mICM range from 0.28 to 0.78. On average, the British metric is best correlated with the mICM while the German method shows the weakest overall relationship. Highest R^2 values are gained in the analysis of the mountain type R-C3.

To improve weak relationships of the mICM with the national metrics I adjusted the national indicator lists of Flanders (including additional disturbance taxa in metric Vw for type R-C1x2) and Germany (re-scoring of indicator taxa for type R-C4x2). Both adjustments lead to coefficients of determination ≥ 0.5 in the regression analysis. Adjustments and results of the regression analyses are specified in Table 25.

In total, 111 common high status sites are identified, with most of these relating to the R-C3 mountain rivers (Table 26). The mICM median values of these sites show a clear difference, distinguishing between the two lowland types on one hand and the mountain rivers on the other. However, the percentiles of the mICM EQR value distributions are rather similar, ranging from 0.79 to 0.84. This indicates that the spread of values is consistently narrow, and that the unit of turnover of the reference population (~0.2) would be appropriate for establishing a series of lower class boundaries.

Table 25: Results of the linear regression analysis of mICM against the national metrics R^2 (orig.) – coefficient of determination using national index with original indicator taxa list, R^2 (amend.) – coefficient of determination using national index with amended indicator taxa list

Stream type	Country	R^2 (orig.)	R^2 (amend.)	Specification of amendment
R-C1x2	Belgium (Flanders)	0.28 (0.21)	0.50 (0.24)	Additional disturbance indicators in metric Vw: *Nymphoides peltata* (Gmel.) Kuntze, *Potamogeton berchtoldii* Fieber, *Potamogeton crispus* L., *Potamogeton perfoliatus* L., *Rorippa amphibia* (L.) Besser, *Sagittaria sagittifolia* L., *Sparganium emersum* Rehmann
	Germany	0.55	-	-
	Poland	0.62	-	-
R-C3	Austria	0.74	-	-
	Belgium (Wallonia)	0.77	-	-
	France	0.76	-	-
	Germany	0.54	-	-
	Great Britain	0.78	-	-
R-C4x2	Belgium (Flanders)	0.59 (0.02)	-	-
	Germany	0.17	0.59	Re-scoring of indicator taxa: *Fontinalis antipyretica* Hedw. (Reference), *Lemna minor* L. (Disturbance)
	France	0.58	-	-
	Great Britain	0.62	-	-
	Poland	0.58	-	-

Table 26: Number of common high status sites (N), mICM reference values (REF), and 5th and 10th percentile values of the mICM EQR distributions

Stream type	N	REF	5th percentile	10th percentile
R-C1x2	27	0.15	0.83	0.84
R-C3	63	0.36	0.79	0.83
R-C4x2	21	0.13	0.80	0.84
all data combined	111	-	0.79	0.83

3.4 Discussion

3.4.1 Description of stream type-specific macrophyte communities

The identification of common high status sites confirms the first hypothesis. For each stream type several surveys are classified in high quality status by most of the national methods. The smaller relative number of such surveys at lowland sites can be attributed to the generally more degraded lowland conditions. The strong linear relation of the abundance of certain macrophyte taxa with the mean index supports the existence of common indicator species. These findings allow stream type-specific communities and their environment to be defined under near-natural and degraded conditions.

The highest quality R-C1x2 sites feature a combination of submerged rooted aquatic species of which *Callitriche hamulata*, *Potamogeton natans* and *Sparganium emersum* are by far the commonest. Scarcer associated species, which characterise high status R-C1x2 rivers include *Potamogeton alpinus*, *Myriophyllum alterniflorum, Elodea canadensis,* various other *Callitriche* spp. and *Ranunculus peltatus*. Emergent vegetation is dominated by *Phalaris arundinacea*, *Sparganium erectum* and *Phragmites australis*, the latter characteristic of high status sites. There are a range of moderate- to small-sized emergent species, of which *Persicaria hydropiper*, *Myosotis scorpioides*, *Glyceria fluitans*, *Berula erecta, Mentha aquatica* and *Veronica anagallis-aquatica* are the most abundant, but it is the less frequent elements, such as *Cardamine amara* and *Caltha palustris* that are characteristic. This assemblage is most likely to be associated with small, active, mesotrophic, shallow, sand-dominated, clear water, moderately fast flowing, partially-shaded streams. The very limited number of characteristic species suggests that this is a type with several geographically distinct variants under high status conditions. With declining quality there is a shift to a community dominated by *Sparganium emersum*, alongside a range of species that are absent from or much scarcer in the highest status sites, including *Potamogeton pectinatus*, *P. trichoides*, *P. perfoliatus* and *P. crispus*, and the duckweeds *Lemna minor* and *L. minuta*. Among the emergent species that overlap with high status sites *Sparganium erectum*, *Persicaria hydropiper*, *Phalaris arundinacea*, *Phragmites australis*, *Myosotis scorpioides* and *Berula erecta* are all much reduced, and are typically replaced by *Rorippa amphibia*, *Glyceria aquatica*, *Sagittaria sagittifolia* and *Alisma plantago-aquatica*. This change in structure suggests a shift to silty, stable, eutrophic, slow flowing, turbid conditions in managed or regulated channels with degraded riparian habitat.

High quality R-C3 rivers feature a combination of leafy liverworts (*Scapania undulata* or *Chiloscyphus polyanthus,* and less frequently*, Marsupella emarginata* or *Jungermannia atrovirens*), acrocarpous mosses (most notably *Racomitrium aciculare*, plus smaller quantities of marginal species, such as *Philonotis fontana* and *Dicranella palustris, Fissidens crassipes* and *F. rufulus*), thallose liverworts (*Pellia epiphylla*), and small macroalgae, including *Lemanea, Oscillatoria* and *Mougeotia* spp. These taxa occur against a backdrop of extensive growths of a range of pleurocarpous mosses, including *Rhynchostegium riparioides*, *Fontinalis squamosa, F. antipyretica*, *Hygrophypnum ochraceum* (and occasionally *H. luridum*)*, Brachythecium rivulare, B. plumosum*, *Hyocomium armoricum*, *Thamnobryum alopecurum* and *Amblystegium fluviatile*. Vascular plants are likely to be restricted to *Callitriche hamulata*, plus occasional marginal growth of species such *Glyceria fluitans*, *Phalaris arundinacea* and *Ranunculus flammula*. The latter often occurs alongside a range of mire forming species, of which the mosses *Sphagnum*, *Mnium hornum*, *Philonotis caespitosa* and *Plagiomnium undulatum* are most characteristic. This is an assemblage of small, shallow, turbulent, flashy, neutral to base-poor, oligotrophic, upland rivers, with a cobble and boulder substrate, often with extensive shading by deciduous trees. Several of these species persist in the lowest quality sites, most notably *Fontinalis antipyretica* and *Rhynchostegium riparioides*, but most bryophytes are replaced by *Amblystegium riparium*. Channel margins are likely to feature more extensive growth of *Phalaris arundinacea*, plus *Sparganium erectum* and a range of smaller amphibious species, such as *Agrostis stolonifera*, *Glyceria fluitans*, *Veronica beccabunga* and *Myosotis scorpioides*. The cover of instream vascular species is generally small, but may include *Callitriche hamulata*, *Ranunculus peltatus*, *Elodea nuttallii*, *Potamogeton crispus*, *Sparganium emersum* and *Ceratophyllum demersum*. Larger green filamentous algae are generally present and will include *Cladophora glomerata* and *Vaucheria* spp. This reflects a shift to more stable, moderate to slow flowing, fertile conditions with reduced shading of the margins and mixed sand-gravel substrates. This change is therefore most likely to be associated with a combination of pollution and siltation from diffuse sources, flow regulation, channel realignment and overgrazing.

High quality R-C4x2 streams are dominated by two species *Fontinalis antipyretica* and *Sparganium emersum*, each of which account for 10 percent of the total plant cover. Other common and widely distributed instream aquatics include *Nuphar lutea*, *Elodea canadensis* and *Amblystegium riparium*, plus the red encrusting alga *Hildenbrandia*, which is highly characteristic of high status R-C4x2 rivers. A diverse range of pondweed species (especially *Potamogeton alpinus*, *P. perfoliatus* and *P. natans*, but occasionally *P. praelongus* or *P. gramineus*) occur alongside batrachids, such as *Ranunculus fluitans* and *R. aquatilis*, plus *Myriophyllum spicatum* and *Callitriche hamulata*. Marginal vegetation is dominated by *Phalaris arundinacea* and *Sparganium erectum*, but other stand forming species are also frequently present and will include *Scirpus lacustris*, *Iris pseudacorus*, *Glyceria aquatica* and *Equisetum fluviatile*. Of the smaller marginal species *Mentha aquatica*, *Veronica anagallis-aquatica*, *Alisma plantago-aquatica*, *Glyceria fluitans*, *Berula erecta* and *Myosotis scorpoides* are especially well represented, and may harbour small patches of various lemnids. Within the marginal zone *Carex rostrata* and *Lysimachia thyrsiflora* are uncommon but are unique to high status sites. This is an assemblage of medium sized, active, moderate to fast-flowing, shallow lowland rivers on neutral to base-rich geology with clear, mesotrophic to eutrophic water. A diversity of substrates occurs, and will include a mix of sand, gravel and unsilted coarser material. The vegetation itself is a major architect of hydromorphological diversity. Remnants of this vegetation occur in rivers in central and north west Europe (e.g. Wiegleb, 1984; Holmes et al., 1999), but it is only in the less densely populated countries of north east Europe that this vegetation can still be found with any regularity (e.g. Paal & Trei, 2004, Baattrup-Pedersen et al, 2008). The most degraded sites are strongly characterised by *Potamogeton pectinatus*, which accounts for 21 percent of the total plant cover in common poor or bad status sites. Of the commoner instream associates *Fontinalis antipyretica*, *Elodea canadensis* and *Sparganium emersum* are all greatly reduced compared to their contribution in high status sites, and are likely to be replaced by *Potamogeton crispus*, *Elodea nuttallii*, *Lemna minor*, *Ranunculus penicillatus*, *Ceratophyllum demersum*, *Persicaria amphibia* or *Zannichellia palustris*, plus various large green filamentous algae, including *Cladophora*, *Rhizoclonium*, *Vaucheria* and *Oedogonium*. The status of *Nuphar lutea* and *Amblystegium riparium* is little changed in comparison to the highest status sites. The margins remain dominated by *Phalaris arundinacea* and *Sparganium erectum* with *Glyceria aquatica* and *Scirpus lacustris* as common associates. However, in place of a range of smaller herbaceous species *Solanum dulcamara*, *Rorippa amphibia*, *Sagittaria sagittifolia* or *Typha latifolia* normally occur. Collectively this assemblage indicates a highly enriched, stable, sluggish, well lit environment dominated by fine sediment. Such vegetation is often associated with streams in urban or intensely agricultural catchments, where management and physical modification of channels and their margins are the norm.

3.4.2 Development of a common metric for intercalibration

The mICM proves to be a suitable common metric for the intercalibration of the national macrophyte methods used in this study. Except for two cases all regressions are characterized by a coefficient of determination ≥ 0.5, thus meeting an important intercalibration criterion given by Owen et al. (2010) in the comparison of benthic invertebrate methods. On average, the mICM is related more strongly to the national methods than the common metric proposed in Chapter 2. However, compared to the outcomes of a diatom intercalibration exercise its performance is poorer (Kelly et al., 2008). This is mainly attributable to the low average relation of the mICM to the Flemish and German methods, underlining their conceptual difference.

This approach to developing a common metric for intercalibration allows differences between methods to be detected at the level of national indicator lists. The mICM taxa scores actually represent a correlation with averaged national indicator ratings. Positive scoring taxa are common indicators of near-natural conditions, negative scoring taxa generally characterize poor or bad quality. Taxa with low correlation to the mean index are either indicative of moderate conditions or rated inconsistently among countries. Obvious discrepancies between the mICM indicator values and the national ratings are easily identified and adjusted. This option provides the opportunity to harmonize the national methods by implementing only minor, thus easily justified changes, rather than seeking wholesale changes in class boundaries which may be politically difficult to achieve.

The mICM indicator scores were derived by Spearman rank correlation. In selecting this analysis I assumed that a linear model best describes the distribution of taxa abundance across the gradient represented by the mean index range. Though not further explored in this chapter I also tested if a unimodal approach using weighted averaging of taxa abundances is more suitable to derive indicator scores. Due to lower correlations of the mICM with the national assessment methods this option was rejected. The linear model seems to better reflect the similarity among national sensitivity metrics that abundant indicator taxa contribute proportionally more to the national EQR score.

I also dismissed alternative approaches based on relative macrophyte abundance data or the exclusion of taxa occurring in less than five surveys since both options also showed weaker relationships. All national metrics use absolute instead of relative abundance values. The indicator scores assigned to rare taxa may be biased by their low occurrence. It is possible that their scores would change when using a larger dataset. However, the better correlation of the mICM in its proposed version justifies the inclusion of rare indicator taxa.

A common value of ~0.8 at the 5th percentile of the mICM EQR provides some reassurance over the potential utility of this approach and demonstrates that there is sufficient commonality in interpretation of high status within each river type for this view to form a robust basis for testing national classifications. A 5th percentile of EQR of ~0.8 is consistent with reference site EQR variability in invertebrate based classification tools, such as RIVPACS (Clarke et al., 1996). It also lends itself to a statistically-based placement of class boundaries from high-good, down to poor-bad, at unit intervals of 0.2.

Growth form metrics add a new dimension to macrophyte based classification which departs strongly from structural assessments in which the indicator value of individual taxa takes priority. Thus, species which share the same growth form may have very different indicator values (and will thus tend to have different mICM scores), while other species representing different growth forms may have similar indicator values. Further work is required to determine how best to integrate elements of national classification methods not shared by other countries within the approach presented here. Thiebaut et al. (2002) reported that the performance of diversity based measures was generally inferior to trophic indices for use in macrophyte classification of rivers. However, Willby et al. (2008) argued that some form of diversity index was desirable within classification in order to differentiate between results based on data of contrasting biological quality, and also to better reflect physical habitat heterogeneity.

3.5 Conclusions

This study represents an important contribution to the intercalibration of river macrophyte classifications in Europe. It defines common reference conditions for three widespread stream types and provides a means to compare the good ecological status of national methods. Furthermore, this work offers a general approach to harmonize the national assessment methods for biological elements of any water category. Based on the differing national assessments of similar transnational ecotypes the approach reveals the common ground of national quality classifications. Basic elements are the common high status sites and the mICM indicator list.

The description of the ecotype-specific communities and their environmental conditions goes beyond these outcomes. It amalgamates the national notions of biological communities at high and bad quality status and establishes an international guiding image that is not influenced by national specialities or biogeographical differences. This image will be of crucial importance in the follow-up process towards harmonization of ecological quality classification.

4 A new procedure for comparing class boundaries of biological assessment methods: a case study from the Danube Basin

4.1 Introduction

Monitoring the biological quality of rivers has a long tradition in the Danube River Basin. In communist times the evaluation of saprobic water quality was standardized in Eastern Europe (Helešic, 2006) and several countries supported research on bioassessment and monitoring (e.g. Zelinka & Marvan, 1961; Rothschein, 1962; Sládeček, 1973; Uzunov, 1979). However, compared to chemical water classification biological assessment played a minor role also in the pan-European context (Newman, 1988). Against this background, the European Water Framework Directive 2000/60/EC (WFD) has set new requirements for water policy. Besides integrated and coordinated river basin management for all European river systems it stipulates ecological quality assessment against near-natural reference conditions specific to each type of water body. For rivers, fish, benthic invertebrates, macrophytes and benthic algae, and phytoplankton are assessed. Results are given in relation to the near-natural reference conditions, thus expressed as numbers between 0 (worst status) and 1 (near-natural reference status), i.e. the ‘Ecological Quality Ratio’ (EQR). The EQR range is split into five classes (high, good, moderate, poor, and bad).

Although individual countries are in charge of modifying their national assessment methods or of developing new methods, the quality classification at the European level is harmonized by intercalibration (Heiskanen et al., 2004). Intercalibration is a legally binding requirement of the WFD. It guarantees the consistent quality classifications despite still diverse assessment methods that countries are applying. European Member States are obliged to compare the results of assessments among countries that share common water body types in similar biogeographical regions. For this, countries are organized in so-called Geographical Intercalibration Groups (GIG). A major policy objective is to achieve good surface water status throughout Europe by 2015. Intercalibration therefore focuses on the EQR values that define good ecological status, i.e. the high-good and good-moderate class boundaries. A list of the main terms and definitions connected with the intercalibration process as meant in this chapter is given in Table 27.

There are three methodological options for intercalibration (European Communities, 2005):

Option 1. Boundaries are compared directly between countries that are using identical assessment methods (e.g. CB GIG Lakes, 2008).

Option 2. The results of national assessment methods are translated into a comparable format using common metrics (e.g. Buffagni et al., 2006). Unlike national methods, common metrics are not optimised for quality assessment but are conversion tools for biological assessment indices.

Option 3. Different national methods are compared directly by assessing the same sampling sites using the participating countries national assessment methods (e.g. Borja et al., 2007; see also Chapter 1 of this thesis).

All these options require data on sites covering the whole range of quality classes to secure statistical robustness of intercalibration results.

Table 27: Definition of main terms dealt with in this chapter

Main term	Definition
1. Intercalibration	Process by which European countries compare and harmonize the quality class boundaries of their biological assessment methods (high-good and good-moderate boundary).
2. Harmonization	If the comparison of biological assessment methods reveals differences between national class boundaries, these differences are harmonized. This is done by adjusting the national boundaries with reference to biological benchmarks.
3. Biological benchmark	Condition of the biological community that represents the transnational reference point for harmonization. The biological benchmark is defined for selected aspects of the biological community measured by common metrics.
4. Common metric	A biological metric widely applicable within a GIG or across GIGs, which can be used to derive comparable information among different countries/stream types (Buffagni et al., 2007).
5. Standardization	Normalization of metric values via transformation to unitless scores. Metrics are divided by the values representing the near-natural condition or the biological benchmark condition.
6. Threshold value	Value of selected environmental parameters/common metrics that influence/indicate the biological condition at the stream site, e.g. conductivity or agricultural land use in the catchment. Threshold values were used to screen for stream sites of at least good environmental status.

In Central Europe, Member States recently intercalibrated river diatom and invertebrate classifications by common metrics (Option 2) (CB GIG Rivers, 2008). These metrics were correlated with the national assessment methods and regression analyses inferred the values of the common metrics that corresponded to the national quality class boundaries. To compare common metrics between countries they had to be standardized. For this purpose the participating countries provided data on undisturbed reference sites, which were selected with harmonized criteria (CB GIG Rivers, 2008). The biological community of these undisturbed sites yielded the reference value of the common metrics and provided EQR scales that were comparable between countries. The principal problem with this approach was the scarcity of reference sites, since unimpacted conditions no longer exist (e.g. Birk et al., 2007b; Gabriels, 2007) or data were not available as monitoring focuses on impacted sites. Several countries could therefore not intercalibrate their methods, especially those applied for large rivers. Therefore, the question arises: Does intercalibration of class boundaries necessarily require data on reference sites or are there alternative approaches?

In this study, I developed a new method for river types of five countries in the Danube River Basin (Figure 7), for which reference data were almost completely unavailable. Benchmarks were therefore established with data from similarly impacted river sites. This approach was tested for both, assessment methods based on benthic diatoms and methods based on benthic invertebrates.

Figure 7: Map of Europe showing the locations of Austria (AT), Slovak Republic (SK), Hungary (HU), Romania (RO) and Bulgaria (BG).

4.2 Materials and Methods

4.2.1 National assessment methods and intercalibration common stream types

I intercalibrated two multimetric diatom indices used in Austria and the Slovak Republic (Table 28). The Austrian method classifies the EQRs of the Trophic Index (TI) (Rott et al., 1999) and Saprobic Index (SI) (Rott et al., 1997) separately and the overall quality status is determined by that index delivering the worst result. The Slovak method integrates the results of three diatom metrics (Indice de Polluosensibilité Spécifique (IPS): CEMAGREF, 1982; Eutrophication/Pollution Index - Diatom-based (EPI-D): Dell'Uomo, 1996; Diatom Index by Descy & Coste (1991) (CEE)). The absolute index values are classified by a five-fold, stream-type specific classification scheme. The overall status is expressed as the averaged class values of each index divided by the maximum obtainable score.

Five invertebrate methods were intercalibrated (Table 28). The multimetric indices of Austria and the Slovak Republic appraise various aspects of the river invertebrate community such as faunal composition, abundance, richness, diversity, sensitivity and ecosystem function (BMLFUW, 2006b). The Bulgarian and Hungarian methods integrate information on taxonomic composition and tolerance to general disturbance, while the Romanian method is a modification of the Saprobic Index. The Saprobic Index indicates biodegradable organic pollution on the basis of species composition and species-specific saprobic indicator values. Further details on these methods are given by Birk & Schmedtje (2005).

Table 28: National assessment methods for rivers using benthic diatoms and invertebrates.

Country	Method name
Benthic diatoms	
Austria	Austrian Phytobenthos Assessment - Component: diatoms
Slovak Republic	Slovak Phytobenthos Assessment - Component: diatoms
Benthic invertebrates	
Austria	Austrian System for Ecological River Status Assessment using Benthic Invertebrates
Bulgaria	Bulgarian Biotic Index for River Quality Assessment (Q-Scheme)
Hungary	Hungarian Average Score Per Taxon (ASPT)
Romania	Romanian Saprobic Index following Pantle & Buck (1955)
Slovak Republic	Slovak System for Ecological River Status Assessment using Benthic Invertebrates

Four common stream types (ECOSTAT, 2004b) (Table 29) were defined for the intercalibration of the selected assessment methods. Ecoregion, catchment area, altitude, geology and dominant channel substrate were used to define the stream types. The common types covered small to medium sized, mid-altitude streams of the Carpathians with coarse bed substrate (Romania and the Slovak Republic), and rivers of different size and altitude ranges in the Hungarian Plains and the Pontic Province (Austria, Bulgaria, Hungary, Romania and the Slovak Republic).

Table 29: Common stream types addressed in this study

Stream type abbreviation	Common stream type	Ecoregion (Illies, 1967)	Catchment area [km^2]	Altitude [m]	Geology	Channel substrate
R-E1	Carpathians: small to medium, mid-altitude	10 (The Carpathians)	10 - 1000	500 - 800	Siliceous	Gravel and boulder
R-E2	Plains: medium-sized, lowland	11 (Hungarian Lowlands) and 12 (Pontic Province)	100 - 1000	< 200	Mixed	Sand and silt
R-E3	Plains: large and very large, lowland	11 (Hungarian Lowlands) and 12 (Pontic Province)	> 1000	< 200	Mixed	Sand, silt and gravel
R-E4	Plains: medium-sized, mid-altitude	11 (Hungarian Lowlands) and 12 (Pontic Province)	100 - 1000	200 - 500	Mixed	Sand and gravel

4.2.2 Data

The analyses were based on national monitoring data from sampling sites at the common stream types. The data included information on composition and abundance of benthic diatoms and invertebrates, selected chemical parameters, the classification of hydromorphological quality (only for invertebrate sampling sites) and catchment land use.

The number of countries included in the intercalibration for the individual stream types varied, depending on the relevance of the type to the country, the availability of national assessment methods and data. The number of sites and samples included into the analysis differed between countries and stream types. In total, data from 356 sampling sites were included, comprising 140 diatom and 543 invertebrate samples (Table 30).

The procedures for sampling benthic diatoms for both compared methods was in line with the European standard EN 13946:2003 or related national protocols. The invertebrate samples were obtained by country-specific, national sampling protocols. Two groups of sampling methods were used (Table 31): Pro-rata Multi-Habitat Sampling (Hering et al., 2003) and Standard Handnet Sampling (EN 27828:1994). The main differences were area-related versus time-related sampling in the field, and, in case of the Multi-Habitat Sampling, the application of sub-sampling procedures in the laboratory. Differences between country-specific methods related to mesh size, recording of abundance and the level of taxonomic identification.

Table 30: The number of sites and samples per country and common intercalibration type, and number of taxa per sample

Stream type abbreviation	Country	Number of sites	Number of samples	Number of taxa per sample	
				median	min / max
Benthic diatoms					
R-E4	Austria	81	117	30	6 / 83
	Slovak Republic	10	23	38	16 / 75
Benthic invertebrates					
R-E1	Romania	52	142	14	2 / 63
	Slovak Republic	39	103	46	1 / 83
R-E2	Romania	24	41	10	2 / 31
	Slovak Republic	11	23	28	6 / 48
R-E3	Bulgaria	32	63	14	3 / 28
R-E4	Austria	46	58	61.5	10 / 105
	Hungary	43	76	13.5	1 / 68
	Slovak Republic	18	37	25	3 / 57

Table 31: National methods for sampling and processing invertebrate samples

Sampling method	Pro-rata multi-habitat-sampling	Standard-handnet-sampling
Country	Austria, Slovak Republic, Hungary	Bulgaria, Romania
Description	Area-related sampling using handnet (20 sampling units taken from all habitat types with >=5% coverage; Hungary: 10 samples)	Time-related using handnet (3 to 5 min., all available habitats)
Mesh size	500 µm Hungary: 950 µm	500 µm
Sampling technique	Kick and sweep	Kick and sweep, additional hand picking
Abundance recording	Individuals per m^2	Bulgaria: 5 abundance classes Romania: number of individuals
Sub-sampling	Yes	No
Identification level	Species	Bulgaria: genus and family Romania: species
Reference	Hering et al. (2003) Nieuwenhuis (2005)	EN 27 828:1994

Table 32 lists the environmental data collected for each sampling site. Physico-chemical parameters were generally measured monthly and averaged over six months (diatom sampling sites) or one year (invertebrate sampling sites).

Table 32: Environmental data collected at each sampling site

Physico-chemical parameters
Conductivity, pH, alkalinity, dissolved oxygen, oxygen saturation, biological oxygen demand (5 day), total phosphorus, ortho-phosphate, nitrate, nitrite, ammonium
Hydromorphological parameter
Hydromorphological quality status (see Table 33)
Catchment land use parameters
% Urban land use
% Intensive agriculture
% Non-intensive agriculture
% Forest
Land use index (Böhmer et al. 2004)

Since no common method for evaluating the structural quality of sites was available, I developed a classification scheme to assess the hydromorphological status of the invertebrate sampling sites. According to the degree of degradation one of three classes was allocated to each site (Table 33). This classification was based on expert judgement of the field staff who sampled the streams.

Share of urban land cover (Corine Land Cover (CLC) class 1), intensive agricultural land cover (CLC codes 2.1, 2.2, 2.4.1, 2.4.2) and non-intensive agricultural land cover (CLC codes 2.3.1, 2.4.3, 2.4.4) in the catchment were taken from CLC data (Bossard et al., 2000). These data were used to calculate the Land Use Index (Böhmer et al., 2004): 4 * urban land use + 2 * intensive agriculture + non-intensive agriculture.

Environmental data were not available for some Romanian and Slovakian sites.

4.2.3 Data analysis

Overview

Figure 8 provides an overview of the analytical procedure. The intercalibration approach was based on the application of common metrics. Using environmental criteria I screened for sampling sites of at least good environmental status. For these sites the distribution of common metric values was calculated. The upper or lower quartile values of this distribution were used as the “biological benchmark”.

Table 33: Classification scheme to assess the hydromorphological quality status of invertebrate sampling sites

Class 1 - near-natural hydromorphological conditions
- Stream type specific variability of channel depth and channel width, shallow profile, close connectivity of the stream and the floodplain
- Natural channel substrate conditions (composition and variability), presence of dead wood
- Bank profile and bank structure unmodified
- Presence of natural riparian vegetation
- Natural hydromorphological dynamic is maintained
- Low degree of anthropogenic land use in the floodplain
Class 2 - moderately altered hydromorphological conditions
- Decreased variability of channel depth and channel width
- Minor changes to bank morphologies, or only one bank is fixed with "soft works"
- Riparian vegetation altered
- Loss of stream length, longitudinal profile is altered by man
Class 3 – severely altered hydromorphological conditions
- Obvious presence of hard engineering
- Severe modifications of instream structures, bed and bank fixation and artificial substrates
- No or only minor variability of channel substrate
- No riparian zone between river and land use
- Channelised, straightened and/or deep-cut river
- Disconnection of river and floodplain

In the next step, I regressed common metrics against national assessment indices; this was done differently for diatom and invertebrate methods:

(1) For comparing the boundaries of diatom methods the national indices represented the predictor variables. Here, I identified the common metric values corresponding to the national good quality boundaries.

(2) For setting the boundaries of invertebrate methods the reverse approach was applied. The common metrics were used as predictor variables from which national boundaries were inferred. I applied this different approach because for only two of the five national methods ecological quality classes were defined.

Selection of common metrics

The results of the national assessment methods were compared against common methods, the so-called “Intercalibration Common Multimetric indices” (ICMi, Buffagni et al., 2005). These indices represented combinations of two or more single metrics that measured different aspects of the biological community. The diatom ICMi comprised the common metrics IPS (CEMAGREF, 1982) and TI (Rott et al., 1999) which are parts of the national assessment methods of Austria and the Slovak Republic, respectively. Kelly et al. (2008) applied this multimetric index to compare the diatom indices of Central European countries.

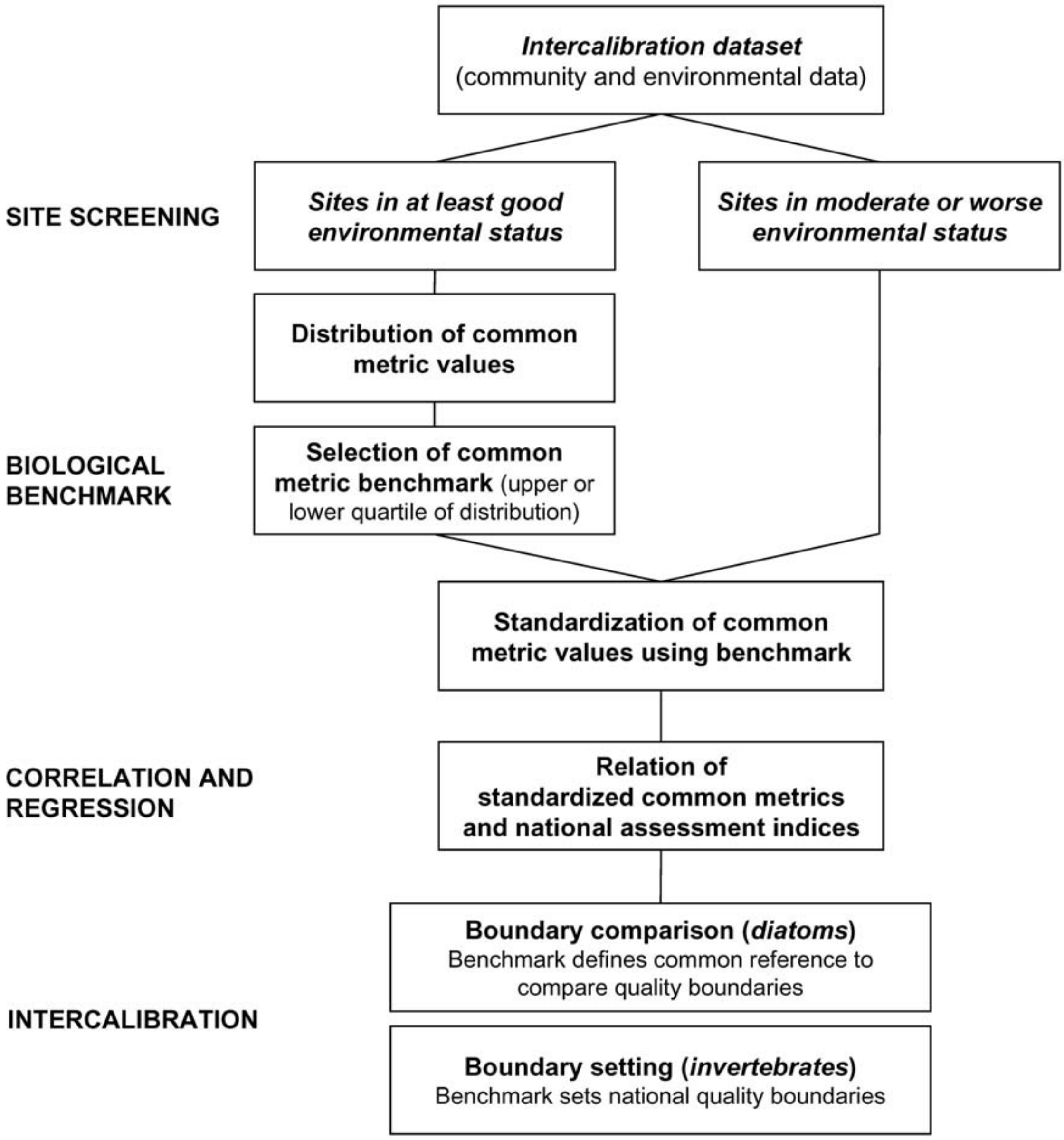

Figure 8: Overview of the analytical procedure

For the invertebrate methods, I developed an invertebrate ICMi using the biological data collated in this study. 140 metrics at the taxonomic level of family were correlated with national indices. Twelve common metrics were selected, which represent different metric types (Hering et al., 2006b) and were among those indices correlating most strongly with all the national indices (average Spearman Rank correlation coefficient > 0.55). Out of these, metrics of different type were combined into four different multimetric indices assessing taxonomical composition, abundance, diversity and sensitivity. The multimetric index that correlated best with all the national indices was used as the invertebrate ICMi.

Data screening

I identified sites of at least good status based on threshold values of selected environmental parameters. These values were mainly taken from scientific literature and environmental standards (Table 34). In the discussion section, I describe the rationale for their selection in detail.

Table 34: Threshold values of environmental parameters used to screen for diatom (DI, only type R-E4) and invertebrate (BI) sampling sites of high or good environmental status (n.a. = not applicable).

Common intercalibration type	Carpathians: R-E1	Plains: R-E2, R-E3, R-E4	
Threshold for	High environmental status	Good environmental status	
Biological quality element	BI	DI	BI
Biological parameter			
Average Score Per Taxon	≥ 6.4	n.a.	≥ 5.1
Chemical parameters			
Total phosphorus [µg l^{-1}]	n.a.	< 100	n.a.
Ortho-phosphate [µg l^{-1}]	n.a.	< 70	n.a.
Biological oxygen demand (5 day) [mg l^{-1}]	≤ 2.5	n.a.	≤ 5.0
Conductivity [µS cm^{-1}]	n.a.	< 1000	
Hydromorphological parameter			
Quality class	1	n.a.	1 and 2
Land use parameter			
Land use index	≤ 50	≤ 140	

For diatoms, I specified limits for total phosphorus, orthophosphate, conductivity and Land Use Index (Table 34). Good environmental status was allocated to every site that showed values below these limits. The invertebrate sampling sites were classified by biotic criteria and environmental data. The biological classification was based on the quality class boundaries of the Austrian Saprobic Index (ÖNORM M6232:1997) proposed for mountain and lowland rivers in the Danube River Basin by Knoben et al. (1999). These boundaries were translated into an index on family level (Average Score Per Taxon, ASPT; Armitage et al., 1983) by linear regression. First, sites were screened by their samples' mean ASPT, and then the abiotic criteria were applied to those sites that passed the ASPT threshold. Different criteria were established for the river types of the Carpathians and the Plains. Data sets of small to medium sized streams in the Carpathian Mountains (R-E1) were screened to identify sites of high environmental status. Data from sites of high status were generally scarce for the common types in the Plains (R-E2, R-E3 and R-E4). Here, I used adapted threshold values to identify sampling sites of good environmental status.

Definition and application of benchmarks

For the purpose of this analysis, “biological benchmarks” are defined as values of the common metrics that correspond to similar levels of disturbance, representing either the high-good or the good-moderate boundary.

Biological benchmarks were derived from the data of sites of at least good environmental status. I calculated the common metrics for all sites and identified the distribution of common metric values that occurred at high or good status sites, respectively. Out of this distribution I selected the quartile values to define biological benchmarks. For the intercalibration of diatom assessment methods I chose the upper quartile for IPS and the lower quartile for TI as benchmarks. Basis was the metric distribution of the combined Austrian and Slovak data, since diatom data were derived by identical sampling protocols. For the invertebrate metrics, the lower quartiles were selected, since metric values increase with degradation. Benchmark calculation was done separately per country and stream type to account for the different national sampling protocols. Standardized common metrics were then combined to the invertebrate ICMi by averaging.

The relation of national diatom indices and the diatom ICMi was calculated by regression models. The quality boundaries high-good and good-moderate of the Austrian and Slovak assessment methods were translated into corresponding values of the ICMi. In invertebrate intercalibration the ICMi was used to harmonize the national quality class boundaries. Depending on the screened dataset, an ICMi value of ‘1’ represented either the high-good or the good-moderate boundary. Based on the relationship of ICMi and national index the national boundaries were inferred by regression analysis. Those boundaries that were not specified by the screened dataset were defined as the 20 percent deviation from the modelled boundaries.

4.3 Results

4.3.1 Selection of common metrics

The ICMi for the intercalibration of the invertebrate methods comprised four metrics: Average Score Per Taxon (ASPT), Austrian Structure Index at Family Level (Structure Index), Total Number of Families (#fam) and Relative abundance of Ephemeroptera, Plecoptera and Trichoptera taxa (%EPT). Type-specific analyses of common metrics and national assessment indices resulted in mean correlation coefficients ranging from R = 0.56 (%EPT) to R = 0.73 (ASPT). The ICMi and its component metrics correlated significantly with catchment land use, hydromorphological status, dissolved oxygen concentration and biological oxygen demand (Table 35). The correlation coefficients of ICMi and national indices varied between R = 0.67 and R = 0.81.

Table 35: Maximum Spearman Rank correlation coefficients for environmental variables and common metrics from national datasets (n.s. = non-significant correlation; *p<0.05, **p<0.01, ***p<0.001).

Environmental parameter	#fam	ASPT	Structure Index	%EPT	EE ICMi
Land use index	n.s.	-0.63***	-0.55***	-0.40*	-0.48***
Hydromorphological quality class	-0.42*	n.s.	-0.57***	-0.71*	-0.75*
Dissolved oxygen concentration	0.74***	0.72***	0.62**	0.41*	0.49**
Biological oxygen demand	-0.51*	-0.59*	-0.51*	-0.53*	-0.62**

Correlation coefficients for national indices against the diatom ICMi were generally higher (R ≥ 0.89). The individual metrics TI and IPS showed highly significant relationships with all of the environmental parameters except temperature, pH, oxygen concentration (only TI) and oxidised nitrogen.

4.3.2 Data screening

27 sites passed the diatom screening thresholds of good environmental status. Total phosphorus represented the most stringent criterion that classified sites as “moderate” or worse. The screening procedure using the harmonized quality criteria for invertebrates resulted in national data subsets of different size. On average, national datasets comprised seven sites of at least good environmental status.

4.3.3 Definition and application of benchmarks

The benchmarks calculated for diatom metrics were IPS = 16.9 (upper quartile value) and TI = 2.44 (lower quartile value). The invertebrate common metrics gave different quartile values for the individual type/country datasets. High status samples of the mountain streams, for instance, showed a quartile value of 9 families (Romania) and 24 families (Slovakia). Another example is the percentage of EPT families in R-E4 samples of good environmental status: 10 percent for Austria versus 50 percent for Hungary. In Figure 9, the ranges of common metrics are juxtaposed for selected data subsets.

I used linear models to translate the national diatom boundaries into diatom ICMi values (Figure 10). For the Austrian method, the overall biological quality class is determined by the worst of the individual module classes. Therefore, the more precautionary boundary values of the Trophic Index (Table 36) were compared to the Slovak quality boundaries. The comparison revealed different settings for the high-good boundary and the near-natural reference value between countries.

Table 36: Quality class boundaries and near-natural reference values of the national diatom indices translated into diatom ICMi values (dICMi = diatom ICMi; TI-AT = Austrian Trophic Index; SI-AT = Austrian Saprobic Index; DI-SK = Slovak Diatom-Index; 95CI = 95 percent confidence interval of regression line).

Class boundary	TI-AT	dICMi	95CI	SI-AT	dICMi	95CI	DI-SK	dICMi	95CI
High-good	0.69	1.00	±0.02	0.85	0.90	±0.02	0.90	0.89	±0.02
Good-moderate	0.41	0.68	±0.01	0.71	0.68	±0.02	0.70	0.67	±0.02
Near-natural reference	1.00	1.35	±0.03	1.00	1.15	±0.04	1.00	0.99	±0.03

The boundaries for the national invertebrate methods were predicted using linear or lognormal regression against the invertebrate ICMi (see Figure 11 for an example). Table 37 shows the boundaries for national assessment methods derived from the comparison to the ICMi. In addition, the table indicates those values that were defined by 20 percent deviation from predicted boundaries.

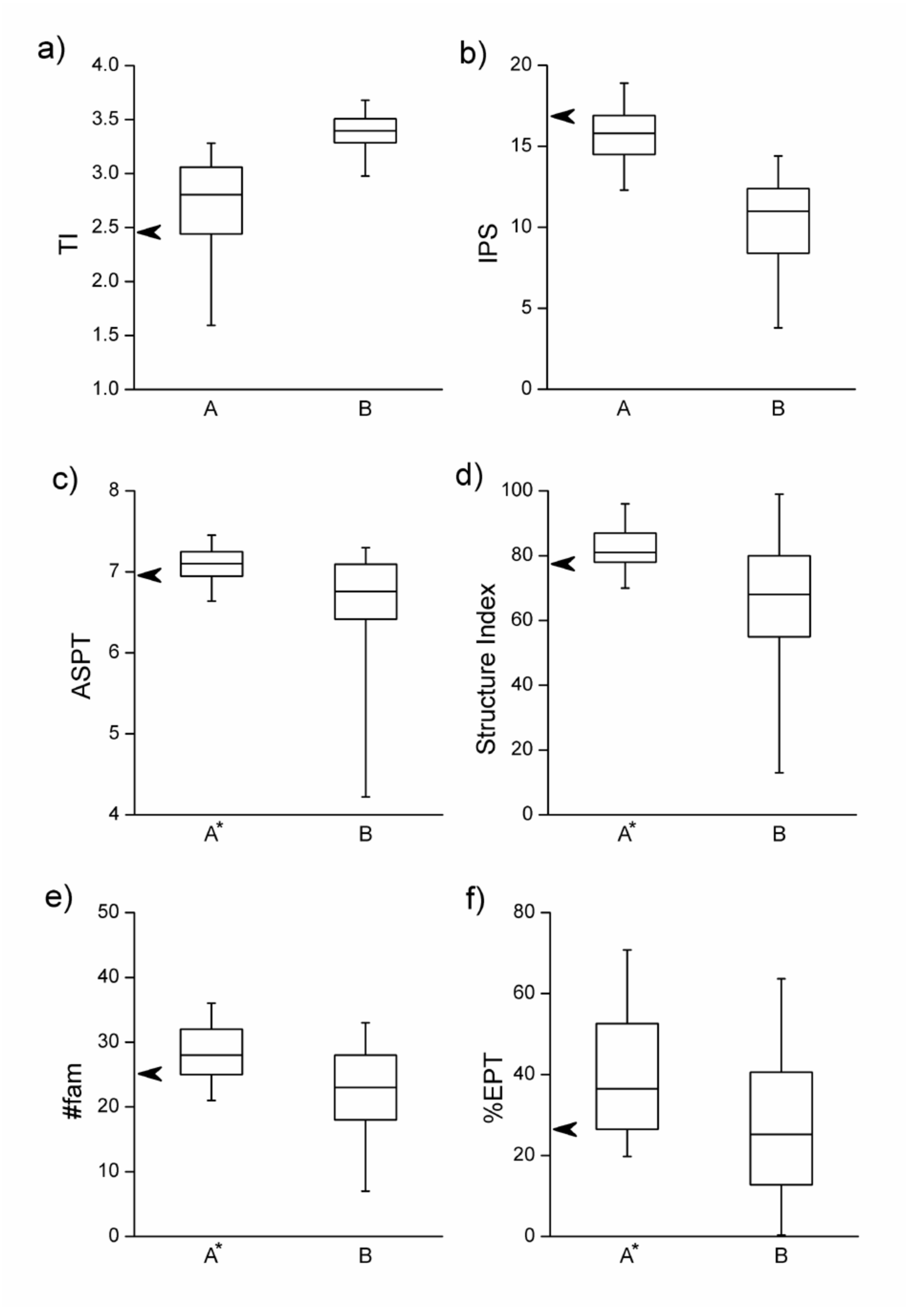

Figure 9: Calculation of benchmarks. Distribution of diatom (a, b) and invertebrate (c to f) common metric values at sites of good (A) or high (A*, Slovak R-E1) and worse (B) environmental status. Metrics were standardized by the quartile values marked with an arrow. Relevant quartiles between groups (A - B) are significantly different at p<0.001 (χ^2-Test).

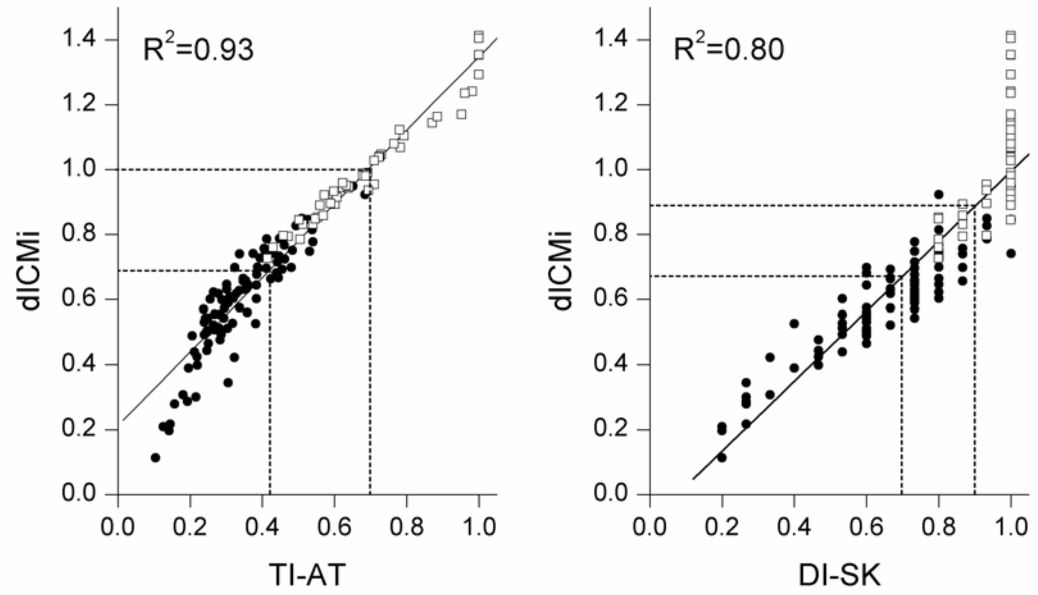

Figure 10: Boundary comparison. Translation of Austrian (TI-AT) and Slovak (DI-SK) good quality boundaries into comparable values of the diatom common metric (dICMi) by linear regression (dashed lines). White squares represent samples of good environmental status.

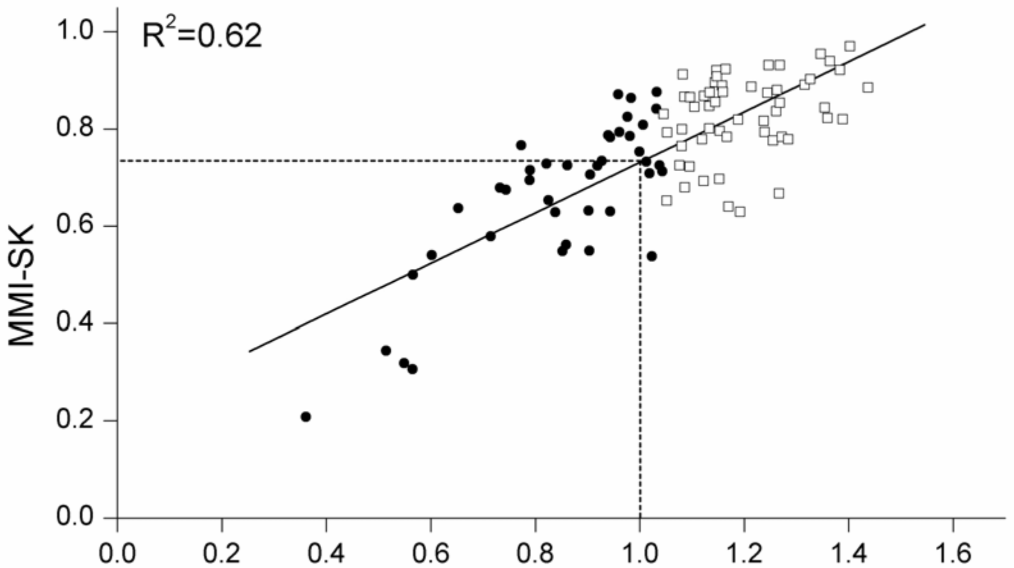

Figure 11: Setting the high-good class boundary for the Slovak invertebrate index (MMI-SK) using the biological benchmark (invertebrate ICMi = 1). White squares represent samples of high environmental status (R^2 = coefficient of determination).

Table 37: Biological class boundaries derived from regression analysis of the invertebrate ICMi against national indices (95CI = 95 percent confidence interval of regression line; * = class boundary defined as 20 percent deviation from predicted boundary value; ‡ = Confidence interval derived from regression analysis using ranks transformed into whole numbers (“1” = 1, “1 to 2” = 2, “2” = 3 etc.).

IC type	Country	Class boundary	Boundary value	95CI
R-E1	Romania	High-good	1.70	± 0.09
		Good-moderate	2.27*	
	Slovak Republic	High-good	0.74	± 0.02
		Good-moderate	0.54*	
R-E2	Romania	High-good	1.81*	± 0.09
		Good-moderate	2.26	
	Slovak Republic	High-good	0.74*	± 0.05
		Good-moderate	0.54	
R-E3	Bulgaria	High-good	4 to 5*	± 0.39‡
		Good-moderate	3 to 4	
R-E4	Austria	High-good	0.79*	± 0.03
		Good-moderate	0.59	
	Hungary	High-good	7.36*	± 0.14
		Good-moderate	5.52	
	Slovak Republic	High-good	0.72*	± 0.04
		Good-moderate	0.52	

4.4 Discussion

4.4.1 Objectives of boundary comparison and setting

The application of benchmarks for comparing and setting boundaries follows different objectives. In boundary comparison, discrepancies between national classifications are identified, but no guidance is given for adjustment. For instance, in this analysis the diatom intercalibration revealed differing high-good boundaries and reference values between the Austrian and Slovak classifications. In the official intercalibration exercise of the Central-Baltic GIG harmonization was achieved in such cases by averaging the boundary values of all participating countries (CB GIG Rivers, 2008). This approach has the character of a committee agreement and is inappropriate if only a small number of national methods are involved. However, with reference to the benchmark, the standardized common metric indicates harmonization requirements: only the high-good boundary of the Austrian Trophic Index is close to a dICMi value of ‘1’. This suggests that a boundary adjustment of the other indices is needed.

In contrast, benchmarks used in boundary setting establish harmonized biological standards directly, without reference to existing national classifications. A somewhat similar approach was described by Sandin & Hering (2004) who applied abiotic descriptors of organic pollution for setting boundaries for invertebrate assessment indices. However, the parameter thresholds were derived from existing national classifications that were actually intercalibrated. Buffagni et al. (2007) proposed an independent scientific classification of biological data to be used as a benchmark in intercalibration. In the present study, the boundaries were mainly derived from abiotic criteria. This concept accounts for the differing status of national method development in the Danube River Basin (Schmedtje, 2005). The boundary setting approach allows for a gradual intercalibration exercise. As soon as individual countries have completed their national assessment methods, the national quality class boundaries can be adapted to the common benchmark.

4.4.2 Rationale for selecting environmental parameters for benchmark definition

A reasonable definition of thresholds requires distinct pressure-impact relationships between environmental parameters and intercalibrated biological metrics. I demonstrated correlations between the environmental parameters and the common metrics and significant differences between the quartiles of the common metric data in the screened data sets. The actual threshold values were derived from scientific literature and environmental standards, and they were confirmed by expert judgement. Because of their importance in the applied intercalibration procedure I briefly describe the rationale for selecting the criteria in the following paragraphs.

The abiotic screening criteria are either factors that affect the stream biota directly (phosphorus concentrations for diatoms; biological oxygen demand, hydromorphology for invertebrates), or indirect, integrative indicators of various human influences (conductivity, land use). Their impact on the invertebrate community is considered by using a biotic index. The Average Score Per Taxon (ASPT) primarily indicates the effects of organic pollution, but it also responds to hydromorphological degradation (Buffagni et al., 2005) and other stressor types. The metric is less influenced by seasonal variation and sampling differences (Armitage et al., 1983; Friberg et al., 2006) and therefore suited for the analysis of data from different sources. The ASPT thresholds used for screening sites corresponded to limits for the Saprobic Index that were proposed by Knoben et al. (1999). I translated these values into ASPT to benefit from the features of a common metric, i.e. minimisation of differences caused by biogeographical variations, the type of degradation, the level of taxonomic identification or the sampling method. The defined thresholds fall 3 percent (for high environmental status in small mountain streams) and 20 percent (for good environmental status in medium-sized lowland streams) below the values for near-natural reference conditions used by the British River Invertebrate Prediction and Classification System (Wright et al., 2000) for small mountain and medium-sized lowland streams, respectively (see Chapter 1 of this thesis).

The phosphorus thresholds for screening the diatom sites correspond to the good-moderate boundary values for these parameters applied in Germany (LAWA-AO, 2007). Compared to the Austrian and Slovak standards (Deutsch & Kreuzinger, 2005) these thresholds are rather precautionary if applied to R-E4 rivers. Sládeček (1973) gave various examples of the direct relationship between biological oxygen demand (BOD) and the benthic invertebrate community. The author allocated a BOD of 2.5 mg l^{-1} to the lower range of oligosaprobic status, while the beta-mesosaprobic status was characterised by BOD values of about 5.0 mg l^{-1}. The latter threshold, as well as the conductivity limit that I used, corresponds to the bottom end of good chemical water quality in various national and international classification schemes (e.g. Newman, 1988, ICPDR, 2004, MMGA, 2006). Parameters relating to hydromorphological status comprise overall quality indicators (longitudinal stream profile) and ratings for specific elements that are relevant to the benthic invertebrates, such as riparian vegetation (extent and degree of shading), instream woody debris and bank modification (Lorenz et al., 2004, Feld & Hering, 2007).

Catchment land use generally represents an integrative measure of human influences on stream ecosystems (Allan, 2004) as it reflects the driving forces impairing river quality. Agricultural and urban land use account for an array of mechanisms that alter the riverine environment. Their extent is related to the proportion of particular land use categories in the catchment. The Land Use Index combines and weights the percentage cover of different land uses. The threshold value that I selected allows, for instance, for a maximum of 12.5 percent (high environmental quality) and 35 percent (good environmental quality) urban land cover in the catchment. However, these numbers are hypothetical because human settlements are usually accompanied by farmland. Therefore, the actual percentage of urban land cover at threshold is much lower.

4.4.3 Consistent and verifiable definition of benchmarks

The official WFD intercalibration exercise is a comprehensive procedure covering large geographical areas and different countries (ECOSTAT, 2004b). If many national methods have to be intercalibrated, the use of common metrics is scientifically sound and convenient (Buffagni et al., 2006, 2007). The crucial steps of the intercalibration analysis are (1) relating national indices to common metrics and (2) standardizing common metrics using a benchmark. Cross-national comparability of common metric values can only be achieved after the second step. Like EQRs it establishes a relative measure (observed condition in relation to benchmark condition) that compensates for the biogeographical and methodological differences of each national method (Buffagni et al., 2005). For convenience, both steps can be completed individually by each country (CB GIG Rivers, 2008).

Because the intercalibration results will influence water management decisions across Europe, the process must be accountable. In this regard, Biggs (2006) commented that standardization based on near-natural reference sites is difficult to verify if these reference sites are identified by the Member States themselves. Following the concept of “zero or insignificant pressure”, any procedure to screen for pristine sites requires extensive datasets. European Communities (2003), for instance, specified 19 general screening criteria. Depending on data availability, only a limited number of criteria can be checked in practice. The guidelines for selecting reference sites in the Central European intercalibration exercise dealt with this by accepting gaps in data and variations in the quality of data to a certain extent (CB GIG Rivers, 2008). Furthermore, even the criteria defined by European Communities (2003) do not necessarily meet the definition of ‘undisturbed’ conditions; large rivers, lowland rivers and almost all rivers in the Mediterranean have been so severely altered by hydraulic engineering, the disconnection of the rivers and its floodplain, water abstraction, pesticides and a multitude of other impacts that undisturbed, pristine conditions do not exist any longer (Moss, 2007). Basing assessment methods and intercalibration on the comparison with “undisturbed” references is therefore a risky approach, as the actual status is compared to an almost unknown status. Precision and confidence is much higher if benchmarks are set in a transparent and verifiable way.

The definition of selected environmental criteria described in this contribution offers a practical solution to problems of data availability that are often encountered in intercalibration (Buffagni et al., 2007). In this approach, the amount of data required is comparatively low. Except for hydromorphological quality, I used parameters that were collected by standard monitoring or satellite remote sensing. The availability of this data accounted for a complete database that enabled consistent site screening, and therefore a verifiable standardization process.

Summary and conclusions

The research presented in this thesis was directed by the question, how the definitions of good ecological status can best be compared between national assessment methods. I approached this problem by investigating three related aspects:

1. Relationships between the biological indices

The relationship between the biological indices employed by the national assessment methods was explored. I found out that invertebrate-based indices were stronger correlated than those used in macrophyte assessment. The ten national invertebrate-indices applied to data of two stream types showed an average coefficient of determination larger than 0.5 (Chapter 1). The four macrophyte-indices were related with a mean R^2 value of less than 0.3 (Chapter 2). Here, nonlinear equations provided better fits in most of the regressions. In both cases, biological indices of the same type (indices sensitive to organic pollution or eutrophication) showed best correlation results ($R^2 > 0.7$).

These outcomes were relevant for selecting the intercalibration approach. The strong relation of invertebrate-indices allowed for a direct comparison of national assessment methods. Two common scales were used: (1) The national index showing the highest mean correlation of all indices. (2) The "Integrative Multimetric Index for Intercalibration" (IMI-IC), an artificial index designed for the purpose of intercalibration. This index was defined as the mean of all national index values calculated for a sample. The average R^2 of the IMI-IC amounted to nearly 0.8, with the indices applied to the data of medium-sized lowland streams performing slightly better. Due to the weaker relationships between macrophyte-indices, I also tested the performance of common metrics besides using the best correlated national index. The trophic index "Ellenberg_N" was considerably related to three out of four assessment methods. However, the average coefficient of determination was below 0.5 due to the poor relationship with the Dutch Macrophyte Score.

In Chapter 3, the "macrophyte Intercalibration Common Metric" (mICM) was developed to compare seven national methods at three Central European stream types. This common metric yielded a mean R^2 value above 0.6, with the correlations of the mountain type data clearly performing above average ($R^2 > 0.7$). The mICM was based on a set of stream type-specific, common indicator species. This concept allowed for amendments of the national indicator lists aiming at harmonized quality classification. Chapter 4 describes the use of common metrics for the intercalibration of diatom and invertebrate methods. The diatom methods were strongly related to the common metric ($R^2 \geq 0.8$). The Spearman correlation coefficients of the invertebrate methods ranged between 0.67 and 0.81. The general approach of using common metrics accounted for the differing status of national method development in the Danube River Basin and thus provided the basis for a gradual intercalibration exercise. As soon as individual countries will have completed their national assessment methods, the national quality class boundaries can be compared via common metrics.

2. Role of reference conditions in the intercalibration exercise

The second aspect of my investigations focused on the role of reference conditions in the intercalibration exercise. The good ecological status is defined as a slight deviation from undisturbed conditions. These undisturbed conditions represent the reference point of assessment. In the intercalibration process a harmonized reference setting is key to successful boundary comparison. In this thesis, I tested various approaches to define comparable reference conditions between national methods. High status sites identified by means of abiotic reference criteria (pre-classification) and biological validation (post-classification) set common stream type-specific reference values for the analyses in Chapter 1. The 75th percentile index values at high status sites established rather stringent conditions compared to the national references. In Chapter 2, the 95th percentile values of the entire data range were set as reference points. This approach resorted to best available conditions and did not require additional environmental data. However, this reference setting corresponded to different quality status according to the national classifications. These outcomes pointed at basic differences in the national conceptions of the reference state.

To solve these incomparabilities the national definitions of high status had to be consolidated. In Chapter 3, I described the common stream type-specific macrophyte communities occurring under undisturbed conditions. This established an international guiding image that was not influenced by national specialities or biogeographical differences. Sites classified in high status by the majority of national methods, and by none of the methods in moderate or worse status, were used to set reference values.

The narrow spread of mICM values among these common high status sites provided some reassurance over the potential utility of this approach and demonstrated that there is sufficient commonality in interpretation of high status within each stream type for this view to form a robust basis for testing national classifications.

In the Danube River Basin case study, data from near-natural reference sites were generally scarce. Therefore, an alternative approach was tested based on sites impacted by similar levels of disturbance. Using environmental variables I screened for sampling sites of at least good environmental status. Common metric values obtained from the screened datasets revealed "biological benchmarks", that represented harmonized points of reference for the intercalibration exercise. Unlike in the previous case studies, the calculation of invertebrate benchmarks was done separately per country and stream type to account for the different national sampling protocols.

3. Comparison and harmonization of national good status boundaries

The third aspect of my investigations covered the actual comparison and harmonization of the national boundaries of good ecological status. In particular cases, the boundary comparisons presented in Chapter 1 revealed discrepancies between national classification schemes of more than 25 percent. The extent of differences between class boundaries was largely dependent on the common scale used for comparison. The use of abiotic pressure data in intercalibration allowed for an additional interpretation of these results, indicating that harmonization is only required between two groups of boundaries with overlapping pressure intervals. For the macrophyte methods both direct and indirect intercalibration options disclosed major differences between the national good status boundaries. However, the nonlinear relationships of the macrophyte indices made the comparison difficult. In the lower quality range Mean Trophic Rank and Ellenberg_N were not responding to changes of the German Reference Index. The German class boundaries thus showed overlapping confidence intervals when transferred into the common scale.

The common grounds in European macrophyte assessment established in Chapter 3 did not comprise any comparison of national quality classes. However, the description of the ecotype-specific communities and their environmental conditions amalgamated the national notions of biological communities at high and bad quality status. Furthermore, the 5th percentile of mICM EQR of ~0.8 was consistent with reference site

EQR variability in invertebrate based classification tools. It thus lends itself to a statistically-based placement of class boundaries from high-good, down to poor-bad, at unit intervals of 0.2. Both guiding image and mICM boundaries will be of crucial importance in the follow-up process towards harmonization of ecological quality classification for river macrophytes.

The intercalibration exercise performed in Chapter 4 comprised the comparison and harmonization of national quality boundaries. The diatom classifications of Austria and the Slovak Republic were compared using the biological benchmarks. This analysis revealed different settings for the high-good boundary and the near-natural reference value between countries. Here, the biological benchmark allowed for the identification of discrepancies between boundaries, but no guidance was given for their adjustment. In contrast, the benchmarking approach was used to set the national good quality boundaries for the invertebrate methods. The benchmarks, that were derived using a set of abiotic criteria characterizing at least good environmental status, established harmonized biological standards directly, without reference to existing national classifications.

Conclusions

The intercalibration exercise plays a prominent role in setting quality standards for European surface waters. The process establishes a transboundary concept of good ecological status that is of high socio-economic relevance: European Member States are obliged to maintain or restore their water bodies to this status. Against this background it is necessary to critically review validity and limitations of intercalibration. In this final section I am widening the scope of this dissertation and touch on relevant aspects excluded from the specific examinations of the main chapters.

The Water Framework Directive literally “frames” the scientific work presented in this thesis, basically by the ecological assumptions that underlie the design of the Directive (Steyaert & Ollivier, 2007, Hatton-Ellis, 2008). The concept of classifying ecological status implies a static, non-dynamic notion of nature. The natural conditions are prized, and human activities are considered as a source of disturbance responsible for status degradation. This status is mainly characterized by the taxonomic composition and abundance of selected organism groups, defining nature in terms of the integrity of the aquatic communities. These statements set out the main aspects of the ecological perception of the WFD. It is far beyond the scope of this thesis to rate the appropriateness and suitability of these assumptions. Nevertheless, they form important preconditions that provide the conceptual basis for ecological status assessment and intercalibration.

More relevant for the actual evaluation of the intercalibration process is the issue of uncertainty related to its results. Various studies highlight effects of sampling method and sample size (Clarke et al., 2006, Vlek et al., 2006), sample processing (Haase et al., 2006) and temporal variation (Sporka et al., 2006) on the outcomes of river macrozoobenthos assessments. Carstensen (2007) specifies monitoring requirements to ensure adequate confidence and precision in classification. However, this is a new, largely unexplored topic. Therefore, few of the national methods covered in this thesis employ schemes for error estimation. In the presented intercalibration analyses uncertainty is only considered in terms of the confidence intervals of the regression performed between assessment indices. Other works (Kelly et al., 2008, Owen et al., 2010) use predefined intervals (“harmonization bands”) instead of single boundary values to account for various indefinite errors in the comparison. Future studies on intercalibration will have to consider in more detail the role of uncertainty and its effects on the harmonization of national classifications.

A general criticism of the intercalibration process is raised by Kelly et al. (2008). The legal requirement to intercalibrate probably contributes to a conservative approach to method development, since radical approaches to ecological status assessment are, by their nature, more difficult to compare with other methods. This becomes evident in the intercalibration of macrophyte classifications. Here, the more innovative growth form appraisal had to be excluded due to its incomparability with the classical assessments based on macrophyte composition. Changing land use practices and industries, but on the other hand measures of the river basin management will reveal new or unforeseen threats to the aquatic environment. In Germany, for instance, the relevance of organic pollution has declined over the past decades as more and more households were connected to sewage treatment and industrial pollution became more strongly regulated. These measures exposed the severe impacts of diffuse agricultural pollutants and structural degradation on riverine ecosystems. Today, also pesticides, organic toxicants or pharmaceuticals threaten the aquatic ecosystems, although their precise impacts are largely unknown. These conditions require a certain flexibility in quality monitoring, but the intercalibration process may act as an administrative barrier constraining future adaptability.

My final remark returns to the fundamental issue already raised in the preface: How can the normative idea of good status be adequately defined by means of scientific approaches? Pollard & van de Bund (2005) outline theoretical options for the boundary setting of good quality status, like identification of discontinuities in the biological response to anthropogenic pressure, or the cross-over of two antagonistic biological metrics. In practice, these approaches are often unfeasible due to the inability to identify clear pressure-impact-relationships from regular monitoring data. But in general, the application of these concepts (e.g. ecological breakpoints, Buffagni et al., 2004) allows for a more defensible, albeit arbitrary boundary setting: What is justifying the selection of these specific criteria and not others? The recent discussion on environmental thresholds (Groffman et al., 2006) provides utilitarian arguments: The ecosystem quality has significantly declined if ecological services, i.e. specific ecosystem functions that are valued by humans, are endangered. However, the same authors question the applicability of this concept since these services are often difficult to measure.

The experience I gained during the last years allows for a concluding appraisal: How do I actually rate the intercalibration exercise? Here, I have to give an ambiguous answer. On the one hand, the pan-European discourse on ecological assessment and environmental quality standards, that intercalibration initiated, cannot be overrated. The exercise promoted transboundary collaboration between scientists and water managers, and already yielded solid outcomes compared to the difficulties the process had to face. On the other hand, the exercise revealed considerable knowledge gaps, for instance, with regard to the ecological processes in aquatic systems and their response to human pressure or restoration. In-depth investigations of these issues were hampered by the tight schedule of the WFD. In the follow-up of intercalibration it is necessary to keep the right balance between the legal demands and the scientific essentials. Moreover, I see the intercalibration work as a continuous requirement within the overall WFD obligations. Many upcoming challenges like climate change will have to be addressed on the international level. The established network of practitioners and applied scientists already proved efficient in dealing with the various issues of water monitoring and ecological assessment. Therefore, I recommend that Member States ensure to maintain this expert network beyond the currently envisaged time frame.

The thesis at hand contains some of the first studies in support of the WFD intercalibration exercise. Therefore, it mainly concentrates on approaches to compare existing classification schemes. Boundary setting was committed to the developers of the national assessment methods. Future research will have to focus on an objective rationale for classification that goes beyond committee agreements or expert consensus. Here, a dialog between ecology and environmental ethics could bring forward a consolidated notion of good quality that satisfies both the needs of nature and society. In this regard the intercalibration process holds the potential to pave the way for an integrated and applicable code of conduct towards the environment.

Zusammenfassung

Hintergrund

„Was soll ich tun?" lautet nach Kant (1800) eine der Grundfragen der Philosophie, welche stellvertretend für die vielfältigen Inhalte ethischer Reflexionen steht. Waren vornehmlich die interpersonellen Beziehungen Gegenstand der Ethik, so fand in der zweiten Hälfte des letzten Jahrhunderts das Verhältnis von Mensch und Natur durch die Wahrnehmung einer Umweltkrise zunehmende Bedeutung (Hardin, 1968, White, 1968). In diesem Zusammenhang kann Kants „Was soll ich tun?" konkretisiert werden als „Wie habe ich mich gegenüber der natürlichen Mitwelt richtig zu verhalten?". Die Ökologie kann keine Antwort auf diese Frage liefern, da sie normative Aussagen fordert, die jenseits des beschreibenden und erklärenden Charakters der Naturwissenschaften stehen (Hume, 1978, Valsangiacomo, 1998). Unsere Vorstellung vom richtigen Umgang mit der natürlichen Mitwelt ist Teil des gesellschaftlichen Diskurses und manifestiert sich, zum Beispiel, in der Umweltpolitik. Hier bildet sie den moralischen Hintergrund, vor dem die angewandte Naturwissenschaft agiert.

Die vorliegende Dissertation beinhaltet angewandte Wissenschaft zur Umsetzung der Europäischen Wasserrahmenrichtlinie (Europäische Kommission, 2000). Dieses Umweltgesetz schafft einen Ordnungsrahmen für Maßnahmen im Bereich der Wasserpolitik für die 27 Mitgliedstaaten der Europäischen Gemeinschaft. Die Richtlinie fordert von den Mitgliedstaaten eine ökologischen Zustandsbewertung ihrer Flüsse, Seen, Küstengewässer und Ästuare (Flussmündungen). Anhand von Bewertungsverfahren bestimmen die Länder den Zustand ausgewählter aquatischer Tier- und Pflanzengruppen, den so genannten biologischen Qualitätskomponenten. Diese Verfahren unterscheiden zwischen verschiedenen Typen von Oberflächengewässern. Bezugspunkt der Bewertung ist der vom Menschen unbeeinträchtigte Gewässerzustand, dass heißt der Referenzzustand, welcher je nach Gewässertyp unterschiedlich ausgeprägt ist. Die Ergebnisse der nationalen Bewertungsverfahren werden als relative Übereinstimmung mit dem Referenzzustand dargestellt im so genannten „Ecological Quality Ratio" (EQR). Je nach Grad der Übereinstimmung erfolgt die Beurteilung des ökologischen Zustands in den Klassen sehr gut, gut, mäßig, unbefriedigend oder schlecht (Birk & Böhmer, 2007).

Die Wasserrahmenrichtlinie fordert den guten ökologischen Zustand für alle Wasserkörper und definiert diesen Zustand über normative Begriffsbestimmungen (Europäische Kommission, 2000, S. 38):

> *„Die Werte für die biologischen Qualitätskomponenten des Oberflächengewässertyps zeigen geringe anthropogene Abweichungen an, weichen aber nur in geringem Maße von den Werten ab, die normalerweise bei Abwesenheit störender Einflüsse mit dem betreffenden Oberflächengewässertyp einhergehen."*

Die Definition des guten ökologischen Zustands stellt ein Schlüsselelement in der Europäischen Wasserpolitik dar. Die Gemeinschaft verpflichtet ihre Mitgliedstaaten zum richtigen Umgang mit der aquatischen Umwelt und schreibt das Ergreifen von geeigneten Maßnahmen vor, wenn diese Zielsetzung verfehlt wird. Das Konzept des guten ökologischen Zustands ist daher von entscheidender Bedeutung bei der Umsetzung der Wasserrahmenrichtlinie. Dennoch überlässt es die Richtlinie den Mitgliedstaaten, auf welche Weise diese recht unkonkrete Definition in die Praxis umgesetzt wird. Die Länder selbst stehen in der Verantwortung, die Bewertungsverfahren zu entwickeln und damit den Zustand ihrer Wasserkörper einzustufen. Um die nationalen Interpretationen des guten ökologischen Zustands zu vergleichen und zu harmonisieren, schreibt die Richtlinie die so genannte Interkalibrierung vor (Heiskanen et al., 2004).

Die Interkalibrierung zielt darauf ab, für alle Mitgliedstaaten einen vergleichbaren Anspruch im Gewässerschutz zu schaffen. Aufgabe der Interkalibrierung ist, die europaweit einheitliche Bewertung des guten ökologischen Zustands durch die nationalen Bewertungsverfahren zu gewährleisten. Vereinfacht ausgedrückt: Die Interkalibrierung soll sicherstellen, dass zum Beispiel ein Wasserkörper in Belgien, der von dem belgischen Verfahren als „gut" bewertet wird, auch vom deutschen und niederländischen Verfahren als „gut" eingestuft würde, wenn sich derselbe Wasserkörper auf deutschem oder holländischem Gebiet befände (Birk & Böhmer, 2007). Allerdings zeigen die Gewässer eines vergleichbaren Typs Unterschiede in Fauna und Flora zwischen den Ländern, auch unter vom Menschen unbeeinflussten Bedingungen. Darüber hinaus sind die nationalen Verfahren durch verschiedene Bewertungskonzepte und -traditionen gekennzeichnet (Birk, 2003, Birk & Schmedtje, 2005). Werden nur Flüsse und Seen berücksichtigt, in denen jeweils vier

Qualitätskomponenten bewertet werden (Phytoplankton, Phytobenthos und Makrophyten, Makrozoobenthos, Fische), sind allein schon über 200 nationale Verfahren zwischen den 27 Mitgliedstaaten zu interkalibrieren (Birk et al., 2009). Dies gibt einen Eindruck von der schwierigen und komplexen Aufgabe der Interkalibrierung.

Die vorliegende Arbeit schafft die methodische Basis für die technische Umsetzung der Interkalibrierung. Die grundlegende Fragestellung lautet: Wie können die nationalen Definitionen des guten ökologischen Zustands am besten verglichen werden? Dabei kommen die Ansätze des direkten Verfahrensvergleichs sowie des indirekten Vergleichs anhand von Allgemeinen Metriks[6] zur Anwendung. In vier Fallstudien untersuche ich (1) die numerischen Zusammenhänge der nationalen Bewertungsverfahren, (2) die Rolle unterschiedlicher Definitionen von Referenzzuständen innerhalb der Interkalibrierung sowie (3) die Möglichkeiten einer einheitlichen Festlegung des guten ökologischen Zustands. Die vier Fallstudien behandeln insgesamt 26 nationale Verfahren zur Bewertung des ökologischen Zustands von Fließgewässern anhand von Makrozoobenthos (15 Verfahren), Makrophyten (9 Verfahren) und benthischen Diatomeen (2 Verfahren). Für die verschiedenen Analysen werden mehr als 1.900 biologische Probenahmen genutzt, welche im Rahmen von Europäischen Forschungsprojekten oder Programmen der nationalen Gewässerüberwachung erhoben wurden. Die Arbeit deckt drei Interkalibrierungstypen[7] Mitteleuropas ab, sowie vier Interkalibrierungstypen in Osteuropa.

1 Direkter Vergleich von nationalen Verfahren zur Bewertung des Makrozoobenthos in Fließgewässern

In der ersten Fallstudie untersuchte ich die numerischen Zusammenhänge von zehn Verfahren zur Bewertung des Makrozoobenthos in Fließgewässern. Datengrundlage bildeten Probenahmen, die durch eine einheitliche Methode in Rahmen der

[6] Unabhängig von den naturräumlichen Gegebenheiten und den spezifischen Formen der Gewässerbelastung eines Landes erfassen Allgemeine Metriks die generelle Belastung eines Gewässers durch den Menschen, wenn auch in etwas unschärferer Weise als die national angepassten Verfahren.

[7] Die Interkalibrierungstypen umfassen Gewässer mit vergleichbaren Merkmalen, die in verschiedenen Ländern vorkommen. Ihre Ausweisung stützt sich auf die Beschreibung ausgewählter Parameter, wie Ökoregion, Größe, Höhenlage, Geologie oder Sohlsubstrat.

Forschungsprojekte AQEM[8] und STAR[9] erhoben wurden (Hering et al., 2004, Furse et al., 2006). Die Daten wurden getrennt für zwei Interkalibrierungstypen Mitteleuropas analysiert. 294 Probenahmen an 125 Stellen in Deutschland, Österreich, Tschechien sowie der Slowakei ließen sich dem Gewässertypen der silikatischen Mittelgebirgsbäche zuordnen. Die kleinen Flüsse des Tieflands waren durch 217 Proben an 71 Gewässerstellen in Dänemark, Deutschland, Großbritannien und Schweden vertreten. Neben den biologischen Daten zu taxonomischer Zusammensetzung und Abundanz des Makrozoobenthos waren diverse physiko-chemische Parameter, die hydromorphologische Qualität sowie Daten zur Landnutzung in Gewässerumfeld und Einzugsgebiet verfügbar. Ferner wurde der ökologische Zustand jeder Probestelle vor Ort durch den jeweiligen Probenehmer voreingestuft. Die untersuchten Bewertungsverfahren umfassten Saprobienindizes und weitere biologische Metriks, die in der nationalen Gewässerüberwachung von Dänemark, Deutschland, Großbritannien, Polen, Schweden, der Slowakei, Tschechien oder Österreich angewendet werden. Diese Länder definierten Grenzwerte für die Einstufung des guten ökologischen Zustands, welche in dieser Studie die Grundlage für den Vergleich der Zustandsbewertungen darstellten.

Auf Basis der Makrozoobenthos-Daten erfolgte für jede Probenahme die Berechnung der biologischen Indizes. Die Auswahl des 75. Perzentilwerts innerhalb der als „sehr gut“ voreingestuften Probestellen ermöglichte eine einheitliche Festlegung von Gewässertyp-spezifischen Referenzwerten. Die Werte eines jeden nationalen Index’ konnten somit als EQR dargestellt werden. Die Korrelation der nationalen EQR sowie die Art ihres Zusammenhangs (linear, nichtlinear) wurden durch Regressionsanalyse bestimmt. Um die Definitionen der nationalen Klassengrenzen des guten ökologischen Zustands zu vergleichen, wurden zwei Vergleichsskalen definiert: (1) Der nationale Index, der die höchste mittlere Korrelation zu allen Bewertungsverfahren aufwies, und (2) der so genannte „Integrative Multimetric Index for Intercalibration“ (IMI-IC), der sich aus dem Mittelwert aller nationalen Bewertungsverfahren pro Probestelle zusammensetzte. Die nationalen Grenzwerte wurden mit Hilfe der

[8] “The Development and Testing of an Integrated Assessment System for the Ecological Quality of Streams and Rivers throughout Europe using Benthic Macroinvertebrates.” Forschungsprojekt im fünften Rahmenprogramm der Europäischen Kommission.

[9] “Standardisation of River Classifications: Framework method for calibrating different biological survey results against ecological quality classifications to be developed for the Water Framework Directive.” Forschungsprojekt im fünften Rahmenprogramm der Europäischen Kommission.

Regressionsfunktionen in Werte der Vergleichsskalen übertragen, und für jeden Wert wurde das 95-Prozent-Konfidenzintervall berücksichtigt. Eine mit den Umweltvariablen durchgeführte Hauptkomponentenanalyse bestimmte, welche Form von anthropogener Belastung im Datensatz eines Interkalibrierungstypen am stärksten ausgeprägt ist. Mittels linearer Regression wurde den nationalen Klassengrenzen des guten Zustands ein korrespondierender Belastungsgrad inklusive des Konfidenzintervalls zugewiesen.

Die Analysen zeigten, dass die einheitlich festgelegten Referenzwerte strenger als die national definierten Werte ausfielen. Der mittlere Determinationskoeffizient (R^2) aller Regressionen der nationalen Verfahren war größer als 0,5. Der deutsche Saprobienindex korrelierte am höchsten für den Datensatz der Mittelgebirgsbäche. Für die Proben der kleinen Tieflandflüsse ergab der dänische Flussfaunaindex die höchsten mittleren R^2-Werte. Die Koeffizienten des IMI-IC lagen im Mittel um 0,8. Vornehmlich wiesen nichtlineare Zusammenhänge gegenüber linearen Beziehungen höhere Determinationskoeffizienten auf. Allerdings waren diese Unterschiede nicht signifikant, daher setzte ich für die weiteren Analysen einen linearen Bezug voraus. Die Hauptkomponentenanalyse zeigte, dass die Daten der Mittelgebirgsbäche durch einen Gradienten der Nährstoffbelastung und organischen Verschmutzung geprägt waren. Die Probenahmen an den kleinen Tieflandflüssen bildeten einen hydromorphologischen Gradienten ab. Generell waren die Bewertungsergebnisse der Mittelgebirgsbäche höher mit dem dort vorherrschenden Belastungsgradienten korreliert. Der Vergleich der nationalen Klassengrenzen über die beiden Vergleichsskalen zeigte Abweichungen von bis zu 25 Prozent. Je nach Skala waren unterschiedliche Abweichungen zu verzeichnen. Anhand der Regression mit den Belastungsgradienten konnten Gruppen mit einheitlicher Grenzdefinition bestimmt werden.

2 Interkalibrierung von Bewertungsverfahren für Makrophyten in Flüssen des Tieflands: direkter Vergleich und Analyse von Allgemeinen Metriks

Die zweite Fallstudie untersuchte zwei verschiedene Optionen des Vergleichs nationaler Bewertungsverfahren anhand von Makrophyten in Fließgewässern. Datengrundlage bildeten 108 nach harmonisiertem Protokoll erhobene Vegetationsaufnahmen an kleinen Flüssen des Tieflands (Dänemark, Deutschland, Großbritannien, Lettland, Polen, Schweden; Furse et al., 2006). Diese Stellen wurden jeweils mit vier nationalen Verfahren bewertet. Hierzu mussten die vorliegenden

Angaben zum Deckungsgrad der Makrophyten-Arten in die nationalen Häufigkeitsskalen übersetzt werden. Einheitliche Referenzwerte leitete ich über das 95. Perzentil der Indexwerte aller Vegetationsaufnahmen ab. Auf Grundlage der Vegetationsaufnahmen wurden 70 biologische Metriks berechnet und mit den nationalen Bewertungsindizes in Relation gesetzt. Ziel war die Bestimmung sowohl von nationalen Indizes als auch Allgemeinen Metriks, die mit allen nationalen Verfahren korrelieren. Aus einer Hauptkomponentenanalyse gewonnene Umweltgradienten dienten zur Dokumentation einer Dosis-Wirkungs-Beziehung zwischen menschlicher Belastung und Reaktion der Makrophytenindizes. Zwei Vergleichsskalen erlaubten die Überprüfung der nationalen Definitionen des guten ökologischen Zustands. Die Übertragung dieser Werte erfolgte durch die aus der Regressionsanalyse resultierenden Kurvengleichungen.

Die nationalen Einstufungen des ökologischen Zustands differierten erheblich zwischen den Bewertungsverfahren. Ebenso fielen die einheitlich definierten Referenzwerte in unterschiedliche nationale Zustandsklassen. Das holländische Verfahren bewertete am strengsten; alle Probestellen wurden als mäßig oder schlechter klassifiziert. Die Bewertungsergebnisse der Indizes aus Frankreich und Großbritannien wiesen einen Determinationskoeffizienten von größer 0,75 auf. Das deutsche und niederländische Verfahren war geringer mit diesen Indizes korreliert. Bei der Regression vor allem des deutschen Verfahrens wurden nichtlineare Zusammenhänge deutlich. Von den 70 Makrophytenmetriks erwies sich nur ein auf Nährstoffzeigern basierender Index (Ellenberg et al., 1992) als Allgemeiner Metrik brauchbar. Dieser Metrik zeigte deutliche Beziehungen zum britischen, deutschen und französischen Verfahren, korrelierte aber schwach negativ mit dem holländischen Index. Aus diesem Grund wurde das niederländische Verfahren vom anschließenden Vergleich der Zustandsklassen ausgeschlossen. Dieser Vergleich offenbarte starke Unterschiede zwischen den nationalen Klassengrenzen. Außerdem erschwerten die nichtlinearen Beziehungen eine Übertragung der nationalen Grenzen in Werte der Vergleichsskalen. Ferner waren alle außer dem niederländischen Verfahren mit dem Nährstoffgradienten korreliert. Der holländische Index reagierte sensitiv gegenüber genereller Degradation.

3 Schaffung einer gemeinsamen Basis für die Europäische Bewertung von Makrophyten in Fließgewässern

Die Ergebnisse des zweiten Kapitels verdeutlichten die Notwendigkeit weiterer Forschung bezüglich der Interkalibrierung von Makrophyten-Verfahren. Vor diesem Hintergrund wurden in einer dritten Fallstudie 609 Vegetationsaufnahmen aus den nationalen Überwachungsprogrammen von zwölf Europäischen Mitgliedstaaten zusammengetragen. Ziel der Studie war die Schaffung einer gemeinsamen Basis für den Vergleich der nationalen Bewertungen anhand von Makrophyten. Untersucht wurden die Verfahren von Belgien, Deutschland, Frankreich, Großbritannien, Österreich und Polen. Die biologischen Daten umfassten taxonomische Zusammensetzung und Häufigkeit von Fließgewässer-Makrophyten für die Interkalibrierungstypen der silikatischen Mittelgebirgsbäche, der silikatischen Sandbäche des Tieflands sowie der kleinen Flüsse des Tieflands. Bei den Tiefland-Typen beschränkten sich die Analysen auf Gewässerstellen mit mittlerem bis hohem Säurebindungsvermögen (Alkalinität).

Im Vorfeld der Analysen wurden Taxonomie und Häufigkeitsskalen harmonisiert und den Arten ein Grad der Wassergebundenheit („level of aquaticity“) zugewiesen. Alle Vegetationsaufnahmen wurden durch die nationalen Verfahren bewertet, dann wurden innerhalb eines Interkalibrierungstypen alle nationalen Bewertungsergebnisse pro Vegetationsaufnahme gemittelt. In einem nächsten Schritt wurde dieser mittlere Index mit den Häufigkeiten der in den Vegetationsaufnahmen vorkommenden Makrophyten-Arten korreliert. Die lineare Beziehung von Arten-Häufigkeit und mittlerem Index wurde über den Korrelationskoeffizienten nach Spearman gemessen. Die Analyse ergab einen Korrelationskoeffizienten für jede Art und umfasste ein Wertespektrum, welches Arten entweder als positiv, negativ oder nicht signifikant korreliert zum mittleren Index auswies. Die Korrelationskoeffizienten wurden zur Festlegung von Art-spezifischen Indikatorwerten genutzt, welche die Beschreibung von Gewässertyp-spezifischen Makrophytengemeinschaften unter ungestörten bzw. degradierten Bedingungen ermöglichte. Die Indikatorwerte wurden außerdem zur Berechnung des Allgemeinen Metriks „macrophyte Intercalibration Common Metric“ (mICM) verwendet.

Auf Grundlage der Vegetationsaufnahmen wurde der mICM gegen die einzelnen nationalen Bewertungsergebnisse aufgetragen. Lineare und nichtlineare (quadratische) Regressionsmodelle wurden angewendet, anschließend die resultierenden

Bestimmtheitsmaße (R^2) überprüft. Im Falle geringer R^2-Werte wurden die mICM-Indikatorwerte mit den jeweiligen nationalen Werten der entsprechenden Arten verglichen. Deutliche Unterschiede beider Indikatorwerte wurden durch Änderungsvorschläge für die nationalen Werte angeglichen. Allerdings fanden nur solche Änderungen statt, die einen wesentlichen Anstieg des Bestimmtheitsmaßes in den wiederholten Regressionsanalysen zur Folge hatten. Einheitliche Referenzwerte wurden über die Definition von Probestellen im allgemein sehr guten Zustand hergeleitet.

Dieser Ansatz erwies sich als tragfähige Methodik zur Schaffung einer gemeinsamen Basis für die Interkalibrierung. Für die drei Gewässertypen konnte eine umfangreiche Beschreibung der Makrophytengemeinschaften und ihrer Umweltbedingungen im naturnahen und belasteten Zustand erstellt werden. Diese Darstellungen fungierten als Leitbild im Prozess der Harmonisierung der Bewertungsverfahren. Mit dem mICM wurde ein geeigneter Allgemeiner Metrik entwickelt. Auf Grundlage des Leitbildes wurden die Indikatorwerte ausgewählter Arten im belgischen und deutschen Verfahren angepasst. In den Regressionsanalysen wies der mICM einen mittleren R^2-Wert von über 0,6 zu allen nationalen Verfahren auf. Die Werte dieses Metriks zeigten eine geringe Spannweite innerhalb der Probestellen im allgemein sehr guten Zustand. Diese Eigenschaft würde die Definition äquidistanter Klassengrenzen zum Zwecke des Vergleichs mit den nationalen Grenzsetzungen erlauben.

4 Eine neue Methode zum Vergleich von Klassengrenzen biologischer Bewertungsverfahren: ein Fallbeispiel aus dem Donau-Stromgebiet

In der vierten Fallstudie dieser Arbeit wurden die Einstufungen des ökologischen Zustands für verschiedene Makrozoobenthos- und Diatomeen-Verfahren in Osteuropa verglichen und harmonisiert. Grundlage für die Analysen bildeten Daten aus den nationalen Überwachungsprogrammen von Österreich, Bulgarien, Rumänien, der Slowakei und Ungarn. Biologische Aufnahmen von Gewässerstellen in naturnahem Zustand waren nicht verfügbar. Deshalb testete ich einen alternativen Ansatz zur Festlegung von Referenzen für die Interkalibrierung. Für vier Interkalibrierungstypen wurden Probestellen im wenigstens guten Umweltzustand ausgewiesen. Hierzu nutzte ich Grenzwerte für die Parameter Gesamt- und Orthophosphat, Biologischer Sauerstoffbedarf, Leitfähigkeit, hydromorphologischer Zustand und Landnutzungsindex (Böhmer et al., 2004). Der biologische Metrik „Average Score Per Taxon“ (Armitage et al., 1983) wurde als zusätzlicher Parameter für die Gewässerstellen mit Makrozoobenthos-Aufnahmen gewählt. Als Skalen für den Vergleich bzw. die Harmonisierung der nationalen Klassengrenzen dienten Allgemeine Metriks. Für das Makrozoobenthos wurde in dieser Studie ein multimetrischer Interkalibrierungs-Index entwickelt. Die Diatomeen-Verfahren verglich ich mit dem Allgemeinen Metrik, der von Kelly et al. (2008) im mitteleuropäischen Interkalibrierungsprozess angewendet wurde.

Anhand der biologischen Daten wurden sowohl die nationalen Bewertungsverfahren als auch die Allgemeinen Metriks berechnet. Die Verteilungen der Ergebnisse der Allgemeinen Metriks innerhalb der Gewässerstellen im wenigstens guten Umweltzustand ermöglichten die Definition transnationaler Bezugspunkte („biological benchmarks") für die Interkalibrierung. Diese Bezugspunkte dienten zur Normalisierung der Werte der Allgemeinen Metriks. Somit konnten die nationalen Klassengrenzen der Diatomeen-Verfahren Österreichs und der Slowakei in Regressionsanalysen verglichen werden. Dabei zeigten sich zwischen den Verfahren Unterschiede sowohl in der Festsetzung der nationalen Referenzwerte als auch der Werte für die Klassengrenze sehr gut - gut. Beim Makrozoobenthos ermöglichte die „Benchmarking"-Methode die einheitliche Festlegung der Klassengrenzen des guten ökologischen Zustands, ohne Rückgriff auf die nationalen Grenzdefinitionen. Diese Vorgehensweise erlaubte die Interkalibrierung von Ländern, deren Verfahren noch in der Entwicklung standen. Die Ergebnisse dieser Studie bildeten Bestandteil der Entscheidung der Europäischen Kommission zur Festlegung der Grenzwerte des guten ökologischen Zustands (European Commission, 2008).

Schlussbetrachtungen

Meine Untersuchungen zur Fragestellung, wie die nationalen Definitionen des guten ökologischen Zustands am besten verglichen werden können, zeigten unterschiedlich starke numerische Zusammenhänge zwischen den Bewertungsergebnissen der nationalen Verfahren. Makrozoobenthos- und Diatomeen-Verfahren waren untereinander und gegenüber Allgemeinen Metriks hoch korreliert, Makrophyten-Verfahren wiesen schwächere Zusammenhänge auf. Differierende Bewertungskonzepte und -traditionen zwischen den biologischen Qualitätskomponenten sind hier von wesentlicher Bedeutung. Des weiteren untersuchte ich verschiedene Ansätze für eine einheitliche Definition von

Referenzzuständen. Neben der Anwendung von nicht-biologischen Kriterien, welche Gewässerstellen als naturnah oder gering gestört auswiesen, wurden auch rein biologisch festgelegte Referenzzustände genutzt (beste, verfügbare Gewässerstellen; Stellen im allgemein sehr guten Zustand). Die Art ihrer Festlegung ist von zentralem Stellenwert für die Interkalibrierung. Der Vergleich und die Harmonisierung des guten ökologischen Zustands bildeten einen dritten Schwerpunkt innerhalb dieser Arbeit. Alle Untersuchungen offenbarten Unterschiede in den nationalen Festsetzungen der Zustandsklassen. Eine Harmonisierung ließ sich sowohl über den Abgleich mit nicht-biologischen Daten zur Gewässerbelastung als auch über die Definition transnationaler Bezugspunkte erreichen.

Durch den Interkalibrierungsprozess wird ein grenzüberschreitendes Konzept für den guten ökologischen Zustand geschaffen, das von hoher sozioökonomischer Bedeutung ist: Die Europäischen Mitgliedstaaten sind verpflichtet, diesen Zustand zu erhalten oder durch geeignete Maßnahmen wieder herzustellen. Vor diesem Hintergrund ist eine kritische Prüfung der Methoden der Interkalibrierung hinsichtlich ihrer Gültigkeit und Beschränkungen unabdingbar. Die Wasserrahmenrichtlinie basiert auf einer bestimmten Naturwahrnehmung, die vom Konzept einer statischen, nicht-dynamischen Umwelt geprägt ist, und in der ein vom Menschen unbeeinflusster Zustand das Leitbild für menschliches Handeln darstellt. Diese Voraussetzungen schaffen den grundsätzlichen Rahmen für ökologische Zustandsbewertung und Interkalibrierung. Der Einfluss von Unsicherheiten auf die Ergebnisse von Bewertung und Interkalibrierung blieb im Prozess weitgehend unberücksichtigt. Ferner kann die Verpflichtung zur Interkalibrierung innovative Ansätze der Gewässerbewertung verhindern, wenn sich diese als unvergleichbar mit den herkömmlichen Verfahren erweisen. Und letztlich bleibt jede naturwissenschaftliche Festlegung des guten Zustands willkürlich: Das Studium der Natur kann uns keine normativen Aussagen zum richtigen Umgang mit der natürlichen Mitwelt liefern.

Der Interkalibrierungsprozess initiierte einen europaweiten Diskurs über biologische Gewässerbewertung und die Definition des guten ökologischen Zustands. Innerhalb dieses Diskurses bildet die vorliegende Arbeit einen wichtigen Beitrag zur wissenschaftlichen Umsetzung der Vorgaben der EG-Wasserrahmenrichtlinie.

References

Alba-Tercedor, J. & A.M. Pujante, 2000. Running-water biomonitoring in Spain: opportunities for a predictive approach. In Wright, J. F., D. W. Sutcliffe & M. T. Furse (Editors), Assessing the biological quality of fresh waters - RIVPACS and other techniques. FBA, Ambleside: 207-216.

Allan, D.J., 2004. Landscapes and Riverscapes: the influence of land use on stream ecosystems. Annual Review of Ecology, Evolution and Systematics 35: 257-284.

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

Baattrup-Pedersen, A. & T. Riis, 1999. Macrophyte diversity and composition in relation to substratum characteristics in regulated and unregulated Danish streams. Freshwater Biology 42: 375-385.

Baattrup-Pedersen, A., G. Springe, T. Riis, S.E. Larsen, K. Sand-Jensen & L.M. Kjellerup-Larsen, 2008. The search for reference conditions for stream vegetation in northern Europe. Freshwater Biology 53: 1890-1901.

Biggs, J., 2006. European Environmental NGO Technical Review of the Water Framework Directive Intercalibration Process. European Environmental Bureau, Royal Society for the Protection of Birds and Pond Conservation, Brussels.

Biggs, J., A. Corfield, D. Walker, M. Whitfield & P. Williams, 1996. A preliminary comparison of European methods of biological river water quality assessment. NRA Thames Region Operational Investigation. Environment Agency Technical Report No. 0I/T/001. National Rivers Authority Thames Region, Reading.

Birk, S., 2003. Review of European assessment methods for rivers and streams using benthic invertebrates, aquatic flora, fish and hydromorphology. Diploma thesis. University of Duisburg-Essen, Essen.

Birk, S. & D. Hering, 2002. Waterview Web-Database: a comprehensive review of European assessment methods for rivers. FBA News 20: 4.

Birk, S. & P. Rolauffs, 2004. A preliminary study comparing the results between the Austrian, Czech and German saprobic systems for the intercalibration of cross-border river basin districts. In Deutsche Gesellschaft für Limnologie (DGL) - Tagungsbericht (Köln). DGL, Werder: 74-79.

Birk, S. & U. Schmedtje, 2005. Towards harmonization of water quality classification in the Danube River Basin: overview of biological assessment methods for running waters. In: J. Bloesch (Editor), River basin management: concepts and transboundary implementation (with special reference to the Danube River). Archiv für Hydrobiologie Supplement Large rivers, 16: 171-196.

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

Birk, S., N. Willby, C. Chauvin, H.C. Coops, L. Denys, D. Galoux, A. Kolada, K. Pall, I. Pardo, R. Pot & D. Stelzer, 2007a. Report on the Central Baltic River GIG Macrophyte Intercalibration Exercise, June 2007.

Birk, S., J. Böhmer, C. Meier, P. Rolauffs, J. Schaumburg & D. Hering, 2007b. EG-Wasserrahmenrichtlinie - Harmonisierung der Berichterstattung zur ökologischen Einstufung nach EG-Wasserrahmenrichtlinie (Interkalibrierung biologischer Untersuchungsverfahren in Deutschland). University of Duisburg-Essen, Essen.

Birk, S., E. Bellack, J. Böhmer, K. Bunzel, F. Fischer, A. Kolbinger, U. Mischke, J. Schaumburg & C. Schütz, 2009. Die Interkalibrierung nach EG-Wasserrahmenrichtlinie - Ergebnisse der ersten Interkalibrierungsphase 2005-2007. Wasserwirtschaft 5: 20-25.

BMLFUW, 2006a. Leitfaden für die Erhebung der Biologischen Qualitätselemente. Arbeitsanweisung Fließgewässer. A4-01a Qualitätselement Makrophyten: Felderhebung, Probenahme, Probenaufbereitung und Ergebnismitteilung. Dezember 2006. Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft, Wien.

BMLFUW, 2006b. Leitfaden für die Erhebung der Biologischen Qualitätselemente. Arbeitsanweisung Fließgewässer. A2-01a Qualitätselement Makrozoobenthos: Felderhebung, Probenahme, Probenaufbereitung und Ergebnismitteilung. Oktober 2006. Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft, Wien.

BMWP (Biological Monitoring Working Party), 1978. Final Report of the Biological Monitoring Working Party: Assessment and presentation of the biological quality of rivers in Great Britain. Department of the Environmental Water Data Unit, London.

Böhmer, J., C. Rawer-Jost, A. Zenker, C. Meier, C.K. Feld, R. Biss & D. Hering, 2004. Assessing streams in Germany with benthic invertebrates: Development of a multimetric invertebrate based assessment system. Limnologica 34: 416-432.

Borja, A., A.B. Josefson, A. Miles, I. Muxika, F. Olsgard, G. Phillips, J.G. Rodríguez & B. Rygg, 2007. An approach to the intercalibration of benthic ecological status assessment in the North Atlantic ecoregion, according to the European Water Framework Directive. Marine Pollution Bulletin 55: 42-52.

Bossard, M., J. Feranec & J. Otahel, 2000. CORINE land cover technical guide - Addendum 2000. European Environment Agency, Copenhagen.

Brabec, K., S. Zahradkova, D. Nemejcova, P. Paril, J. Kokes & J. Jarkovsky, 2004. Assessment of organic pollution effect considering differences between lotic and lentic stream habitats. Hydrobiologia 516: 331-346.

Braun-Blanquet, J., 1928: Pflanzensoziologie. Grundzüge der Vegetationskunde. Springer, Berlin.

Buffagni, A., S. Erba, M. Cazzola & J. L. Kemp, 2004. The AQEM multimetric system for the southern Italian Apennines: assessing the impact of water quality and habitat degradation on pool macroinvertebrates in Mediterranean rivers. Hydrobiologia 516: 313-329.

Buffagni, A., S. Erba, S. Birk, M. Cazzola, C. Feld, T. Ofenböck, J. Murray-Bligh, M. T. Furse, R. T. Clark, D. Hering, H. Soszka & W. v. d. Bund, 2005. Towards European Inter-calibration for the Water Framework Directive: Procedures and examples for different river types from the E.C. project STAR. 11th STAR deliverable. STAR Contract No: EVK1-CT 2001-00089. Quaderni Istituto di Ricerca sulle Acque 123: 1-468.

Buffagni, A., S. Erba, M. Cazzola, J. Murray-Bligh, H. Soszka & P. Genoni, 2006. The STAR common metrics approach to the WFD intercalibration process: full application for small, lowland rivers in three European countries. In: M.T. Furse, D. Hering, K. Brabec, A. Buffagni, L. Sandin and P.F.M. Verdonschot (Editors), The ecological status of European rivers. Evaluation and intercalibration of assessment methods. Hydrobiologia, 566: 379-399.

Buffagni, A., S. Erba & M.T. Furse, 2007. A simple procedure to harmonize class boundaries of assessment systems at the pan-European scale. Environmental Science and Policy 10: 709-724.

Carstensen, J., 2007. Statistical principles for ecological status classification of Water Framework Directive monitoring data. Marine Pollution Bulletin 55: 3-15.

CB GIG Lakes, 2008. Central Baltic GIG Lakes. WFD intercalibration technical report. Part 2 - Lakes. Section 2 – Chlorophyll-a concentration. Joint Research Centre, Ispra.

CB GIG Rivers, 2008. Central Baltic GIG Rivers. WFD intercalibration technical report. Part 1 - Rivers. Section 2 - Benthic macroinvertebrates. Joint Research Centre, Ispra.

CEMAGREF (Centre National du Machinisme Agricole du Génie Rural, des Eaux et des Forets), 1982. Etude des méthodes biologiques d'appréciation quantitative de la qualité des eaux. Rapport Q.E. Lyon. Division Qualité des Eaux - Pêche et Pisciculture, Lyon.

Clarke, R.T., M.T. Furse, J.F. Wright & D. Moss, 1996. Derivation of a biological quality index for river sites : comparison of the observed with the expected fauna. Journal of Applied Statistics 23: 311-332.

Clarke, R., A. Lorenz, L. Sandin, A. Schmidt-Kloiber, J. Strackbein, N. T. Kneebone & P. Haase, 2006. Effects of sampling and sub-sampling variation using the STAR-AQEM sampling protocol on the precision of macroinvertebrate metrics. In: M.T. Furse, D. Hering, K. Brabec, A. Buffagni, L. Sandin and P.F.M. Verdonschot (Editors), The ecological status of European rivers. Evaluation and intercalibration of assessment methods. Hydrobiologia, 566: 441-459.

Council of the European Communities, 1991. Urban Waste Water Treatment Directive 91/271/EEC. Official Journal of the European Communities, L135/40-52, 30 May 1991, Brussels.

CSN 757716, 1998. Water quality, biological analysis, determination of saprobic index. Czech Technical State Standard, Czech Standards Institute, Prague.

Dell'Uomo, A., 1996. Assessment of water quality of an Apennine river as a pilot study for diatom-based monitoring of Italian watercourses. In: B.A. Whitton & E. Rott (Editors), Use of algae for monitoring rivers II: proceedings of an international symposium held at the Volksbildungsheim Grillhof, Vill near Innsbruck. Rott, Innsbruck: 65-72.

Descy, J.-P. & M. Coste, 1991. A test of methods for assessing water quality based on diatoms. Verhandlungen der internationalen Vereinigung für theoretische und angewandte Limnologie 24: 2112-2116.

Deutsch, K. & N. Kreuzinger, 2005. Leitfaden zur typspezifischen Bewertung der allgemeinen chemisch/physikalischen Parametern in Fließgewässern. Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft, Wien.

ECOSTAT (CIS WG 2.A Ecological Status), 2004a. Guidance on the intercalibration process. Agreed version of WG 2.A Ecological Status meeting held 7-8 October 2004 in Ispra. Version 4.1 - 14 October 2004. ECOSTAT, Ispra.

ECOSTAT (CIS WG 2.A Ecological Status), 2004b. Overview of common intercalibration types. Final version for finalisation of the intercalibration network spring 2004. Version 5.1 - 23 April 2004. ECOSTAT, Ispra.

Elbersen, J.W.H., P.F.M. Verdonschot, B. Roels & J.G. Hartholt, 2003. Definitiestudie KaderRichtlijn Water (KRW). I. Typologie Nederlandse Oppervlaktewateren. Alterra-rapport 669. ALTERRA, Wageningen.

Ellenberg, H., H.E. Weber, R. Düll, V. Wirth, W. Werner & D. Paulißen, 1992. Indicator values of plants in Central Europe. Erich Goltze, Göttingen.

EN 13946:2003. Water quality. Guidance standard for the routine sampling and pre-treatment of benthic diatoms from rivers. European Committee for Standardization, Brussels.

EN 14184:2003, Water quality - Guidance standard for the surveying of aquatic macrophytes in running waters. European Committee for Standardization, Brussels.

EN 27828:1994. Water quality - Methods of biological sampling - Guidance on handnet sampling of aquatic benthic macro-invertebrates (ISO 7828:1985). European Committee for Standardization, Brussels.

European Commission, 2000. Directive 2000/60/EC. Establishing a framework for community action in the field of water policy. Official Journal of the European Communities L327/1:1-73.

European Commission, 2008. Commission Decision of 30 October 2008 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. Official Journal of the European Union L332: 20-44.

European Communities, 2003. River and lakes – Typology, reference conditions and classification systems. Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Guidance document No. 10. Office for Official Publications of the European Communities, Luxembourg.

European Communities, 2005. Guidance on the intercalibration process 2004-2006. Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Guidance document No. 14. Office for Official Publications of the European Communities, Luxembourg.

Europäische Kommission, 2000. Richtlinie 2000/60/EG des Europäischen Parlaments und des Rates vom 23. Oktober 2000 zur Schaffung eines Ordnungsrahmens für Maßnahmen der Gemeinschaft im Bereich der Wasserpolitik. Amtsblatt der Europäischen Gemeinschaften L 327/1: 1-73.

Feld, C.K. & D. Hering, 2007. Community structure or function: effects of environmental stress on benthic macroinvertebrates at different spatial scales. Freshwater Biology 52: 1380–1399.

Ferreira, M.T., P.M. Rodríguez-González, F.C. Aguiar & A. Albuquerque, 2005. Assessing biotic integrity in Iberian rivers: Development of a multimetric plant index. Ecological Indicators 5: 137-149.

Friberg, N., L. Sandin, M. Furse, S.E. Larsen, R.T. Clark & P. Haase, 2006. Comparison of macroinvertebrate sampling methods in Europe. In: M.T. Furse, D. Hering, K. Brabec, A. Buffagni, L. Sandin and P.F.M. Verdonschot (Editors), The ecological status of European rivers. Evaluation and intercalibration of assessment methods. Hydrobiologia, 566: 365-378.

Friedrich, G., E. Coring & B. Küchenhoff, 1995. Vergleich verschiedener europäischer Untersuchungs- und Bewertungsmethoden für Fließgewässer. Landesumweltamt Nordrhein-Westfalen, Essen.

Friedrich, G. & V. Herbst, 2004. Eine erneute Revision des Saprobiensystems - weshalb und wozu? Acta Hydrochimica et Hydrobiologica 32: 61-74.

Furse, M., D. Hering, O. Moog, P. Verdonschot, L. Sandin, K. Brabec, K. Gritzalis, A. Buffagni, P. Pinto, N. Friberg, J. Murray-Bligh, J. Kokes, R. Alber, P. Usseglio-Polatera, P. Haase, R. Sweeting, B. Bis, K. Szoszkiewicz, H. Soszka, G. Springe, F. Sporka & I. Krno, 2006. The STAR project: context, objectives and approaches. In: M.T. Furse, D. Hering, K. Brabec, A. Buffagni, L. Sandin and P.F.M. Verdonschot (Editors), The ecological status of European rivers. Evaluation and intercalibration of assessment methods. Hydrobiologia, 566: 3-29.

Gabriels, W., 2007. Multimetric assessment of freshwater macroinvertebrate communities in Flanders, Belgium. PhD thesis. Ghent University, Ghent.

Ghetti, P. F. & G. Bonazzi, 1977. A comparison between various criteria for the interpretation of biological data in the analysis of the quality of running waters. Water Research 11: 819-831.

Ghetti, P. F. & G. Bonazzi, 1980. Biological water assessment methods: Torrente Parma, Torrente Stirone, Fiume Po. 3rd Technical Seminar. Final Report. Commission of the European Communities, Brussels.

Groffman, P., J. Baron, T. Blett, A. Gold, I. Goodman, L. Gunderson, B. Levinson, M. Palmer, H. Paerl, G. Peterson, N. Poff, D. Rejeski, J. Reynolds, M. Turner, K. Weathers & J. Wiens, 2006. Ecological thresholds: The key to successful environmental management or an important concept with no practical application? Ecosystems 9: 1-13.

Haase, P., J. Murray-Bligh, S. Lohse, S. Pauls, A. Sundermann, R. Gunn & R. Clarke, 2006. Assessing the impact of errors in sorting and identifying macroinvertebrate samples. In: M.T. Furse, D. Hering, K. Brabec, A. Buffagni, L. Sandin and P.F.M. Verdonschot (Editors), The ecological status of European rivers. Evaluation and intercalibration of assessment methods. Hydrobiologia, 566: 505-521.

Hardin, G., 1968. The tragedy of the commons. Science 162: 1243-1248.

Hatton-Ellis, T., 2008. The Hitchhiker's Guide to the Water Framework Directive. Aquatic Conservation-Marine and Freshwater Ecosystems 18: 111-116.

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

Helešic, J., 2006. Biological monitoring of running waters in Eastern and Central European countries (former Communist Block). In: G. Ziglio, M. Siligardi & G. Flaim (Editors), Biological monitoring of rivers: applications and perspectives. John Wiley and Sons, Chichester, pp. 327-350.

Hering, D., A. Buffagni, O. Moog, L. Sandin, M. Sommerhäuser, I. Stubauer, C.K. Feld, R.K. Johnson, P. Pinto, N. Skoulikidis, P.F.M. Verdonschot & S. Zahrádková, 2003. The development of a system to assess the ecological quality of streams based on macroinvertebrates – design of the sampling programme within the AQEM project. International Review of Hydrobiology 88: 345-361.

Hering, D., O. Moog, L. Sandin & P.F.M. Verdonschot, 2004. Overview and application of the AQEM assessment system. Hydrobiologia 516: 1-20.

Hering, D., R.K. Johnson, S. Kramm, S. Schmutz, K. Szoszkiewicz & P.F.M. Verdonschot, 2006a. Assessment of European rivers with diatoms, macrophytes, invertebrates and fish: A comparative metric-based analysis. Freshwater Biology 51: 1757-1785.

Hering, D., C.K. Feld, O. Moog & T. Ofenböck, 2006b. Cook book for the development of a Multimetric Index for biological condition of aquatic ecosystems: experiences from the European AQEM and STAR projects and related initiatives. In: M.T. Furse, D. Hering, K. Brabec, A. Buffagni, L. Sandin and P.F.M. Verdonschot (Editors), The ecological status of European rivers. Evaluation and intercalibration of assessment methods. Hydrobiologia, 566: 311-324.

Holmes, N.T.H. & B.A. Whitton, 1975. Macrophytes of the river Tweed. Transactions of the Botanical Society of Edinburgh 42: 369-381.

Holmes, N.T.H., J.R. Newman, S. Chadd, K.J. Rouen, L. Saint & F.H. Dawson, 1999. Mean Trophic Rank: A User's Manual. R & D Technical Report E38. Environment Agency, Bristol.

Hume, D., 1978. A treatise of human nature. Oxford University Press, Oxford.

ICPDR (International Commission for the Protection of the Danube River), 2004. Water quality in the Danube River Basin - TNMN yearbook 2001. ICPDR, Vienna.

Illies, J. (Editor), 1967. Limnofauna Europaea. G. Fischer, Stuttgart.

Janauer, G.A., 2001. Is what has been measured of any direct relevance to the success of the macrophyte in its particular environment? Journal of Limnology 60 (Suppl.): 33-38.

Janauer, G.A., P. Hale & R. Sweeting (Editors), 2003. Macrophyte inventory of the river Danube: A pilot study. Archiv für Hydrobiologie, Supplement “Large Rivers” 147 (1-2): 1-229.

Just, I., F. Schöll & T. Tittitzer, 1998. Versuch einer Harmonisierung nationaler Methoden zur Bewertung der Gewässergüte im Donauraum am Beispiel der Abwässer der Stadt Budapest. Umweltbundesamt, Berlin.

Kant, I., 1800. Immanuel Kants Logik. Ein Handbuch zu Vorlesungen. Friedrich Nicolovicus, Königsberg.

Kelly, M.G. & B.A. Whitton, 1998. Biological monitoring of eutrophication in rivers. Hydrobiologia 384: 55-67.

Kelly, M., C. Bennett, M. Coste, C. Delgado, F. Delmas, L. Denys, L. Ector, C. Fauville, M. Ferréol, M. Golub, A. Jarlman, M. Kahlert, J. Lucey, B. Ní Chatháin, I. Pardo, P. Pfister, J. Picinska-Faltynowicz, J. Rosebery, C. Schranz, J. Schaumburg, H. van Dam & S. Vilbaste, 2008. 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. DOI 10.1007/s10750-008-9641-4

Knoben, R.A.E., C. Roos & M.C.M. van Oirschot, 1995. Biological assessment methods for watercourses. UN/ECE Task Force on Monitoring and Assessment, Lelystad.

Knoben, R.A.E., L. Bijlmakers & P. van Meenen, 1999. Water Quality Enhancement in the Danube River Basin; sub action 2A: Water quality classification/characterisation. IWACO, 's-Hertogenbosch.

Kohler, A., 1978. Methoden der Kartierung von Flora und Vegetation von Süßwasserbiotopen. Landschaft & Stadt 10: 73-85.

Korte, T. & K. van de Weyer, 2005. Die Bewertung von Fließgewässern mit Makrophyten gemäß EU-WRRL - Ergebnisse des Vergleichs von zwei Bewertungsverfahren. Wasser und Abfall 9/2005: 46-49.

Kownacki, A., H. Soszka, D. Kudelska & T. Fleituch, 2004. Bioassessment of Polish rivers based on macroinvertebrates. In Geller, W. et al. (Editors), Proceedings of the international 11th Magdeburg Seminar on Waters in Central and Eastern Europe: Assessment, Protection, Management. 18-22 October 2004, UFZ Leipzig: 250-251.

Lacoul, P. & B. Freedman, 2006. Environmental influences on aquatic plants in freshwater ecosystems. Environmental Reviews 14: 89-136.

LAWA-AO, 2007. RaKon Monitoring Teil B. Arbeitspapier II: Hintergrund- und Orientierungswerte für physikalisch-chemische Komponenten. Stand 07.03.2007.

Ständiger Ausschuss "Oberflächengewässer und Küstengewässer" der Bund/Länder-Arbeitsgemeinschaft Wasser, Mainz.

Leyssen, A., P. Adriaens, L. Denys, J. Packet, A. Schneiders, K. Van Looy & L. Vanhecke, 2005. Toepassing van verschillende biologische beoordelingssystemen op Vlaamse potentiële interkalibratielocaties overeenkomstig de Europese kaderrichtlijn water : partim 'Macrofyten'. Instituut voor Natuurbehoud, Brussels.

Lorenz, A., D. Hering, C.K. Feld, C.K. & P. Rolauffs, 2004. A new method for assessing the impact of hydromorphological degradation on the macroinvertebrate fauna of five German stream types. Hydrobiologia 516: 107-127.

Meilinger, P., S. Schneider & A. Melzer, 2005. The reference index method for the macrophyte-based assessment of rivers – a contribution for the implementation of the European Water Framework Directive in Germany. International Review of Hydrobiology 90: 322-342.

Melzer, A., R. Harlacher, K. Held, R. Sirch & E. Vogt, 1986. Die Makrophytenvegetation des Chiemsees. Informationsbericht des Bayerischen Landesamts für Wasserwirtschaft 4/86. BLfW, München.

Metcalfe-Smith, J.L., 1994. Biological water-quality assessment of rivers: Use of macroinvertebrate communities. In Calow, P. & G.E. Petts (Editors), The Rivers Handbook - hydrological and ecological principles. Blackwell Scientific Publications, Oxford: 144-170.

MMGA (Ministeriul Mediului si Gospodaririi Apelor), 2006. Anexa la Ordinul ministrului mediului si gospodaririi apelor nr. 161/2006 pentru aprobarea Normativului privind clasificarea calitaii apelor de suprafata in vederea stabilirii starii ecologice a corpurilor de apa. Ministry of Environment and Water Management, Bucharest.

Molen, D. T. van der, M. Beers, M.S. v. d. Berg, T. v. d. Broek, R. Buskens, H.C. Coops, H. v. Dam, G. Duursema, M. Fagel, T. Ietswaart, M. Klinge, R.A.E. Knoben, J. Kranenbarg, J. d. Leeuw, R. Noordhuis, R.C. Nijboer, R. Pot, P.F.M. Verdonschot & T. Vriese, 2004. Referenties en maatlatten voor rivieren ten behoeve van de Kaderrichtlijn Water - version July 2004. Alterra, Wageningen.

Moog, O., A. Chovanec, J. Hinteregger & A. Römer, 1999. Richtlinie zur Bestimmung der saprobiologischen Gewässergüte von Fliessgewässern. Bundesministerium für Land- und Forstwirtschaft, Wien.

Morpurgo, M., 1996. Confronto fra Indice Saprobico (Friedrich e DIN, 1990) e Indice Biotico Esteso (Ghetti e IRSA, 1995). Biologia Ambientale 14: 30-36.

Moss, B. 2007. Shallow lakes, the water framework directive and life. What should it all be about? Hydrobiologia 584: 381-394.

National Rivers Authority, 1994. The quality of rivers and canals in England and Wales (1990 to 1992) as assessed by a new general quality assessment scheme. HMSO, London.

Newman, P.J., 1988. Classification of surface water quality. Review of schemes used in EC member states. Heinemann Professional Publishing Ltd., Oxford.

NF T90-395, 2003. Water quality - Determination of the Macrophytes biological index for rivers (IBMR). Association Française de Normalisation (AFNOR), Saint Denis La Plaine.

Nieuwenhuis, R., 2005. ECOSURV Manual for Sampling and Determination. Hungarian Ministry for Environment and Water, Budapest.

Nixon, S.C., C.P. Mainstone, T. Moth Iversen, P. Kristensen, E. Jeppesen, N. Friberg, E. Papathanassiou, A. Jensen & F. Pedersen, 1996. The harmonized monitoring and classification of ecological quality of surface waters in the European Union. Final Report. European Commission Directorate General XI, Brussels.

ÖNORM M6232:1997. Richtlinien für die ökologische Untersuchung und Bewertung von Fließgewässern. Österreichisches Normungsinstitut, Wien.

Owen, R., C. Bennett, S. Birk, A. Buffagni, S. Erba, N. Mengin, J. Murray-Bligh, G. Ofenböck, I. Pardo, W. van de Bund, F. Wagner & J.-G. Wasson, 2010. Intercalibration of ecological class boundaries for different European river macro-invertebrate assessment methods used to implement the Water Framework Directive. Submitted to Hydrobiologia.

Paal, J. & T. Trei, 2004. Vegetation of Estonian watercourses; the drainage basin of the southern coast of the Gulf of Finland. Annales Botanici Fennici 41: 157-177.

Pall, K., Moser, V., Schaumburg, J., Schranz, C., & P. Meilinger, 2005. Ergebnisse zur Interkalibrierung der Fließgewässerbewertung mit Makrophyten (Option 3: Vergleich Deutschland-Österreich). Oral presentation held at the conference of the “Deutsche Gesellschaft für Limnologie” in Karlsruhe, 28 September 2005.

Pantle, R. & H. Buck, 1955. Die biologische Überwachung der Gewässer und die Darstellung der Ergebnisse. Gas- und Wasserfach 96: 604.

Pielou, E.C., 1966. The measurement of diversity in different types of biological collections. Journal of Theoretical Biology 13: 131-144.

Pollard, P. & W. van de Bund, 2005. Template for the development of a boundary setting protocol for the purposes of the intercalibration exercise. Common Implementation Strategy - Working Group A ECOSTAT. Version 1.2.

Pot, R., 2005. QBWat - ecologische beoordeling van waterkwaliteit conform de Europese Kaderrichtlijn Water. Version 1.01.

Rico, E., A. Rallo, M.A. Sevillano & M.L. Arretxe, 1992. Comparison of several biological indices based on river macroinvertebrate benthic community for assessment of running water quality. Annales de Limnologie 28: 147-156.

Rolauffs, P., D. Hering, M. Sommerhäuser, S. Rödiger & S. Jähnig, 2003. Entwicklung eines leitbildorientierten Saprobienindexes für die biologische Fließgewässerbewertung. Umweltbundesamt, Berlin.

Rothschein, J., 1962. Saprobiologische Charakteristik der fließenden Gewässer im Einzugsgebiete des Flusses Bodrog auf der Basis von Zoobenthosanalysen. Technology of Water (Institute Chemical Technology, Prague) 6: 227-275.

Rott, E., G. Hofmann, K. Pall, P. Pfister & E. Pipp, 1997. Indikationslisten für Aufwuchsalgen in Österreichischen Fliessgewässern, Teil 1: Saprobielle Indikation. Bundesministerium für Land- und Forstwirtschaft, Wasserwirtschaftskataster, Wien.

Rott, E., P. Pfister, H. van Dam, E. Pipp, K. Pall, N. Binder, N. & K. Ortler, 1999. Indikationslisten für Aufwuchsalgen in Österreichischen Fliessgewässern, Teil 2: Trophieindikation und autökologische Anmerkungen. Bundesministerium für Land- und Forstwirtschaft, Wasserwirtschaftskataster, Wien.

Sandin, L. & D. Hering, 2004. Comparing macroinvertebrate indices to detect organic pollution across Europe: a contribution to the EC Water Framework Directive intercalibration. Hydrobiologia 516: 55-68.

Schaumburg, J., C. Schranz, J. Foerster, A. Gutowski, G. Hofmann, P. Meilinger, S. Schneider & U. Schmedtje, 2004. Ecological classification of macrophytes and phytobenthos for rivers in Germany according to the Water Framework Directive. Limnologica 34: 283-301.

Schaumburg, J., U. Schmedtje, B. Köpf, C. Schranz, S. Schneider, P. Meilinger, D. Stelzer, G. Hofmann, A. Gutowski & J. Foerster, 2005. Makrophyten und Phytobenthos in Flüssen und Seen. Leitbildbezogenes Bewertungsverfahren zur Umsetzung der EG-Wasserrahmenrichtlinie. Informationsbericht Heft 1/05. Bayerisches Landesamt für Wasserwirtschaft, München.

Schaumburg, J., C. Schranz, D. Stelzer, G. Hofmann, A. Gutowski & J. Foerster, 2006. Instruction Protocol for the ecological Assessment of Running Waters for Implementation of the EC Water Framework Directive: Macrophytes and Phytobenthos. Bavarian Water Management Agency, Munich.

Schmedtje, U. (Editor), 2005. The Danube River Basin District. Part A - Basin-wide overview (WFD Roof Report 2004). International Commission for the Protection of the Danube River, Vienna.

Schmidt-Kloiber, A., A. Lorenz & W. Graf, 2006. The AQEM/STAR taxalist - a pan-European macro-invertebrate ecological database and taxa inventory. In: M.T. Furse, D. Hering, K. Brabec, A. Buffagni, L. Sandin and P.F.M. Verdonschot (Editors), The ecological status of European rivers. Evaluation and intercalibration of assessment methods. Hydrobiologia, 566: 325-342.

Schneider, S., 2000. Entwicklung eines Makrophytenindex zur Trophieindikation in Fließgewässern. Shaker Verlag, Aachen.

Shannon, C.E. & W. Weaver, 1949. Mathematical theory of communication. University of Illinois Press, Urbana.

Simpson, E. H., 1949. Measurement of diversity. Nature 163: 688.

Skriver, J., N. Friberg & J. Kirkegaard, 2000. Biological assessment of running waters in Denmark: Introduction of the Danish Stream Fauna Index (DSFI). Verhandlungen der Internationalen Vereinigung für theoretische und angewandte Limnologie 27: 1822-1830.

Sládeček, V., 1973. System of water quality from the biological point of view. Archiv für Hyrdrobiologie Beihefte (Ergebnisse der Limnologie) 7: 1-218.

Sporka, F., H. E. Vlek, E. Bulánková & I. Krno, 2006. Influence of seasonal variation on bioassessment of streams using macroinvertebrates. In: M.T. Furse, D. Hering, K. Brabec, A. Buffagni, L. Sandin and P.F.M. Verdonschot (Editors), The ecological status of European rivers. Evaluation and intercalibration of assessment methods. Hydrobiologia, 566: 543-555.

Steyaert, P. & G. Ollivier, 2007. The European Water Framework Directive: How ecological assumptions frame technical and social change. Ecology and Society 12: 25.

STN (Slovenská Technická Norma) 83 0532-1 to 8, 1978/79. Biologický rozbor povrchovej vody. (Biological analysis of surface water quality.) Slovak Standardisation Institute, Bratislava.

Stubauer, I. & O. Moog, 2000. Taxonomic sufficiency versus need for information - comments based on Austrian experience in biological water quality monitoring. Internationale Vereinigung für theoretische und angewandte Limnologie: Verhandlungen 27: 1-5.

Swedish Environmental Protection Agency, 2000. Environmental quality criteria: Lakes and watercourses. Swedish Environmental Protection Agency, Stockholm.

SYSTAT Software Inc., 2002. TableCurve 2D - Version 5.01. SSI, Richmond CA.

Szoszkiewicz, K., T. Ferreira, T. Korte, A. Baattrup-Pedersen, J. Davy-Bowker & M. O'Hare, 2006a. European river plant communities: the importance of organic pollution and the usefulness of existing macrophyte metrics. In: M.T. Furse, D. Hering, K. Brabec, A. Buffagni, L. Sandin and P.F.M. Verdonschot (Editors), The ecological status of European rivers. Evaluation and intercalibration of assessment methods. Hydrobiologia, 566: 211-234.

Szoszkiewicz K., J. Zbierska, S. Jusik, T. Zgoła, 2006b. Development of the methodological basis for the biological monitoring of water bodies using macrophytes and its application to water bodies of certain categories and types. Second phase, volume 2: Rivers. Institute of Environmental Protection, Agricultural Academy "A. Cieszkowicz", Warsaw. [in Polish]

Tansley, A.G., 1946: Introduction to plant ecology. Allen & Unwin, London.

Thiebaut, G., F. Guérold & S. Muller, 2002. Are trophic and diversity indices based on macrophyte communities pertinent tools to monitor water quality? Water Research 36: 3602-3610.

Tittizer, T., 1976. Comparative study of biological-ecological water assessment methods. Practical demonstration on the river Main. 2-6 June, 1975 (summary report). In Amavis, R.-J. (Editor), Principles and methods for determining ecological criteria on hydrobiocoenosis: Proceedings of the European Scientific Colloquium Luxembourg, Nov. 1975. Pergamon Press, Oxford: 403-463.

Uzunov, J., 1979. Aquatic Oligochaeta - a supplement to the list of limnosaprobic bioindicators. Doklady Bolgarskoi Akademii Nauk 32: 1101-1103.

Valsangiacomo, A., 1998. Die Natur der Ökologie: Anspruch und Grenzen ökologischer Wissenschaften. Hochschulverlag, ETH Zürich, Zürich.

Van de Weyer, K., 2003. Kartieranleitung zur Erfassung und Bewertung der aquatischen Makrophyten der Fließgewässer in Nordrhein-Westfalen gemäß den Vorgaben der EU-Wasser-Rahmenrichtlinie. LUA-Merkblätter Nr. 39. Landesumweltamt (LUA) NRW, Düsseldorf.

Van den Berg, M.S., H.C. Coops, R. Pot, W. Altenburg, R. Nijboer, T. v. d. Broek, M. Fagel, G. Arts, R. Bijkerk, H. v. Dam, T. Ietswaart, J. v. d. Molen, K. Wolfstein, D. d. Jong & H. Hartholt, 2004. Achtergronddocument referenties en maatlatten waterflora. RIZA, Lelystad.^

Vlek, H. E., F. Sporka & I. Krno, 2006. Influence of macroinvertebrate sample size on bioassessment of streams. In: M.T. Furse, D. Hering, K. Brabec, A. Buffagni, L. Sandin and P.F.M. Verdonschot (Editors), The ecological status of European rivers. Evaluation and intercalibration of assessment methods. Hydrobiologia, 566: 523-542.

White, L. T., 1967. The historical roots of our ecologic crisis. Science 155: 1203-1207.

Wiegleb, G., 1984. A study of habitat conditions of the macrophytic vegetation in selected river systems in Western Lower Saxony (Federal Republic of Germany). Aquatic Botany 18: 313-352.

Wiegleb, G., 1988. Analysis of flora and vegetation in rivers: concepts and applications. In Symoens, J.J. (Editor), Vegetation of inland waters. Kluwer Academic Publishers, Dordrecht: 311-340.

Wiegleb, G., 1991. Die Lebens- und Wuchsformen der makrophytischen Wasserpflanzen und deren Beziehung zur Ökologie, Verbreitung und Vergesellschaftung der Arten. Tuexenia 11: 135-147.

Willby, N., J. Hilton, J.-A. Pitt & G. Philipps, 2006. Summary of approach used in LEAFPACS for defining ecological quality of rivers and lakes using macrophyte composition. Interim Report June 2006. University of Stirling, Stirling.

Willby, N., J.-A. Pitt & G. Philipps, 2008. Development of a system for the classification of lakes and rivers in the UK using aquatic macrophytes. Part II, Rivers. Scientific report. University of Stirling, Stirling.

Woodiwiss, F.S., 1978. Comparative study of biological-ecological water quality assessment methods. Second practical demonstration. Summary Report. Commission of the European Union, Brussels.

Wright, J.F., D.W. Sutcliffe & M.T. Furse (Editors), 2000. Assessing the biological quality of fresh waters: RIVPACS and other techniques. Freshwater Biological Association, Ambleside.

Zelinka, M. & P. Marvan, 1961. Zur Präzisierung der biologischen Klassifikation der Reinheit fließender Gewässer. Archiv für Hydrobiologie 57: 389-407.

Appendix:
Common type-specific mICM indicator taxa scores analysed in Chapter 3

Taxon name	Aquaticity	mICM 1x2	mICM 3	mICM 4x2
Acorus calamus L.	4	-	-	0.021
Agrostis stolonifera L.	4	0.086	-0.421	-0.079
Alisma lanceolatum With.	4	0.053	-	-0.213
Alisma plantago-aquatica L.	4	-0.184	-	0.169
Amblystegium fluviatile (Hedw.) Schimp.	1	-	-0.086	0.323
Amblystegium riparium (Hedw.) B.S.G.	2	-0.091	-0.866	0.153
Amblystegium tenax (Hedw.) C. E. O. Jensen	1	-	0.008	-
Aneura pinguis (L.) Dumort.	2	-	-0.004	-
Angelica sylvestris L.	5	-0.139	-	-0.105
Apium nodiflorum (L.) Lag.	2	0.198	-0.145	-
Audouinella sp. Bory	1	-	0.034	-
Batrachospermum sp. Roth	1	-	-0.034	-
Berula erecta (Huds.) Coville	2	-0.011	-	0.160
Bidens cernua L.	5	0.109	-	-0.170
Bidens frondosa L.	5	0.045	-	0.086
Bidens tripartita L.	5	-0.293	-	-0.056
Brachythecium plumosum (Hedw.) B.S.G.	1	-	0.570	-
Brachythecium rivulare Schimp.	2	-	0.388	0.277
Butomus umbellatus L.	4	-0.195	-	-0.125
Caltha palustris L.	4	0.450	0.142	0.038
Callitriche cophocarpa Sendtn.	1	-0.295	-0.070	0.007
Callitriche hamulata Kuetz. ex W.D.J. Koch	1	1.000	0.014	0.336
Callitriche obtusangula Le Gall	1	0.125	-0.277	-0.183
Callitriche palustris L.	1	0.231	-	-
Callitriche platycarpa Kuetz.	1	0.193	-0.280	0.236
Callitriche stagnalis Scop.	1	0.202	-0.121	-
Cardamine amara L.	5	0.467	-0.067	0.076
Carex rostrata Stokes	4	0.144	-0.050	0.332
Carex vesicaria L.	4	-	-	-0.085
Ceratophyllum demersum L.	1	-0.205	-0.291	-0.492
Chara sp. L. ex Vaillant	1	-	-	0.190
Chiloscyphus polyanthos (L.) Corda.	1	-	0.595	-
Cinclidotus fontinaloides (Hedw.) P. Beauv.	1	-	-0.172	-
Cladophora glomerata (L.) Kuetz.	1	-	-0.228	-0.070
Cladophora sp. Kuetz.	1	-0.197	-0.428	-0.198
Collema fluviatile (Huds.) Steud.	3	-	-0.104	-
Conocephalum conicum (L.) Dum.	5	-	0.139	-
Cratoneuron filicinum (Hedw.) Spruce	1	-	-0.109	-
Dermatocarpon sp. Eschw.	3	-	0.159	-

Taxon name	Aquaticity	mICM 1x2	mICM 3	mICM 4x2
Diatoma sp. Lyngb.	1	-	-0.100	-0.242
Draparnaldia sp. Bory de St Vincent	1	-	0.166	-
Drepanocladus aduncus (Hedw.) Warnst.	3	-	0.125	-
Eleocharis acicularis (L.) Roem & Schult	5	0.202	-	-
Eleocharis palustris (L.) Roem & Schult	4	-0.096	-0.129	-0.102
Elodea canadensis Michx.	1	-0.152	-0.138	0.027
Elodea nuttallii (Planch.) H. St. John	1	-0.267	-0.286	-0.543
Enteromorpha sp. Link	1	-	-	-0.331
Epilobium hirsutum L.	5	0.092	-0.215	-0.301
Epilobium palustre L.	5	0.068	-	-0.145
Equisetum arvense L.	5	-0.084	-0.202	-
Equisetum fluviatile L.	2	0.207	-0.041	0.434
Equisetum palustre L.	2	0.179	-0.145	0.009
Eupatorium cannabinum L.	5	-0.109	-0.076	-0.054
Fissidens crassipes Wils. ex B.S.G.	1	-	-0.046	-0.071
Fissidens pusillus (Wils.) Milde	2	-	0.080	-
Fissidens rivularis (Spruce) B.S.G.	2	-	-0.023	-
Fissidens rufulus B.S.G.	2	-	0.218	-
Fontinalis antipyretica Hedw.	1	0.018	-0.413	0.933
Fontinalis squamosa Hedw.	1	-	0.525	-
Galium palustre L.	4	-0.170	-0.173	-0.140
Glyceria aquatica (L.) Wahlb.	4	-0.421	-0.203	-0.316
Glyceria fluitans (L.) R. Br.	2	0.192	-0.039	0.345
Hildenbrandia sp. Nardo	1	-0.043	-0.060	0.795
Hottonia palustris L.	1	0.087	-	-
Hydrocharis morsus-ranae L.	1	-0.174	-	0.085
Hygrohypnum duriusculum (De Not.) Jamieson	2	-	-0.010	-
Hygrohypnum luridum (Hedw.) Jenn.	2	-	0.292	-
Hygrohypnum ochraceum (Wils.) Loeske	1	-	0.501	-
Hydrodictyon sp. Roth	1	-	-	-0.304
Hyocomium armoricum (Brid.) Wijk & Marg.	3	-	0.559	-
Hydrocotyle ranunculoides L.	5	-0.354	-	-
Hydrurus sp. C.A. Agardh	1	-	0.102	-
Iris pseudacorus L.	4	0.083	-0.182	-0.007
Isothecium myosuroides Brid.	2	-	0.249	-
Juncus acutiflorus Ehrh. ex Hoffm.	5	-	-0.127	-
Juncus articulatus L.	5	0.156	0.110	0.076
Juncus bufonius L.	5	-	-	-0.032
Juncus bulbosus L.	4	-	0.170	-
Juncus conglomeratus L.	5	-	0.128	-
Juncus effusus L.	4	-0.134	-0.123	-0.093
Lemanea sp. Bory de St Vincent	1	-	0.173	-
Leersia oryzoides (L.) Sw.	5	-	-	0.181
Lemna gibba L.	1	-0.001	-	-0.369
Lemna minor L.	1	-0.735	-0.410	-0.581

Taxon name	Aquaticity	mICM 1x2	mICM 3	mICM 4x2
Lemna minuta Kunth	1	-0.521	-	-0.211
Lemna trisulca L.	1	-0.106	-	0.153
Littorella uniflora (L.) Asch.	1	-	0.078	-
Lunularia cruciata (L.) Dum.	5	-	-0.224	-
Luronium natans (L.) Rafin.	2	0.086	-	-
Lycopus europaeus L.	4	0.096	-0.218	0.001
Lyngbya sp. C.A. Agardh ex Gomont	1	-	0.052	-
Lysimachia nummularia L.	5	-	-0.009	-
Lysimachia thyrsiflora L.	4	-	-	0.370
Lysimachia vulgaris L.	5	0.070	-0.108	0.067
Lythrum salicaria L.	5	-0.121	-0.154	-0.052
Marchantia polymorpha L.	5	-	0.099	-
Marsupella emarginata (Ehrh.) Dum.	2	-	0.426	-
Melosira sp. C.A. Agardh	1	-	-0.293	-0.228
Mentha aquatica L.	4	0.195	-0.232	0.558
Mentha longifolia (L.) Huds. em. Harley	4	-	-0.212	-
Microspora sp. Balbiani	1	-	-0.097	-
Mnium hornum Hedw.	2	-	0.321	-
Montia fontana L.	2	-	0.108	-
Mougeotia sp. C.A. Agardh	1	-	0.225	-
Myosotis scorpioides L.	2	0.050	-0.249	-0.139
Myriophyllum alterniflorum DC.	1	0.233	0.120	-
Myriophyllum aquaticum (Vell.) Verdc.	1	-0.030	-	-0.204
Myriophyllum spicatum L.	1	-0.159	-0.321	-0.092
Myriophyllum verticillatum L.	1	0.078	-0.200	-
Nasturtium officinale R. Br.	2	-0.014	0.007	-0.181
Nitella flexilis (L.) C.A. Agardh	1	0.134	-	-
Nitella sp. C.A. Agardh	1	0.205	-	-
Nostoc sp. Vaucher ex Born& & Flahault	1	-	0.025	-
Nuphar lutea (L.) Sibth. & Sm.	1	-0.260	-0.206	0.030
Nymphoides peltata (Gmel.) Kuntze	1	-0.308	-	-
Octodiceras fontanum (La Pyl.) Lindb.	1	-	-	-0.083
Oedogonium sp. Link	1	-	-0.398	-0.546
Oenanthe aquatica (L.) Poiret	1	0.179	-	-0.223
Oenanthe crocata L.	2	-	0.103	-
Oscillatoria sp. Vaucher ex Gomont	1	-	0.030	-0.401
Pellia endiviifolia (Dicks) Dumort	2	0.301	-0.002	0.268
Pellia epiphylla L. Corda	2	-	0.598	-
Petasites hybridus (L.) Gaertn., Mey. & Scherb.	5	-	-0.037	-
Peucedanum palustre (L.) Moench	5	0.029	-	0.223
Phalaris arundinacea L.	4	0.052	-0.540	-0.284
Philonotis caespitosa Jur.	2	-	0.219	-
Philonotis gr. fontana (Hedw.) Brid.	1	-	0.191	-
Phormidium sp. Kuetz. ex Gomont	1	-	0.127	-
Phragmites australis (Cav.) Trin. ex Steud	4	0.364	-	-0.167

Taxon name	Aquaticity	mICM 1x2	mICM 3	mICM 4x2
Plagiomnium rostratum (Schrad.) T.Kop.	3	-	0.012	-
Plagiomnium undulatum (Hedw.) Kop.	3	-	0.199	-
Polygonum amphibium L.	2	-0.191	-0.168	-0.246
Polygonum hydropiper L.	4	0.181	-0.319	0.087
Potamogeton alpinus Balbis	1	0.348	-	0.566
Potamogeton berchtoldii Fieber	1	-0.222	-	0.038
Potamogeton crispus L.	1	-0.333	-0.299	-0.444
Potamogeton lucens L.	1	-	-	0.083
Potamogeton natans L.	1	0.029	-0.185	0.053
Potamogeton obtusifolius Mert. & Koch	1	-0.044	-	0.213
Potamogeton panormitanus Biv.	1	-0.004	-	-0.006
Potamogeton pectinatus L.	1	-0.740	-	-1.000
Potamogeton perfoliatus L.	1	-0.541	-	0.186
Potamogeton polygonifolius Pourret	1	0.134	0.097	-
Potamogeton praelongus Wulfen	1	-	-	0.325
Potamogeton trichoides Cham. & Schltdl	1	-0.514	-	-0.315
Racomitrium aciculare (Hedw.) Brid.	3	-	0.788	-
Ranunculus aquatilis L.	1	-	-0.022	0.358
Ranunculus circinatus Sibth.	1	0.051	-	-0.084
Ranunculus flammula L.	4	-	0.340	-
Ranunculus fluitans Lamk.	1	-	-0.147	0.099
Ranunculus lingua L.	5	0.287	-	0.160
Ranunculus omiophyllus Ten.	1	-	0.009	-
Ranunculus peltatus Schrank	1	0.230	-0.100	0.242
Ranunculus penicillatus (Dumort.) Bab.	1	0.042	-0.302	-0.224
Ranunculus penicillatus (Dumort.) Bab. *var. penicillatus*	1	-	-0.158	-
Ranunculus penicillatus ssp. pseudofluitans (Syme) S.D. Webster	1	-	-0.164	-
Ranunculus sceleratus L.	5	0.075	-	-0.295
Ranunculus trichophyllus (Chaix) Grey	1	-	-0.245	0.301
Rhizoclonium sp. Kuetz.	1	-	-0.330	-0.342
Rhynchostegium riparioides (Hedw.) Cardo	1	0.229	0.103	0.165
Rhizomnium punctatum (Hedw.) T. Kop.	3	-	0.107	-
Riccardia chamedryfolia (With.) Grolle	2	-	0.097	-
Riccia fluitans L.	2	-0.107	-	-0.299
Rorippa amphibia (L.) Besser	4	-0.524	-0.077	-0.351
Rumex hydrolapathum Huds.	5	-0.043	-	-0.244
Sagittaria sagittifolia L.	2	-0.247	-	-0.571
Scapania undulata (L.) Dum	1	-	1.000	-
Scirpus fluitans L.	1	0.217	-	-
Scirpus lacustris L.	1	-	-	0.116
Scirpus sylvaticus L.	5	0.232	-0.166	0.027
Scrophularia auriculata L.	5	-	-	-0.051
Schistidium rivulare (Brid.) Podp.	3	-	-0.062	-
Scytonema sp. C.A. Agardh ex Bornet & Flahault	1	-	0.036	-
Solanum dulcamara L.	5	-0.088	-0.293	-0.104

Taxon name	Aquaticity	mICM 1x2	mICM 3	mICM 4x2
Sparganium emersum Rehmann	2	-0.445	-0.216	0.179
Sparganium emersum Rehmann *f. longissimum*	2	-	-0.301	-
Sparganium erectum L.	4	-0.164	-0.250	-0.032
Sphagnum sp. L.	2	-	0.333	-
Spirogyra sp. Link	1	-	-0.172	-
Spirodela polyrhiza (L.) Schleid	1	-0.374	-0.266	-0.335
Stigeoclonium sp. Kuetz.	1	-	-0.123	-
Stigeoclonium tenue (C.A. Agardh) Kuetz.	1	-	-0.112	-
Tetraspora sp. Link ex Descaux	1	-	0.006	-
Thamnobryum alopecurum (Hedw.) Gang.	2	-	0.249	-
Thelypteris palustris (Gray) Schott	5	-0.110	-	-
Tolypothrix sp. Kuetz. ex Bornet & Flahault	1	-	0.041	-
Tribonema sp. Drebes & Solier	1	-	0.044	-
Typha angustifolia L.	4	-0.199	-	-
Typha latifolia L.	4	-0.141	-0.153	-0.243
Ulothrix sp. Kuetz.	1	-	-0.078	-
Utricularia sp. L.	1	-0.232	-	-
Utricularia vulgaris L.	1	-0.171	-	-
Vaucheria sp. DC.	2	-	-0.272	-0.173
Veronica anagallis-aquatica L.	2	0.106	-0.157	0.388
Veronica beccabunga L.	2	0.184	-0.338	0.161
Verrucaria sp. F.H. Wigg.	3	-	0.008	-
Zannichellia palustris L.	1	-	-	-0.314

Lebenslauf

Persönliche Daten

- Name: Sebastian Birk
- Geburtsdatum: 10. April 1974
- Geburtsort: Bottrop
- Anschrift: Hüttengasse 23, Hamm (Sieg)
- Familienstand: verheiratet, 3 Kinder

Ausbildung und Berufstätigkeit

- Seit 08/03 - Wissenschaftlicher Mitarbeiter: Universität Duisburg Essen, FB Biologie & Geographie, Abteilung Hydrobiologie; Mitarbeit an verschiedenen Forschungsprojekten zur Umsetzung der EG-Wasserrahmenrichtlinie
- 08/03 - Abschluss, Universität Duisburg-Essen: Diplom-Umweltwissenschaftler (Note: sehr gut)
- 01/03 – 08/03 - Diplomarbeit Universität Duisburg-Essen: 'Review of European assessment methods for rivers and streams'. Beitrag zu EG-Forschungsprojekt (Note: 1,1)
- 09/95 – 08/03 - Universität Duisburg-Essen: Studium der Ökologie, Biologie und Philosophie
- 01/95 – 09/95 - Reisen im Europäischen Ausland sowie mehrwöchige Auftrittstournee in Deutschland (Musik)
- 10/93 – 12/94 - Zivildienst beim Sozialdienst Katholischer Männer, Bottrop.
- 09/84 – 07/93 - Josef-Albers-Gymnasium Bottrop; Abschluss Abitur (Note: 2,0)
- 08/80 – 07/84 - Droste-Hülshoff Grundschule Bottrop

Erklärung

Hiermit erkläre ich, gem. § 6 Abs. 2, Nr. 6 der Promotionsordnung der Math.-Nat.-Fachbereiche zur Erlangung des Dr. rer. nat., dass ich die vorliegende Dissertation selbständig verfasst und mich keiner anderen als der angegebenen Hilfsmittel bedient habe.

Essen, den 9. April 2009

Sebastian Birk

Erklärung

Hiermit erkläre ich, gem. § 6 Abs. 2, Nr. 7 der Promotionsordnung der Math.-Nat.-Fachbereiche zur Erlangung der Dr. rer. nat., dass ich das Arbeitsgebiet, dem das Thema *„Intercalibration of national methods to assess the ecological quality of rivers in Europe using benthic invertebrates and aquatic flora“* zuzuordnen ist, in Forschung und Lehre vertrete und den Antrag von Herrn Sebastian Birk befürworte.

Essen, den 9. April 2009

Prof. Dr. Daniel Hering

Erklärung

Hiermit erkläre ich, gem. § 6 Abs. 2, Nr. 8 der Promotionsordnung der Math.-Nat.-Fachbereiche zur Erlangung des Dr. rer. nat., dass ich keine anderen Promotionen bzw. Promotionsversuche in der Vergangenheit durchgeführt habe und dass diese Arbeit von keiner anderen Fakultät abgelehnt worden ist.

Essen, den 9. April 2009

Sebastian Birk